PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Gap between Real-world Data and Clinical Research within Hospitals in China : A Qualitative Study
AUTHORS	Jin, Feifei; Yao, Chen; Yan, Xiaoyan; Dong, Chongya; Lai, Junkai; Li, Li; Wang, Bin; Tan, Yao; Zhu, Sainan

VERSION 1 – REVIEW

REVIEWER	Dr. Mary McHugh Retired, United States of America
REVIEW RETURNED	15-May-2020

GENERAL COMMENTS	Thank you for submitting this very interesting work. Given the hope that electronic medical records will facilitate and speed much important research, a study of the problems of using EMR data for research is important and timely. This reviewer thinks that the results and conclusions need to more clearly state the problems found. The discussion is especially concerning because it fails to mention that the biggest barrier is that the great hope of EMR data--that it can be directly downloaded into the research files was not achieved should be more strongly reported in the conclusions section--the biggest problems were: 1. Much data were stored in PDF and image files which mean those data cannot be directly downloaded into a computer research data file. They must be manually transcribed. 2. The different platforms used in different hospitals present a barrier to research because it makes it difficult to match variables across settings. These two problems alone mean that the benefits of EMR data storage for research have been lost and that should be made more clear to readers. And the recommendation should be that the systems should be revised so that virtually all data are entered as numeric data in clinical reports, lab and other ancillary dept. reports, and in so far as possible, all clinician notes should be entered as check lists or numbers in a clinical flow sheet rather than as narrative. Storing numeric patient care data as images or PDF files is disastrous for data retrieval for any purpose other than patient care. These are the real findings of this study which should be clearly presented in both the abstract and the conclusions section, as well as recommendations section. The writing is problematic. First, the author's writing style is excellent. The manuscript is very interesting and easy to read because the logic and flow are quite good. That said, there are many syntax, grammar, and word usage problems. It is strongly recommended that the authors employ a native English speaking

	editor to correct the many English language problems in the manuscript. Just two examples of the edits required are the following: Page 4, line 9: Change the word, "is" to "are" Page 4, line 10. Change, " path a way for exploring benefit and limitation" to "provide a way to explore the benefits and limitations"

REVIEWER	Karen Day The University of Auckland
REVIEW RETURNED	06-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to review your manuscript titled 'The Gap between Real-world Data and Clinical Research Initiated by Doctors in Chinese Hospitals: A Qualitative Study'. This paper is of interest to researchers, clinicians and data analysts in countries that are emerging with EMRs and other technologies for collecting and using patient data. The research question is clearly stated and worth exploring. Grounded Theory is a good way to do this when gathering data about issues, challenges and what works for different people in the hospital workforce. You could improve this manuscript in the following ways. - There is a bigger body of literature than mentioned in the introduction section. You use the term 'real world data' which gave you a small set of research literature to use. However, if you had searched using 'routinely collected data' and/or 'big data' you would have found a similar study to your own plus a large number of articles to help you (1) derive your research question and (2) do constant comparison between the findings and the literature. - You state that you have chosen to use Grounded Theory (GT), which is appropriate for your research question. However, there are few signs that you actually used it. It looks like you identified a 'community concern' (that doctors want to use EMR data but cannot because of certain barriers), but you do not state that it's a concern or define what a GT community concern is. You should indicate which type of GT you used (the type frames what you do and how you do it), why you used GT and not general qualitative research, how you decided on the set of questions and the order in which you did the interviews, how you analysed the data (in a way that distinguishes it from 'usual' qualitative data analysis processes), if you used theoretical saturation (if not, why), how you used memo-writing in the analysis, and if your results enable you to create a tentative theory or model. You should at the least cite GT theory in your description of the methods. - The findings are interesting and it looks like you've got something good to write about. Please avoid presenting raw data in your findings (the tables) – the reader is interested in the higher analytical levels of what you saw and how you made sense of the data in your analysis. The occasional quote to underline a point you make is acceptable but full tables are not useful to a reader who is left feeling like they should analyse the data for you. The figure is interesting and potentially useful for readers but you do not introduce or discuss it in the text. - Since the findings are not clearly enough presented, it is not easy for me to judge the value of the discussion. Once you have provided a GT framework for your research, and have tightened the findings section you will need to rewrite the discussion. All the best with your research.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Dr. Mary McHugh
Institution and Country: Retired, United States of America
Competing interests: None

Reviewer's Comment:
Thank you for submitting this very interesting work. Given the hope that electronic medical records will facilitate and speed much important research, a study of the problems of using EMR data for research is important and timely.

1.This reviewer thinks that the results and conclusions need to more clearly state the problems found. The discussion is especially concerning because it fails to mention that the biggest barrier is that the great hope of EMR data--that it can be directly downloaded into the research files was not achieved should be more strongly reported in the conclusions section--the biggest problems were:
1) Much data were stored in PDF and image files which mean those data cannot be directly downloaded into a computer research data file. They must be manually transcribed.
2) The different platforms used in different hospitals present a barrier to research because it makes it difficult to match variables across settings.
These two problems alone mean that the benefits of EMR data storage for research have been lost and that should be made more clear to readers.
Response:
Thank you very much for your suggestions. We have emphasized the biggest barrier and the reasons in results and conclusion as suggested by the reviewer. Two reasons are summarized as unstructured EMR data and lack of data interoperability. At the same time, we explained in detail why these reasons affect the use of EMR data and added research results from other countries in discussion. In addition, data security concerns by information technology departments and managers is also one of the reasons, especially in China. Therefore, we added this reason in results and explain it in discussion.

2.And the recommendation should be that the systems should be revised so that virtually all data are entered as numeric data in clinical reports, lab and other ancillary dept. reports, and in so far as possible, all clinician notes should be entered as check lists or numbers in a clinical flow sheet rather than as narrative. Storing numeric patient care data as images or PDF files is disastrous for data retrieval for any purpose other than patient care. These are the real findings of this study which should be clearly presented in both the abstract and the conclusions section, as well as recommendations section.
Response:
Thank you very much for this important suggestion. We have added the recommendations in results and conclusion. Two suggestions are summarized as update hospital information system and promote data standards. At the same time, we explained them in detail and added research results from other countries in discussion. In addition, the suggestion for data security concerns was establishment of an independent clinical research platform.

3.The writing is problematic. First, the author's writing style is excellent. The manuscript is very interesting and easy to read because the logic and flow are quite good. That said, there are many syntax, grammar, and word usage problems. It is strongly recommended that the authors employ a native English speaking editor to correct the many English language problems in the manuscript.

Just two examples of the edits required are the following:
Page 4, line 9: Change the word, "is" to "are"
Page 4, line 10. Change, " path a way for exploring benefit and limitation" to "provide a way to explore the benefits and limitations"
Response:
Thank you very much for your suggestion. We have carefully revised the manuscript and the writing of the manuscript was improved by a native speaker.

Reviewer: 2
Reviewer Name: Karen Day
Institution and Country: The University of Auckland
Competing interests: None declared

Reviewer's Comment:
Thank you for the opportunity to review your manuscript titled 'The Gap between Real-world Data and Clinical Research Initiated by Doctors in Chinese Hospitals: A Qualitative Study'. This paper is of interest to researchers, clinicians and data analysts in countries that are emerging with EMRs and other technologies for collecting and using patient data.
The research question is clearly stated and worth exploring. Grounded Theory is a good way to do this when gathering data about issues, challenges and what works for different people in the hospital workforce.
You could improve this manuscript in the following ways.
1.There is a bigger body of literature than mentioned in the introduction section. You use the term 'real world data' which gave you a small set of research literature to use. However, if you had searched using 'routinely collected data' and/or 'big data' you would have found a similar study to your own plus a large number of articles to help you (1) derive your research question and (2) do constant comparison between the findings and the literature.
Response:
Thank you very much for your suggestion. RWD contains many kinds of data as suggested by the reviewer. Therefore, we have added research results from other scholars and countries in introduction by searching 'EMR data', 'routinely collected data' and 'big data' in pubmed. At the same time, we compared the results of this study and other research in discussion.

2.You state that you have chosen to use Grounded Theory (GT), which is appropriate for your research question. However, there are few signs that you actually used it. It looks like you identified a (that doctors want to use EMR data but cannot because of certain barriers), but you do not state that it's a concern or define what a GT community concern is.
Response:
Thank you very much for this important suggestion. We have added the description of qualitative research and constructivist grounded theory (CGT) to illustrate this study is suitable for GCT and state the study concern in method.

3.You should indicate which type of GT you used (the type frames what you do and how you do it), why you used GT and not general qualitative research, how you decided on the set of questions and the order in which you did the interviews, how you analysed the data (in a way that distinguishes it from 'usual' qualitative data analysis processes), if you used theoretical saturation (if not, why), how you used memo-writing in the analysis, and if your results enable you to create a tentative theory or model. You should at the least cite GT theory in your description of the methods.
Response:
Thank you very much for this important suggestion. CGT theory and it's three stages coding method have be described in design and analysis respectively. We explained the coding and memoing method and how we make sure theoretical saturation in analysis.

4.The findings are interesting and it looks like you've got something good to write about. Please avoid presenting raw data in your findings (the tables) – the reader is interested in the higher analytical levels of what you saw and how you made sense of the data in your analysis. The occasional quote to underline a point you make is acceptable but full tables are not useful to a reader who is left feeling like they should analyse the data for you. The figure is interesting and potentially useful for readers but you do not introduce or discuss it in the text.
Response:
Thank you very much for your suggestion. We deleted the table 2-table 4 and quote the interviewer's reply under the specific subtheme to underline the point in results. This will make it easier for readers to understand. In addition, a simplified summary table is also provided to readers.

5.Since the findings are not clearly enough presented, it is not easy for me to judge the value of the discussion. Once you have provided a GT framework for your research, and have tightened the findings section you will need to rewrite the discussion.
All the best with your research.
Response:
Thank you very much for your suggestion and patience. We have provided a framework for this study in figure and explained every theme and subtheme in the discussion according to the revised results.

VERSION 2 – REVIEW

REVIEWER	Dr. Mary McHugh Retired, USA
REVIEW RETURNED	28-Sep-2020

GENERAL COMMENTS	Many excellent improvements have been made to this manuscript. It is greatly improved. However, more improvement is still needed. The manuscript still carries a rather large number of English language errors. It is strongly recommended that the authors retain the services of a native English speaking editor who can correct the errors. Just one example is found as follows: Page 26 of 45, line 26. Change "sight of the value of EMR, many clinical research" to "light of the value of EMR, many clinical research projects" Due to the English language errors, some parts of the manuscript are confusing or make little sense.

REVIEWER	Karen Day The University of Auckland, New Zealand
REVIEW RETURNED	19-Aug-2020

GENERAL COMMENTS	Thank you for the changes you have made to this manuscript. It reads much better and it is easier to see what you have done in terms of grounded theory. There are still some improvements that can be made. Grammar and typos: There are still grammatical issues in this manuscript. At times it is difficult to make out what you are writing about because of long sentences. I have left comments in the margins of the attached document for you to help you make the necessary changes. Use of Grounded Theory: Thank you for clarifying this. You mention a framework and you refer to Figure 1 but there is no Figure 1 in the manuscript or on the supplementary documents. I would be

	interested to see how your data categories work in the framework and how the different categories are linked (if there are links). Methods section: There are several statements that should be supported by references to the literature. Please indicate why you chose semi-structured interviews and what the difference is between group interviews and focus groups – dynamics change how and what data you get. Findings section: I'm still not convinced that a table with raw quotes is needed. You have provided good descriptions of the analysed data in the text and repeated some of the quotes, which seems redundant. Discussion: Overall, the discussion section has improved. I like how you've matched the discussion to the findings and in turn to the research purpose. However, there are several minor issues that still need to be addressed (see my comments in the margins). In particular, I recommend that you discuss the points about interoperability in light of global information standards that are currently in use so that it is clear that the hospitals you used either use or do not use these standards. This would be form the basis of any conclusions you make about aspects of interoperability in your findings. I also recommend that you find out what privacy legislative requirements already exist in China and how they inform the policies and protocols that the IT department have put in place to protect personal information in EMRs. The same laws should apply to research databases and clinical practice – you should also reference standard privacy protection behaviour in clinical practice and how it overlaps EMR and research database use. The conclusion section should be rewritten in order to be strengthened.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Dr. Mary McHugh
Institution and Country: Retired, United States of America
Competing interests: None
Reviewer's Comment:
Many excellent improvements have been made to this manuscript. It is greatly improved. However, more improvement is still needed.
1.The manuscript still carries a rather large number of English language errors. It is strongly recommended that the authors retain the services of a native English speaking editor who can correct the errors. Just one example is found as follows:
Page 26 of 45, line 26. Change "sight of the value of EMR, many clinical research" to "light of the value of EMR, many clinical research projects"
Due to the English language errors, some parts of the manuscript are confusing or make little sense.
Response:
Thank you very much for your suggestion. We have carefully revised the manuscript and asked a professional English editing company to modify it. All changes have been highlighted in red (or yellow) in the revised manuscript.

Reviewer: 2
Reviewer Name: Karen Day
Institution and Country: The University of Auckland
Competing interests: None declared

Reviewer's Comment:
Thank you for the changes you have made to this manuscript. It reads much better and it is easier to see what you have done in terms of grounded theory. There are still some improvements that can be made.
1.Grammar and typos: There are still grammatical issues in this manuscript. At times it is difficult to make out what you are writing about because of long sentences. I have left comments in the margins of the attached document for you to help you make the necessary changes.
Response:
Thank you very much for your suggestion and careful comments. We have deleted this table. We have carefully revised the manuscript and asked a professional editing company for help with language revision.

2.Use of Grounded Theory: Thank you for clarifying this. You mention a framework and you refer to Figure 1 but there is no Figure 1 in the manuscript or on the supplementary documents. I would be interested to see how your data categories work in the framework and how the different categories are linked (if there are links).
Response:
Thank you very much for your suggestion. We checked the uploaded manuscript, and Figure 1 is attached; we are not sure why you were unable to access it. In the revised manuscript, we have confirmed that Figure 1 has been attached to the manuscript.

3.Methods section: There are several statements that should be supported by references to the literature. Please indicate why you chose semi-structured interviews and what the difference is between group interviews and focus groups – dynamics change how and what data you get.
Response:
Thank you very much for your suggestion. We have modified the text according to your comments and added references. Please see below for detailed modifications.

4.Findings section: I'm still not convinced that a table with raw quotes is needed. You have provided good descriptions of the analysed data in the text and repeated some of the quotes, which seems redundant.
Response:
Thank you very much for your suggestion. We have deleted this table.

5.Discussion: Overall, the discussion section has improved. I like how you've matched the discussion to the findings and in turn to the research purpose. However, there are several minor issues that still need to be addressed (see my comments in the margins). In particular, I recommend that you discuss the points about interoperability in light of global information standards that are currently in use so that it is clear that the hospitals you used either use or do not use these standards. This would be form the basis of any conclusions you make about aspects of interoperability in your findings. I also recommend that you find out what privacy legislative requirements already exist in China and how they inform the policies and protocols that the IT department have put in place to protect personal information in EMRs. The same laws should apply to research databases and clinical practice – you should also reference standard privacy protection behaviour in clinical practice and how it overlaps EMR and research database use.
Response:

Thank you very much for your suggestion. We have modified the text according to your comments and added references. Please see below for detailed modifications.

6.The conclusion section should be rewritten in order to be strengthened.
Response:
Thank you very much for your suggestion. We have rewritten the conclusion according to the revised results. We have reiterated the purpose and design of the research and added the next research plan.

KD1: Avoid using acronyms in the abstract
Response:
Thank you very much for your suggestion. All acronyms in the abstract have been replaced with full names.

KD2: Avoid using contractions, i.e. write the words out in full in formal writing
Thank you very much for your suggestion. We have revised all contractions in the manuscript.

KD3: Grammar issue in this sentence.
Response:
Thank you very much for your suggestion. We have revised the sentence.

KD4: Roles'
Response:
Thank you very much for this suggestion.

KD5: Feels like a contradiction when you say you used stratified purposive sampling and then you tell us that selection bias occurred. Didn't you try to reduce selection bias by using stratified purposive sampling?
Response:
Thank you very much for your suggestion. Stratified purposive sampling is a way to reduce selection bias. We have revised this sentence in accordance with your suggestion.

KD6: How is this relevant as a strength or weakness of the study? One would normally assume that ethics committee members remain separated from a research project.
Response:
Thank you very much for your suggestion. We have deleted this sentence.

KD7: Light?
Response:
Thank you very much for your suggestion. We have revised this sentence.

KD8: One should not use words like 'enormous' unless one can prove that the gap is indeed enormous. It is sufficient to say in your introduction that there is a gap, because you will describe it in your findings. 'Enormous' simply exaggerates where an exaggeration is not necessary.
Response:
Thank you very much for this suggestion. We have deleted the 'enormous'.

KD9: Is not
Response:
Thank you very much for this suggestion. We have revised this sentence.

KD10: 'Cannot'. I won't correct any more of your contractions – you know what to do.
Response:

Thank you very much for this suggestion. We have revised this sentence and carefully revised the manuscript.

KD11: Please cite literature to support this decision, i.e. why is it good to gain multiple perspectives
Response:
Thank you very much for your suggestion. We have added 2 papers to the references.
1. Bentley T, Rizer M, McAlearney AS, et al. The journey from precontemplation to action: Transitioning between electronic medical record systems. Health Care Manage Rev 2016;41(1):22-31. doi: 10.1097/HMR.0000000000000041 [published Online First: 2014/10/18]
2. Denny E, Weckesser A. Qualitative research: what it is and what it is not: Study design: qualitative research. BJOG 2019;126(3):369. doi: 10.1111/1471-0528.15198 [published Online First: 2018/06/20]

KD12: Why stratified purposive sampling? Please cite the literature to support your decision.
Response:
Thank you very much for your suggestion. We have added 2 papers to the references.
1. Setia MS. Methodology Series Module 5: Sampling Strategies. Indian J Dermatol 2016;61(5):505-9. doi: 10.4103/0019-5154.190118 [published Online First: 2016/10/01]
2. Moser A, Korstjens I. Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. Eur J Gen Pract 2018;24(1):9-18. doi: 10.1080/13814788.2017.1375091 [published Online First: 2017/12/05]

KD13: Please cite literature to support your definition.
Response:
Thank you very much for your suggestion. We have added 1 paper to the references.
1. Corbin JM, Strauss AL. The Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory: SAGE.

KD14: Please explain ‘Clinical research experience (Doctors specific)’
Response:
Thank you very much for your suggestions. Our objective is to investigate the gap between real-world data (RWD) and clinical research initiated by doctors in China, explore the potential reasons for this gap and collect different stakeholders’ suggestions. When doctors conduct clinical research, they need to extract the data in EMRs as research data and be able to identify problems with data extraction. Doctors who do not conduct clinical research may not be able to provide useful information. Therefore, the doctors included in this study had clinical research experience that allowed them to express their views on the subject of this study. We have added this explanation the information on the participants.

KD15: Why semi-structured interviews? Please supply a citation to support your decision
Response:
Thank you very much for your suggestion. We have added 3 papers to the references.
1. Whiting LS. Semi-structured interviews: guidance for novice researchers. Nurs Stand 2008;22(23):35-40. doi: 10.7748/ns2008.02.22.23.35.c6420 [published Online First: 2008/03/08]
2. Peters K, Halcomb E. Interviews in qualitative research. Nurse Res 2015;22(4):6-7. doi: 10.7748/nr.22.4.6.s2 [published Online First: 2015/03/19]
3. Britten N. Qualitative interviews in medical research. BMJ 1995;311(6999):251-3. doi: 10.1136/bmj.311.6999.251 [published Online First: 1995/07/22]

KD16: When does a group interview become a focus group? Is there a difference? Please cite references to support your decision.
Response:
Thank you very much for your comment. Focus group interviews can provide information about a

range of ideas and feelings that individuals have about certain issues and can illuminate the differences in perspectives between groups of individuals. One of the distinct features of focus group interviews is their group dynamics; hence, the type and range of data generated through the social interactions of the group are often deeper and richer than those obtained from one-on-one interviews. We have changed “group interviews” to “focus group interviews” and added 1 paper to the references.
34. Rabiee F. Focus-group interview and data analysis. Proc Nutr Soc 2004;63(4):655-60. doi: 10.1079/pns2004399 [published Online First: 2005/04/16]

KD17: How did you design the interview guides, e.g. did you use the literature to inform your questions or did you have conversations with clinicians about their issues to establish a broad description of the problem, i.e. find out what the ‘community concern’ is as one would in a CGT project? How did the questions change over the
course of the interviews?
Response:
Thank you very much for your suggestion. We conducted conversations with the interviewees about their issues to establish a broad description of the problem. During the interviews, we asked different questions based on the interviewee’s answers.

KD18: Of what research group? Is your research a subset of a larger research project?
Response:
Thank you very much for your comment. Our research is not a subset of a larger research project. JFF, DCY, TY and LHQ were a group in this qualitative study. To avoid confusion, we have deleted “the research group”.

KD19: Not relevant to state this.
Response:
Thank you very much for your suggestion. We have deleted this sentence.

KD20: How does this structure fit into your CGT process? Is it necessary to have a supplementary document with details about how you went about setting up and conducting the interviews, or is it more expedient to summarise the process in a few sentences?
Response:
Thank you very much for your comment. The structured process can provide detailed operational guidance for some researchers who are conducting qualitative interviews for the first time. Before we conducted qualitative interviews, we spent considerable time reviewing information and formulating a detailed structured interview process. This ensured that every interview was conducted in the same way. We hope to provide readers with a reference manual to help researchers who have no interview experience develop their own interview plans. Therefore, we hope that this structured process can be attached to the paper as a Supplement.

KD21: Nice description

KD22. Individual? You need to resolve the difference between group and individual interviews – there are different dynamics (e.g. discussion among participants) that occur when more than one person is talking.
Response:
Thank you very much for your comment. We have deleted the word “personal”. In-depth interviews and focus group interviews are two methods of qualitative interviews.

KD23. In the methods section you call them ‘clinical managers’ and now you call them ‘clinical directors’. Which one is it? Managers are different from directors.
Response:

Thank you very much for your suggestion. It should be ‘clinical managers’.

KD24: Are you sure you need to exaggerate the concept of ‘hope’?
Response:
Thank you very much for your suggestion. We have deleted the word ‘great’.

KD25: If you use this strong word, you’re obliged to provide proof from your data that the concerns were actually excessive.
Response:
Thank you very much for your suggestion. We have deleted the word ‘excessive’.

KD26: I’m assuming that the interviews were translated for this manuscript. Please correct the grammar in the translations.
Response:
Thank you very much for your suggestion. We have corrected this issue.

KD27: Why use the same quote that is already in the table above?
Response:
Thank you very much for your comment. We have deleted the table.

KD28: Is?
Response:
Thank you very much for your suggestion. We have revised the word.

KD29: Long sentence. I can’t make sense of what you’re saying here.
Response:
Thank you very much for your suggestion. We have revised the sentence.

KD30: It’s not clear if the platform already exists or if the participants are suggesting that a new one needs to be built.
Response:
Thank you very much for your comment. The hospital plans to build a platform. We have revised the sentence.

KD31: EMR?
Response:
Thank you very much for your suggestion. We have revised the word.

KD32: Your discussion opens with a silent assumption that you confirmed that there is indeed a gap – please make this explicit.
Response:
Thank you very much for your suggestion. We have revised the sentence.

KD33: I can’t work out if you’re citing research to support your finding or if you’re summarizing research that explains your finding that EMR data can’t be used in a separate research platform. This is because the causes you describe in the findings section are the three points you make in this sentence.
Response:
Thank you very much for your question. We are citing research to support our finding.

KD34: Interoperability is an issue inside an EMR where several software products are compiled to deliver an EMR. If you don’t solve the internal EMR interoperability problem you’re not likely to solve

data standards problems when migrating data into a clinical research platform. Your thoughts?
Response:
Thank you very much for your suggestion. In the same hospital, the inpatient system, outpatient system and examination system are independent. Therefore, when a clinician wants to search a patient's outpatient data, the data are difficult to find. This leads to a lack of interoperability of data in the hospital. Moreover, because of the large number of patients, the EMR system only saves the information of patients who have visited within six months or one year. If all medical information of a patient is migrated to the clinical research platform, the doctor can easily search all medical information of the patient.

KD35: Exchange?
Response:
Thank you very much for your suggestion. We have revised the word.

KD36: Supply reference. Research is considered a secondary use of clinical data.
Response:
Thank you very much for your suggestion. We agree that research is a secondary use of clinical data. EMRs typically contain general information such as the treatment and medical history of a patient as it is collected by the individual medical practice. Therefore, the primary role of EMRs is to help clinicians complete clinical practice. We have revised “EMRs are mainly used for clinical practice” and added 2 papers to the references.

KD37: Don’t Chinese hospitals and laboratories use laboratory coding systems, e.g. LOINC? I’d be surprised if they don’t. Therefore why is lab data not directly imported into the EMR and why are pdf formats used for these reports?
Response:
Thank you very much for your comment. Some laboratories already use coding systems in China, but the data storage time is very short. After a period of time, the laboratory results are printed and saved as a PDF file. Therefore, when a doctor wants to search the information of a past patient, he or she can only see the data in PDF format.

KD38: I agree with your point here but it needs to be clearer in terms of the interoperability requirement for diagnostic data and images such as radiology (which comes with a DICOM standard and should be readable and exportable to a research platform). Are you suggesting that over and above information standards such as
DICOM and LOINC that there should also be use of Artificial Intelligence for unstructured text analysis in the EMR?
Response:
Thank you very much for your suggestion. As you said, we believe that artificial intelligence should be used for unstructured text analysis, and we have added this to the manuscript.

KD39: Is SNOMED-CT or something similar already being used? Your statement assumes that no information standards are being used at all, yet information standards are internationally accepted to some extent in most EMRs
Response:
Thank you very much for your suggestion. We have revised the description of this paragraph. Similar to the laboratory coding system, SNOMED-CT has also been used in some hospitals. The problem is that the data storage time is very short. Most of the patient treatment records in EMR are saved in an unstructured text format, which cannot be directly used for data analysis.

KD40: Please supply a reference to support this point.
Response:

Thank you very much for your suggestion. We have added 3 papers to the references.
1. Takenouchi K, Yuasa K, Shioya M, et al. Development of a new seamless data stream from EMR to EDC system using SS-MIX2 standards applied for observational research in diabetes mellitus. Learn Health Syst 2019;3(1):e10072. doi: 10.1002/lrh2.10072 [published Online First: 2019/06/28]
2. Jung SY, Kim JW, Hwang H, et al. Development of Comprehensive Personal Health Records Integrating Patient-Generated Health Data Directly From Samsung S-Health and Apple Health Apps: Retrospective Cross-Sectional Observational Study. JMIR Mhealth Uhealth 2019;7(5):e12691. doi: 10.2196/12691 [published Online First: 2019/05/30]
3. Aldaz G, Shluzas LA, Pickham D, et al. Hands-free image capture, data tagging and transfer using Google Glass: a pilot study for improved wound care management. PLoS One 2015;10(4):e0121179. doi: 10.1371/journal.pone.0121179 [published Online First: 2015/04/23]

KD41: This point was not indicated in your findings. You can't introduce new findings in the discussion.
Response:
Thank you very much for your suggestion. We have deleted the sentence.

KD42: You make good points in this paragraph, mostly about the potential of a research platform to protect patient data privacy. What is interesting is that you have not cited any privacy legislation that affects EMR structures and privacy protocols in the different hospitals –what legislation exists and what implications does that legislation have on the ability to protect patient data privacy in the EMR? How does that legislation transfer to the research platform?
Response:
Thank you very much for your suggestion. We have added the legislation. In 2017, the Chinese government issued the "The Cybersecurity Law of the People's Republic of China" and "Regulations for the Application of Electronic Medical Records (Trial)". According to document requirements, medical institutions should strictly protect patient privacy. Data sharing can only be performed if the safety of patient electronic data is ensured. The platform would be a good way to integrate research-specific data and routinely collected data as well as to de-identify the data to be used for statistical analysis to reduce the risk of a data leak.

KD43: A good conclusion (1) reiterates the purpose of the research (not done here),(2) summarises the research strategy/design (not done) (3) summarises the findings (done), (4) indicates some actual conclusions (partially done), and (5) finishes with a call to action, not necessarily what future research might be about, but could be the next step now that this part of the research has been completed.
Response:
Thank you very much for your suggestion. We have rewritten the conclusion according to the revised results. We have reiterated the purpose and design of the research and added the future research plan.

VERSION 3 – REVIEW

REVIEWER	Dr. Mary McHugh None--retired
REVIEW RETURNED	10-Nov-2020

GENERAL COMMENTS	Review of the manuscript: The Gap between Real-world Data and Clinical Research within Hospitals in China : A Qualitative Study. 1. The first paragraph of any research report should clearly state purpose of the study. The purpose (aim) is beautifully stated in the abstract, and this sentence from the abstract should be appended to the end of the first paragraph of the paper's introduction.

	2. On page 4 (lines 26-27) the author states: “the current problem is that EMRs are generally a separate system that is not directly connected to EDC systems”. However, given concerns about confidentiality of data, it is unlikely that any responsible person would allow these two systems to be directly connected. Anyway, even if you were to put a cable between the two computers and directly connect them, that would not solve the problem. It would be better to state that the two systems cannot communicate—and as your sources note, that has been a common problem in many countries. Later on you discover that too much of the EMR is narrative, but that is for the results section, not for the introduction section. The reviewer is an informaticist, and in America, we have been concerned with this problem for years. We have identified that the real problem is that data stored in the EMR uses data coding and data storage schemes that are incompatible with the coding and storage formats of EDC systems. We don’t have as much of a problem with narrative in the EMR systems. So the author might want to consider amending this paper to conclude: Both systems should be worked on to ensure that first, the clinical systems use data recording formats that are primarily in coded formats. In the U.S., a lot of hospitals have clinicians store data through use of DBMS tables that format and store the data, and user interfaces that clinicians use to quickly chart. They can do this because the user interface presents screens that allow most charting to be entered via radio buttons or numeric data entry. For example, instead of writing, “Ciprofloxacin 500 mg. twice a day” the clinician would order via a pharmacy system. In that system, the physician clicks on “antibiotics,” then clicks on “Cipro,” then clicks on the dosage, and finally clicks on the frequency of administration. That system then automatically transcribes the order onto the nurses’ medication administration record (MAR). The pharmacy and nursing systems code and store those data in exactly the same formats. So it is then possible for researchers to use those formats in setting up their EDC. So, the real solution here is two-fold: First, the EMR must be converted from narrative to codified data formats, and then the EDC must be designed to use the same coding and storage formats as the EMR. Then when permission is given to mine the clinical data system, an interface between the two systems can be written so that the data can be electronically extracted from the EMR and placed in the EDC. 3. Overall, the introductory section does a good job of introducing the research problem. I think the importance of the problem could be better explained with few sentences explaining that there is a wealth of information stored in clinical practice data. That information could be used to evaluate clinical practice and discover which care protocols are most successful so that information could be disseminated to practitioners who then would have evidence for how to quickly improve care and care outcomes. This approach might provide much faster improvement in care than can traditional prospective research protocols. 4. The author chose an excellent research design to investigate the research question. 5. The research methods are very well described and were

	appropriate to the research design and research questions. 6. The reviewer particularly appreciated the thorough and clear explanation of the results and analysis, which was exactly the correct way to analyze this kind of qualitative data. The results strongly support the conclusions and recommendations presented by the author. The paper is extremely well written, and the entire study is strong, coherent, and congruent with the science of conducting qualitative studies.

REVIEWER	Karen Day University of Auckland, New Zealand
REVIEW RETURNED	01-Nov-2020

GENERAL COMMENTS	Thank you for the adjustments you have made in response to the reviews. The manuscript flows much better and the clarity is improved. Thank you for addressing the English language errors that have been pointed out to you. The revision has created new errors that you need to correct.

VERSION 3 – AUTHOR RESPONSE

Reviewer: 2

Reviewer Name: Karen Day
Institution and Country: University of Auckland, New Zealand
Competing interests: None declared

Comments to the Author:

Thank you for the adjustments you have made in response to the reviews. The manuscript flows much better and the clarity is improved. Thank you for addressing the English language errors that have been pointed out to you. The revision has created new errors that you need to correct.
Response:
Thank you very much for your suggestion. We carefully revised the manuscript and asked a professional English editing company to modify the manuscript again. All changes are highlighted in red in the revised manuscript.

Reviewer: 1

Reviewer Name: Dr. Mary McHugh
Institution and Country: None--retired
Competing interests: None declared

Comments to the Author:

Review of the manuscript: The Gap between Real-world Data and Clinical Research within Hospitals in China : A Qualitative Study.
1. The first paragraph of any research report should clearly state purpose of the study. The purpose (aim) is beautifully stated in the abstract, and this sentence from the abstract should be appended to

the end of the first paragraph of the paper's introduction.
Response:
Thank you very much for your suggestion. We revised the introduction. The research purpose was added in the last sentence of the introduction.

2. On page 4 (lines 26-27) the author states: "the current problem is that EMRs are generally a separate system that is not directly connected to EDC systems". However, given concerns about confidentiality of data, it is unlikely that any responsible person would allow these two systems to be directly connected. Anyway, even if you were to put a cable between the two computers and directly connect them, that would not solve the problem. It would be better to state that the two systems cannot communicate—and as your sources note, that has been a common problem in many countries. Later on you discover that too much of the EMR is narrative, but that is for the results section, not for the introduction section.
Response:
Thank you very much for your suggestion. We stated that EMRs and EDC systems cannot communicate in the results.

The reviewer is an informaticist, and in America, we have been concerned with this problem for years. We have identified that the real problem is that data stored in the EMR uses data coding and data storage schemes that are incompatible with the coding and storage formats of EDC systems. We don't have as much of a problem with narrative in the EMR systems. So the author might want to consider amending this paper to conclude: Both systems should be worked on to ensure that first, the clinical systems use data recording formats that are primarily in coded formats. In the U.S., a lot of hospitals have clinicians store data through use of DBMS tables that format and store the data, and user interfaces that clinicians use to quickly chart. They can do this because the user interface presents screens that allow most charting to be entered via radio buttons or numeric data entry. For example, instead of writing, "Ciprofloxacin 500 mg. twice a day" the clinician would order via a pharmacy system. In that system, the physician clicks on "antibiotics," then clicks on "Cipro," then clicks on the dosage, and finally clicks on the frequency of administration. That system then automatically transcribes the order onto the nurses' medication administration record (MAR). The pharmacy and nursing systems code and store those data in exactly the same formats. So it is then possible for researchers to use those formats in setting up their EDC.
So, the real solution here is two-fold: First, the EMR must be converted from narrative to codified data formats, and then the EDC must be designed to use the same coding and storage formats as the EMR. Then when permission is given to mine the clinical data system, an interface between the two systems can be written so that the data can be electronically extracted from the EMR and placed in the EDC.
Response:
Thank you very much for your suggestions and detailed guidance. The experience and attempts in the United States are very meaningful for China's electronic medical data management. We revised the results and discussion. We stated that the data stored in the EMR uses data coding and data storage schemes that are incompatible with the coding and storage formats of EDC systems in the discussion of the causes. Additionally, we stated that EDC systems and EMR systems should be revised and that both systems should use the same coding and storage formats in the discussion.

3. Overall, the introductory section does a good job of introducing the research problem. I think the importance of the problem could be better explained with few sentences explaining that there is a wealth of information stored in clinical practice data. That information could be used to evaluate clinical practice and discover which care protocols are most successful so that information could be

disseminated to practitioners who then would have evidence for how to quickly improve care and care outcomes. This approach might provide much faster improvement in care than can traditional prospective research protocols.
Response:
Thank you very much for your suggestion. We revised and deleted some sentences in the introduction.

4. The author chose an excellent research design to investigate the research question.
Response:
Thank you very much for your comments.

5. The research methods are very well described and were appropriate to the research design and research questions.
Response:
Thank you very much for your comments.

6. The reviewer particularly appreciated the thorough and clear explanation of the results and analysis, which was exactly the correct way to analyze this kind of qualitative data. The results strongly support the conclusions and recommendations presented by the author. The paper is extremely well written, and the entire study is strong, coherent, and congruent with the science of conducting qualitative studies.
Response:
Thank you very much for your comments.