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)	Analysis of multimorbidity networks associated with different factors in Northeast China: A cross-sectional analysis
AUTHORS	Yu, Jianxing; Li, Yingying; Zheng, Zhou; Jia, Huanhuan; Cao, Peng; Qiangba, Yuzhen; Yu, Xihe;

VERSION 1 – REVIEW

REVIEWER	Kalgotra, Pankush Auburn University
REVIEW RETURNED	23-May-2021

GENERAL COMMENTS	The paper presented an analysis of multimorbidity network classified by gender, age and hospital duration. Tis study has contextual contributions i.e. the network method was applied on a new dataset. However, studying multimorbidity through network analysis is not novel. The paper has several limitations in writing, explaining the dataset and networks, and in conclusions. Please find my comments below: • The paper has several grammatical mistakes and should go through a comprehensive con-editing. • The introduction section presents the purpose of this study but how it contributes to the existing literature is missing. Several papers have used the network approach. Some of them are: Hidalgo et al. (2009), Kalgotra et al. (2017), Kalgotra et al. (2020), Glicksberg et al. (2016) and Aguado et al. (2020). Discussing these papers and how the current study adds to the literature is required. • The data collection part is missing. “The respondents were surveyed…” What does this mean? Were patients were asked about the diseases they have? • Need more information about multi-stage stratified sampling. Normally, stratified sampling is based on a variable. Which variable/s were used to sampling and why? • In the analysis, the networks are weighted. Which measure was used to create the weight? For instance, Hidalgo et al. (2009) used Pearson correlation but Kalgotra, P., Sharda, R., & Luse, A. (2020) suggested Salton cosine index to be a better measure. Did you use any of these or just the frequency? How your approach is better than theirs need to be discussed because the most important part of the network building is the definition of an edge. • In the results, a new term has been used: non-multimorbidity. What does it mean? May be the patients with only one disease are labeled as non-multimorbid patients? If that is the case, it becomes obvious that the patients with one disease will stay shorter than patients with multiple disease. What is surprising here? • The results needed to be tied back on the current literature. Use the papers suggested in the references. May be also use the syndemic theory explained by Singer (2000) to discuss why a patient

	with multimorbid conditions are likely to have bad outcomes. Overall, the paper needs better writing. The authors need to add more papers and discuss how this paper is different or how this paper is adding to the body of knowledge. More discussion about the findings and implications is required. Hope my comments will help improve the paper. Good luck! References Aguado, A., Moratalla-Navarro, F., López-Simarro, F., & Moreno, V. (2020). Morbinet: multimorbidity networks in adult general population. Analysis of type 2 diabetes mellitus comorbidity. Scientific reports, 10(1), 1-12. Glicksberg, B. S., Li, L., Badgeley, M. A., Shameer, K., Kosoy, R., Beckmann, N. D., ... & Dudley, J. T. (2016). Comparative analyses of population-scale phenomic data in electronic medical records reveal race-specific disease networks. Bioinformatics, 32(12), i101-i110. Hidalgo, C. A., Blumm, N., Barabási, A. L., & Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. PLoS Comput Biol, 5(4), e1000353. Kalgotra, P., Sharda, R., & Croff, J. M. (2017). Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. International journal of medical informatics, 108, 22-28. Kalgotra, P., Sharda, R., & Croff, J. M. (2020). Examining multimorbidity differences across racial groups: a network analysis of electronic medical records. Scientific reports, 10(1), 1-9. Kalgotra, P., Sharda, R., & Luse, A. (2020). Which similarity measure to use in network analysis: Impact of sample size on phi correlation coefficient and Ochiai index. International Journal of Information Management, 55, 102229. Singer, M. A dose of drugs, a touch of violence, a case of AIDS: conceptualizing the SAVA syndemic. Free Inq. Creat. Sociol.28, 13–24 (2000).

REVIEWER	Lu, Jiao Shanxi Medical University
REVIEW RETURNED	16-Jul-2021

GENERAL COMMENTS	Comments to the Author GENERAL COMMENTS: The manuscript aimed to explore multimorbidity patterns, especially those associated with different factors, using data from 431,295 inpatients in Jilin Province, China in 2018. They found the central point of disease clusters and the relationship between the disease clusters and different factors. And the effective interventions for the diseases at the central point of the disease cluster should be taken. However, in this manuscript, some sections are difficult to follow, along with some parts still in need of further clarification. REQUESTED REVISIONS: Introduction 1. There is a great deal of information in the introduction, but the sentence structures and expressions seem disordered and merit further amendments. For example, the second paragraph seems to describe the necessity of studying multimorbidity at the individual level, but the language and logic need to be modified. 2. It may be necessary to define the concept of “multimorbidity”. Does multimorbidity refer to co-existing diseases or co-existing

chronic diseases? Please read the second paper you cite carefully and try to find some authoritative definition to support your opinion.

Methods

3. Please list the assessment criteria from the sentence “through comprehensive assessment of the geographical location, economic level and health service status of each city, the cities finally included in the sample area were as follows” in detail.

4. Why set the criteria for rare diseases as disease frequency less than 1000? What is the basis for this?

5. The R package deals only with setting parameters and drawing. Please specify/justify how the weighted network was calculated.

6. In fact, there are many ways to identify multimorbidity patterns. What are the advantages of weighted networks over others and why did you choose to use this method? Please explain the reason.

Results

7. The authors point out that “The distribution of the diseases differed significantly by gender, age and hospital duration ($P < 0.001$)”, but it is unclear whether “the distribution of the diseases” means each disease or multimorbidity. This should be clear and specific.

8. In Table 2, what are the proper nouns for each disease based on, a dictionary or a thesaurus? They don’t seem to be consistent with ICD-10. In addition, do members of the general practitioner group be good at English? By the way, ICD is an abbreviation and should be capitalized when it first appears in the text.

9. The variable of ‘total cost’ in Figure 1 is not represented in the paper. In addition, the network graph shows the strength of the relationship between variables, but how to judge the relationship between sex and hospital duration is not indicated.

Discussion

10. In the first paragraph, what’s the meaning of “the inclusion of all the potential factors may have signified a greater complexity, but could also better explain the connection with diseases”?

11. In the fifth paragraph, the adverbial of comparison in this sentence of “Specogna et al found that patients with spontaneous intracerebral hemorrhage arriving at the hospital with hypertension were 31% more likely to stay in the hospital beyond 1 week per visit than non-hypertensive patients” seems inappropriate. Please explain how it is described in the original text, and what do you mean here?

Conclusions

12. There are no specific recommendations in the conclusion. The meaning of “effective interventions” and “different strategies” should be specified.

13. What does the last sentence do? Is it a suggestion or a fact?

14. The format of the references is not uniform. For example, whether the first letter of each word in the title should be capitalized; Whether the journal name should be uppercase, and so on. Please check the guideline for careful modification.

	15. In addition, I suggest you correct the English by a native. The authors may benefit from proof-reading to reduce typos and grammatical errors throughout the manuscript.

VERSION 1 – AUTHOR RESPONSE

A reply to the comments of reviewer 1 on
"Analysis of multimorbidity networks associated with different factors in Northeast China: A cross-sectional analysis"

Thank you very much for your helpful and constructive comments on our manuscript. We have revised our manuscript following your valuable suggestions. If there are other questions about the paper, we hope you could not hesitate to point out and to help us improve the quality of my paper. The following is the point-by-point responses to your comments.

1. The paper has several grammatical mistakes and should go through a comprehensive con-editing.
Response: Thank you very much for your good suggestion and constructive comments to this paper. And we have revised English language of our manuscript by Editage (www.editage.cn).
2. The introduction section presents the purpose of this study but how it contributes to the existing literature is missing. Several papers have used the network approach. Some of them are: Hidalgo et al. (2009), Kalgotra et al. (2017), Kalgotra et al. (2020), Glicksberg et al. (2016) and Aguado et al. (2020). Discussing these papers and how the current study adds to the literature is required.
Response: Thank you for pointing out this and offering pretty advice. We have revised the introduction, and added the current studies, see details on Page 3 Line 86-94.
3. The data collection part is missing. "The respondents were surveyed…" What does this mean? Were patients were asked about the diseases they have?
Response: We are so sorry for missing the data collection part, and we have added it. These patients were inpatients, their diagnosis was given by the doctor, and then we exported information about their hospitalization from the electronic medical record system (EMRs) or hospital information system (HIS). See detail on Page 4 Line 106-108.
4. Need more information about multi-stage stratified sampling. Normally, stratified sampling is based on a variable. Which variable/s were used to sampling and why?
Response: Thank you for pointing out this question. The multi-stage stratified sampling according to administrative region. First, through comprehensive assessment of the geographical location, economic level and health service status of each city, the cities finally included in the sample area were as follows: Changchun, Jilin, Yanbian, Tonghua and Baicheng. And these cities are located in different locations in Jilin Province (Please see the map below). And the GDP ranking of these cities is also representative (See Table 1a). At the same time, we consulted related experts and managers in the field of health services. Finally, we choose these cities. Next, one district and two counties were selected from these locations. Finally, 15 general hospitals were selected as the monitoring institutions for this study based on the administrative division.

Table1a GDP of different cities in Jilin Province in 2019 (100 million \Yuan)
ranking City GDP
1 Changchun 5904.14
2 Jilin 1414.5
3 Songyuan 729.78
4 Tonghua 725.76
5 Yanbian 723.37

6 Baishan 509.66
7 Siping 492
8 Baicheng 491.55
9 Liaoyuan 410.38

5. In the analysis, the networks are weighted. Which measure was used to create the weight? For instance, Hidalgo et al. (2009) used Pearson correlation but Kalgotra, P., Sharda, R., & Luse, A. (2020) suggested Salton cosine index to be a better measure. Did you use any of these or just the frequency? How your approach is better than theirs need to be discussed because the most important part of the network building is the definition of an edge.
Response: Thank you for your good suggestion and pointing out this question. The edge in the network represented the co-occurrence of a multimorbidity pair, and the weight of the edge was proportional to the prevalence of each multimorbidity pair. Hidalgo et al. (2009) and Kalgotra, P., Sharda, R., & Luse, A. (2020) researches are powerful and groundbreaking. In Hidalgo et al.'s paper, node size is proportional to disease prevalence. Link color indicates correlation strength, and Kalgotra et al.'s research, a circle is a disease; an edge represents a comorbidity. Size of a node explains how well it is connected to other nodes. However, we consider disease as counts variable, not quantitative data, and we also wanted to describe the association between diseases in a relatively simple and clear form. Therefore, we chose this method, node size is proportional to disease prevalence and the thickness of each line, which connects the nodes, represents the degree of association between two nodes. Most importantly, the methods may be different, but the results are consistent. For example, comparing the results of the present study to those by Hidalgo et al on human phenotype using a dynamic network approach can also be used to verify the reliability of the results of this study. Hidalgo et al found that many diseases were associated with hypertensive diseases (HyD) or ischaemic heart diseases (IHD), consistent with the findings of the present study. Hidalgo et al. further demonstrated higher comorbidity for diabetes mellitus (DME) and HyD in black men compared to white men, and this study also confirmed more notable connections for DME in male inpatients compared to female inpatients. See detail on Page 5 Line 135-142, Page 9 Line 229-235.
6. In the results, a new term has been used: non-multimorbidity. What does it mean? May be the patients with only one disease are labeled as non-multimorbid patients? If that is the case, it becomes obvious that the patients with one disease will stay shorter than patients with multiple disease. What is surprising here?
Response: Thank you for pointing out this question. We have revised table 1 and used the number of diseases to replace non-multimorbidity/ multimorbidity. And the description about Table 1 has also been modified. See detail on Page 5-6 Line 155-164.
7. The results needed to be tied back on the current literature. Use the papers suggested in the references. May be also use the syndemic theory explained by Singer (2000) to discuss why a patient with multimorbid conditions are likely to have bad outcomes.
Response: Thank you for your good suggestion. We totally agree. And we have revised this paper according to the references you suggested. See detail on Page 9 Line 219-221. Thank you again for your helpful and constructive comments.
8. Overall, the paper needs better writing. The authors need to add more papers and discuss how this paper is different or how this paper is adding to the body of knowledge. More discussion about the findings and implications is required. Hope my comments will help improve the paper. Good luck!
Response: Thank you very much for your positive comments of our manuscript and giving a lot of valuable advice on our manuscript. We have added many papers and discuss about how this paper is different or how this paper is adding to the body of knowledge, and the findings and implications also are added. See detail on Page 9, Line 221-228.

We appreciate your insightful comments that greatly improved the quality and presentation of the manuscript. We hope the revised version is acceptable to BMJ Open. Thank you again for your helpful and constructive comments.

A reply to the comments of reviewer 2 on
"Analysis of multimorbidity networks associated with different factors in Northeast China: A cross-sectional analysis"

Thank you very much for your helpful and constructive comments on our manuscript. We have revised our manuscript following your valuable suggestions. If there are other questions about the paper, we hope you could not hesitate to point out and to help us improve the quality of my paper. The following is the point-by-point responses to your comments.

Introduction

1. There is a great deal of information in the introduction, but the sentence structures and expressions seem disordered and merit further amendments. For example, the second paragraph seems to describe the necessity of studying multimorbidity at the individual level, but the language and logic need to be modified.

Response: Thank you for pointing out this and offering pretty advice. We have revised the introduction, see details on Page 3 Line 78-85.

2. It may be necessary to define the concept of "multimorbidity". Does multimorbidity refer to co-existing diseases or co-existing chronic diseases? Please read the second paper you cite carefully and try to find some authoritative definition to support your opinion.

Response: Thank you for pointing out this. Multimorbidity refers to the occurrence of two or more conditions (diseases) in one individual (Van den, A. M., Buntinx, F., & Knottnerus, J. A. (1996). Comorbidity or multimorbidity: What's in a name? A review of the literature. The European Journal of General Practice, 2, 65–70.). Comorbidity signifies the occurrence of one or more conditions beyond an index condition of interest (Feinstein, A. R. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. Journal of Chronic Diseases, 23, 455–468.). These conditions can be acute or chronic in nature, although in the vast majority of cases, reference is made to chronic conditions. Some have suggested that the definition of multimorbidity should be expanded, so that discrete medical conditions are considered alongside more integrated functional domains, such as physical functional status, mental health and well-being, and social functional status (Bayliss, E. A., Ellis, J. L., & Steiner, J. F. (2009). Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated into two other data-based morbidity measures. Journal of Clinical Epidemiology, 62(5), 550–557. el.) Therefore, we have defined the concept of multimorbidity on Page 3 Line 68-69.

Methods

3. Please list the assessment criteria from the sentence "through comprehensive assessment of the geographical location, economic level and health service status of each city, the cities finally included in the sample area were as follows" in detail.

Response: Thank you for pointing out this question. Changchun, Jilin, Yanbian, Tonghua and Baicheng are located in different locations in Jilin Province (Please see the map). And the GDP ranking of these cities is also representative (See Table 1a). At the same time, we consulted related experts and managers in the field of health services. Finally, we choose these cities.

Table1a GDP of different cities in Jilin Province in 2019 (100 million \Yuan)

ranking	City	GDP
1	Changchun	5904.14
2	Jilin	1414.5
3	Songyuan	729.78
4	Tonghua	725.76
5	Yanbian	723.37
6	Baishan	509.66

7 Siping 492
8 Baicheng 491.55
9 Liaoyuan 410.38
4. Why set the criteria for rare diseases as disease frequency less than 1000? What is the basis for this?
Response: Thank you for pointing out this. For the sake of accuracy of the study results and the clear presentation of the disease association results, 8426 inpatients were excluded with disease frequency less than 1000, accounting for a ratio of 1.95% (8426/431295=1.95%) in the analysis population. Among this population, the maximum number of patients with one disease was 912. Therefore, the data of disease frequency less than 1000 were excluded in this study. Besides, the Kalgotra et al. study also selected diseases with a frequency of more than 1000 (Kalgotra, P., Sharda, R., & Luse, A. (2020). Which similarity measure to use in network analysis: Impact of sample size on phi correlation coefficient and Ochiai index).
5. The R package deals only with setting parameters and drawing. Please specify/justify how the weighted network was calculated.
Response: Thank you for pointing out this question. The nodes represent diseases/factors, and the size of the nodes serves as the weight when they occur with all other diseases/factors. The edge in the network represented the co-occurrence of a multimorbidity pair, and the weight of the edge was proportional to the prevalence of each multimorbidity pair. When an inpatient carried more than 2 diseases, the count of every multimorbidity pair would have an increment of 1 (e.g., when an inpatient carried ischaemic heart diseases (IHD), hypertensive diseases (HyD) and cerebrovascular diseases (CD), then the multimorbidity pair IHD& HyD, IHD& CD and HyD& CD would have an increment of 1). See detail on Page 5, Line 135-142.
6. In fact, there are many ways to identify multimorbidity patterns. What are the advantages of weighted networks over others and why did you choose to use this method? Please explain the reason.
Response: Thank you for pointing out this question. Aguado et al pointed out that multimorbidity is a complex phenomenon and can be studied with network analysis. In Hidalgo et al.'s paper, node size is proportional to disease prevalence. Link color indicates correlation strength, and Kalgotra et al.'s research, a circle is a disease; an edge represents a comorbidity. Size of a node explains how well it is connected to other nodes. Further, networks offer a more global picture because it includes not only direct connections but also indirect associations which give more accurate information about multimorbidity. However, we consider disease as counts variable, not quantitative data, and we also wanted to describe the association between diseases in a relatively simple and clear form. Therefore, we chose this method, and node size is proportional to disease prevalence and the thickness of each line, which connects the nodes, represents the degree of association between two nodes in this study.
Results
7. The authors point out that "The distribution of the diseases differed significantly by gender, age and hospital duration ($P < 0.001$)", but it is unclear whether "the distribution of the diseases" means each disease or multimorbidity. This should be clear and specific.
Response: Thank you for pointing out this question. We have revised table 1 and used the number of diseases to replace the distribution of the diseases. And the description about Table 1 has also been modified. See detail on Page 5-6 Line 155-164.
8. In Table 2, what are the proper nouns for each disease based on, a dictionary or a thesaurus? They don't seem to be consistent with ICD-10. In addition, do members of the general practitioner group be good at English? By the way, ICD is an abbreviation and should be capitalized when it first appears in the text.
Response: Firstly, an ICD-10 code of a disease has three, four or five digits (xxx.xx). The first three digits represent the broader category of a disease. We chose the first three ICD-10 codes range in the study and we also have added the ICD-10 codes range in the Table 2. Here is the link to the ICD-10 website: https://icd.who.int/browse10/2010/en. Secondly, the English level of the general practitioner group is acceptable, but not relevant to the study. The important thing is that they need to know the

ICD-10 codes for the disease. Thirdly, we have given the full name about ICD-10 when it first appears in the text on Page 4 Line 110-111.

9. The variable of 'total cost' in Figure 1 is not represented in the paper. In addition, the network graph shows the strength of the relationship between variables, but how to judge the relationship between sex and hospital duration is not indicated.

Response: Thank you for pointing out this and we have deleted the variable of "total cost" in Figure 1. Since the connection between sex and hospital duration was similar, we considered the brevity of the Figure and did not give the connection between sex and hospital duration, which was indeed an oversight on our part. We have revised the Figure 1 and added the connection between sex and hospital duration, see Figure1 please.

Discussion

10. In the first paragraph, what's the meaning of "the inclusion of all the potential factors may have signified a greater complexity, but could also better explain the connection with diseases"?

Response: We are very sorry that this sentence has given you a misunderstanding. We want to say that these factors (ex, age and hospital duration) may make the figure more complex, but these factors also better explain the connection with disease. Knowledge of the connection between factors and diseases may help in planning the health services needed in the future. In addition, the healthcare system can further modifications to accurately identify and effectively manage patients with multimorbidity according to the combinations of multimorbidity. Moreover, knowledge of the common combinations of multimorbidity not only improve the quality of patients' life, but also can decrease the heavy economic burden to individuals, families and the society. Finally, we have revised this section, see detail on Page 9 Line 226-229.

11. In the fifth paragraph, the adverbial of comparison in this sentence of "Specogna et al found that patients with spontaneous intracerebral hemorrhage arriving at the hospital with hypertension were 31% more likely to stay in the hospital beyond 1 week per visit than non-hypertensive patients" seems inappropriate. Please explain how it is described in the original text, and what do you mean here?

Response: The description in the original text is that "This study demonstrated that additional morbidities may be associated with longer stay and higher cost of hospital care overall after hemorrhagic stroke; with hypertension being the most frequent and costly multimorbidity. Patients with Spontaneous intracerebral hemorrhage arriving at hospital with hypertension were 31% more likely to stay in hospital beyond 1 week". We want to say similar with Specogna' study, hypertension being the most frequent with other diseases, and the multimorbidity patients with hypertension, the length of hospital duration increases. Therefore, hypertensive diseases (HyD) showed more marked connections with other diseases and a stronger connection with the longer hospital duration group, which is consistent with Specogna' research.

Conclusions

12. There are no specific recommendations in the conclusion. The meaning of "effective interventions" and "different strategies" should be specified.

Response: This study shows the differences in multimorbidity according to gender, age group and hospital duration through visual form. Adjusting the analysis of multimorbidity patterns to the individual level revealed that ischemic heart diseases, hypertensive diseases, cerebrovascular diseases and other forms of heart disease were the central points of disease clusters and directly or indirectly related to other diseases and factors. Therefore, we can only propose effective interventions for these diseases, different intervention strategies for different age groups. However, the study did not research the specific "effective interventions" and "different strategies". Therefore, we think give the specific "effective interventions" and "different strategies" in the conclusion is not appropriate.

13. What does the last sentence do? Is it a suggestion or a fact?

Response: Firstly, the analysis was based on actual inpatient data, so the results are consistent with clinical practice, the sentence "the multimorbidity patterns were more consistent with clinical practice" is a fact in the conclusion. And the sentence" allowing the effective management of patients with multimorbidity" is a suggestion or extension based on fact.

14. The format of the references is not uniform. For example, whether the first letter of each word in

the title should be capitalized; Whether the journal name should be uppercase, and so on. Please check the guideline for careful modification.
Response: Thank you for your careful reminder, we have checked all the references and modified them.
15. In addition, I suggest you correct the English by a native. The authors may benefit from proof-reading to reduce typos and grammatical errors throughout the manuscript.
Response: Thank you very much for your good suggestion and constructive comments. And we have revised English language of our manuscript by Editage (www.editage.cn).

Thank you very much for your positive comments of our manuscript and giving a lot of valuable advice on our manuscript. We appreciate your insightful comments that greatly improved the quality and presentation of the manuscript. We hope the revised version is acceptable to BMJ Open.

VERSION 2 – REVIEW

REVIEWER	Kalgotra, Pankush Auburn University
REVIEW RETURNED	16-Sep-2021

GENERAL COMMENTS	I appreciate the authors for their efforts in revising the paper. It has certainly improved, especially, the English. The abstract needs to be re-written. For instance, the objective says "different factors". You must talk about the specific factors considered in this paper so that the reader gets your point here. In addition, the limitations are not discussed in detail. Do not hesitate to elaborate on all possible limitations at the problem level, method level and the results level. Finally, I still found some typos. For instance, the numbers in the paper do not have commas. For example, "516399 patients" should be "516,399 patients". Similarly, other numbers should be corrected. Good luck with the revision! Thanks!

REVIEWER	Lu, Jiao Shanxi Medical University
REVIEW RETURNED	The authors gave an extensive and polite response to my suggestions, and most of them have been addressed. These changes have already appeared in their revised manuscript. Overall, although I disagree with many of the points made by the authors in my comments, I think it is worth publishing with further revisions. 1. The authors have listed the location and the GDP ranking on which the sampling was based, but how you choose among them remains ambiguous. 2. The reason why the criterion for rare diseases was set at a frequency of less than 1000 is not clearly stated. What were the results of the Kalgotra et al. that led the authors to believe that 1000 was an appropriate criterion. I think what you should write clearly is: what are the advantages of using a disease with 1000 or more patients compared to a disease with fewer than 1000 patients? In other words, what can be achieved with a sample size of 1000? Please use quantitative data to illustrate. Tip: The study of Kalgotra et al. was conducted to find the influence of sample size on the network. 3. The ultimate purpose of a study is to guide practice, and there is nothing inappropriate about giving specific policy implications or intervention strategies in conclusion. 4. There is no reference added on page 3, line 87.

VERSION 2 – AUTHOR RESPONSE

A reply to the comments of reviewer 1 on

“Analysis of multimorbidity networks associated with different factors in Northeast China: A cross-sectional analysis”

Thank you very much for your helpful and constructive comments on our manuscript. We have revised our manuscript following your valuable suggestions. The following is the point-by-point responses to your comments.

I appreciate the authors for their efforts in revising the paper. It has certainly improved, especially, the English. The abstract needs to be re-written. For instance, the objective says "different factors". You must talk about the specific factors considered in this paper so that the reader gets your point here. In addition, the limitations are not discussed in detail. Do not hesitate to elaborate on all possible limitations at the problem level, method level and the results level. Finally, I still found some typos. For instance, the numbers in the paper do not have commas. For example, "516399 patients" should be "516,399 patients". Similarly, other numbers should be corrected. Good luck with the revision! Thanks!

Response: Thank you for offering pretty advice. Firstly, we have revised the abstract, see details on Page 2 Line 42-44. Secondly, we also have revised the limitations, see details on Page 11 Line 276-280. Thirdly, we have referred to the publihed articles of the journal BMJ Open and found that the numbers in the journal's paper do not have commas. Thank you again for your help.

We appreciate your insightful comments that greatly improved the quality and presentation of the manuscript. We hope the revised version is acceptable to BMJ Open. Thank you again for your helpful and constructive comments.

A reply to the comments of reviewer 2 on

“Analysis of multimorbidity networks associated with different factors in Northeast China: A cross-sectional analysis”

Thank you very much for your helpful and constructive comments on our manuscript. We have revised our manuscript following your valuable suggestions. The following is the point-by-point responses to your comments.

The authors gave an extensive and polite response to my suggestions, and most of them have been addressed. These changes have already appeared in their revised manuscript. Overall, although I disagree with many of the points made by the authors in my comments, I think it is worth publishing with further revisions.

1. The authors have listed the location and the GDP ranking on which the sampling was based, but how you choose among them remains ambiguous.

Response: Thank you for your careful reminder, three general hospitals from each of these locations (Jilin, Changchun, Baicheng, Yanbian, and Tonghua) were then random selected. Finally, based on the administrative division, 15 general hospitals were selected as the monitoring institutions for this study. Please see detail on Page 4 Line 124-125.

2. The reason why the criterion for rare diseases was set at a frequency of less than 1000 is not clearly stated. What were the results of the Kalgotra et al. that led the authors to believe that 1000 was an appropriate criterion. I think what you should write clearly is: what are the advantages of using a disease with 1000 or more patients compared to a disease with fewer than 1000 patients? In other words, what can be achieved with a sample size of 1000? Please use quantitative data to illustrate. Tip: The study of Kalgotra et al. was conducted to find the influence of sample size on the network.

Response: Thank you for pointing out this. Relevant studies have not defined the exclusion criteria for diseases with less incidence at present. At the beginning of the study, we included these diseases with less incidence into the model for analysis, and found that these diseases appear to be relatively isolated in the entire model due to its small number of patients and little association with other diseases. Besides, we refer to the study of Kalgotra et al. and chose to exclude diseases with fewer than 1,000 patients in the analysis, and we found that the exclusion of diseases with fewer than 1000 do not affect the overall results of the study. However, this may be a limitation of this study, we have added it, please see detail on Page 11 Line 279-280.

3. The ultimate purpose of a study is to guide practice, and there is nothing inappropriate about giving specific policy implications or intervention strategies in conclusion.

Response: Thank you very much for your good suggestion and constructive comments to this paper. And we have given specific policy implications or intervention strategies in conclusion. Please see detail on Page 11 Line 286-289.

4. There is no reference added on page 3, line 87.

Response: We are so sorry for missing it. Thank you for pointing out this and we have added it. Please see detail on Page 3 Line 88.

We appreciate your insightful comments that greatly improved the quality and presentation of the manuscript. We hope the revised version is acceptable to BMJ Open. Thank you again for your helpful and constructive comments.

VERSION 3 – REVIEW

REVIEWER	Kalgotra, Pankush Auburn University
REVIEW RETURNED	05-Oct-2021

GENERAL COMMENTS	Good Luck!

REVIEWER	Lu, Jiao Shanxi Medical University
REVIEW RETURNED	16-Oct-2021

GENERAL COMMENTS	I was very glad to re-view the paper in greater depth. The submission has been greatly improved and is worthy of publication.