

## 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

<b>TITLE (PROVISIONAL)</b>	Medical Expenditure Clustering and Determinants of the Annual Medical Expenditures of Residents: A Population-Based Retrospective Study from Rural China
<b>AUTHORS</b>	Zhang, Yan; Lu, Shan; Yadong, Niu; ZHANG, LIANG

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Mingsheng Chen Nanjing Medical University, China
<b>REVIEW RETURNED</b>	09-Mar-2018

<b>GENERAL COMMENTS</b>	<p>Chinese health sector traditionally confronted with the difficulty of high medical cost. Examination on the medical expenditure and its determinants will provide useful evidence and policy implication on medical cost calculation and control. This study aims to identify the characteristics of the patients with high cost and seek to explore the determinants of the annual medical expenditure for rural residents. This is a significant contribution to the global literature on the study on medical expenditure. The methodology is sound and consistent with the literature, allowing the findings to be readily benchmarked against findings worldwide. Below are some comments/questions and suggestions for improvement. Please feel free to treat these as you see fit, taking into account what further analysis can be conducted using your data:</p> <ol style="list-style-type: none"><li>1. At the beginning of background section, it might be proper to introduce some crucial background information on China's medical expenditure. When introducing medical expenditure clustering in the initial part, the readers may feel confused with the main purpose and main problem that this manuscript hopes to cope with.</li><li>2. Statistical analysis was not well understood. Please introduce the procedure of statistical analysis in detail.</li><li>3. Personally, I do not recommend that the Discussion section is divided by different parts because these parts may interconnect with each other. You could add some proper sentences and make the Discussion section as an integral one.</li></ol>
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Hans V Hogerzeil Department of Health Sciences, Global Health, University Medical Centre Groningen, The Netherlands
<b>REVIEW RETURNED</b>	12-Mar-2018

<b>GENERAL COMMENTS</b>	Good article, much work went into this. Good conclusions, relevant
-------------------------	--------------------------------------------------------------------

	<p>I am not sure about the statistics; I get the feeling not all statistics are really necessary; the formula is a bit too complex for me.</p> <p>Interesting observation that 5% of patients consume over 60% of all resources - that is really very important news. Efforts to reduce costs and promote cost-efficiency should focus on these patients. I think they cannot be identified before or on the basis of risk factors; they should only be identified by their expenditure / health claims</p> <p>Some comments:</p> <p>Line 261: I do not understand the importance of the Gini coefficient in this context.</p> <p>268: I do not understand the 7,35% in this context</p> <p>277: SD much higher than the mean: not clear to me</p> <p>280: Some indication of dollar value is interesting for the reader</p> <p>314: So only 8,5% of the HCgroups are people above 60 years? That is against my intuition. If true, that is interesting; it would show that the majority of HC patients are young adults. Please check</p>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### VERSION 1 – AUTHOR RESPONSE

Our responses to your comments are below:

1. At the beginning of background section, it might be proper to introduce some crucial background information on China's medical expenditure. When introducing medical expenditure clustering in the initial part, the readers may feel confused with the main purpose and main problem that this manuscript hopes to cope with.

Response: Thanks for your excellent concern, we agree it's necessary. We've added a paragraph to introduce the China's medical expenditure, make clear the relationship between medical expenditure and medical expenditure clustering (Line 69-76).

Line 69-76: The rapid increase in health expenditures greatly impedes the development of the New Rural Cooperative Medical System (NRCMS), the largest basic social health insurance system in rural China that covers 603.46 million rural residents. Specifically, the health expenditures per capita in China have increased from 513.8¥ (83.6\$) in 2012 to 1279.2¥ (208.2\$) in 2017 with an annual growth rate of 25.6%, which is much higher than the annual growth in fundraising per capita (16.02%).

Medical expenditure clustering is considered an important factor that motivates such rapid increase in health expenditures.

2. Statistical analysis was not well understood. Please introduce the procedure of statistical analysis in detail.

Response: It's a question. We re-examined the level of research methods and further provided details of statistical methods for the reader's understanding (Line 194-210).

3. Personally, I do not recommend that the Discussion section is divided by different parts because these parts may interconnect with each other. You could add some proper sentences and make the Discussion section as an integral one.

Response: It's a very good suggestion. We rebuild the discussion, and have added some proper sentences as your recommend.

Dear professor Hans V Hogerzeil,

Our responses to your comments are below:

1. I am not sure about the statistics; I get the feeling not all statistics are really necessary; the formula is a bit too complex for me.

Response: It's a question. We re-examined the level of research methods and further provided details of statistical methods for the reader's understanding (Line 194-210).

2. I think they cannot be identified before or on the basis of risk factors; they should only be identified by their expenditure / health claims

Response: It's a question. Although HC patients are identified based on their medical expenditure or health claims, a patient can be predicted as HC high probability in advance based on several risk factors (Line 376-383).

Line 376-383: Although HC patients are identified based on their medical expenditure or health claims, a patient can be predicted as HC high probability in advance based on several risk factors. Robst et al. and Wodchis et al. found that a patient identified as HC in a year is more than 40% likely to be identified as an HC patient in the following year given that these patients often maintain a high level of medical expenditure for the following year. Therefore, those residents with a remarkably high healthcare utilisation, are exposed to many risk factors and have been identified as HC patients in the previous year warrant special attention.

3. Line 261: I do not understand the importance of the Gini coefficient in this context.

Response: Yes, we recognized it. The Gini coefficient is a digitised representation of medical expenditure clustering. A larger Gini coefficient corresponds to a higher degree of medical expenditure clustering. We are not only focusing on high-cost groups, we also have to pay attention to the medical expenditure clustering of the entire population (Line 194-197).

Line 194-197: Firstly, the medical expenditures of the residents were clustered by using the Gini coefficient and Lorentz curve. The Gini coefficient is a digitised representation of medical expenditure clustering. A larger Gini coefficient corresponds to a higher degree of medical expenditure clustering.

4. 268: I do not understand the 7,35% in this context;

Response: It's our negligence, here we'd like to express that the annual medical expenditure per capita of the entire population was only 7.35% of the annual medical expenditure per capita of the HC group (Line 295-298).

Line 295-298: The annual medical expenditure per capita of the entire population was 1,222.49¥ (199.01\$), which was nearly similar to that of the MC group (1261.36¥, 205.18\$). However, this value was only 7.35% of the annual medical expenditure per capita of the HC group.

5. 277: SD much higher than the mean: not clear to me

Response: Yes, it's a good suggestion. This is an indicator on the statistical description of data distribution, the larger the standard deviation, the more dispersed the data.

6. 280: Some indication of dollar value is interesting for the reader

Response: Good point – we agree. We have added the dollar value for all expenditure.

7. 314: So only 8,5% of the HCgroups are people above 60 years? That is against my intuition. If true, that is interesting; it would show that the majority of HC patients are young adults. Please check

Response: That's not what we meant, here we want to express that residents aged above 60 years account for 34.48% of HC group. We have adjusted my expression (Line344-346).

Line344-346: This case is particularly true for those residents aged above 60 years, who account for 34.48% of HC group, in other words, 8.5% of the elderly population was defined as HC, while only 5% of total population was defined as HC.

Editorial Requests:

1. Can the methods section of the abstract be more informative? For example, it does not include the study's sample size.

Response: Yes, it's a good suggestion. We added the study design and study's sample size in the participants section (Line 37-38).

Line 37-38: A total of 478,051 residents who availed healthcare services were recruited for the retrospective study in 2014.

2. Can you please clarify the study design in the abstract and title? "Empirical study" is vague. We note that it is later described as a retrospective cohort study.

Response: Yes, it's a good suggestion. We adjusted the title as below, a population-based retrospective study in China.

3. Please work on improving the quality of English. We recommend consulting a native English speaker or professional copy-editing service, if possible.

Response: Thank you for this, as you recommended, we have got a professional editing service from a professional editing company to improve the readability of the manuscript. We thank Audrey Holmes, MA, from Liwen Bianji, Edanz Group China ([www.liwenbianji.cn/ac](http://www.liwenbianji.cn/ac)), for editing the English text of a draft of this manuscript.

4. Please explain in the methods section of the manuscript why your study did not require approval from a local ethics committee.

Response: OK, as you recommended, we added the ethical approval section, and provided the Chinese Clinical Trial Registry (ChiCTR-OOR-14005563) (Line 214-217).

5. Along with your revised manuscript, please provide a completed copy of the STROBE checklist (<http://www.strobe-statement.org/>).

Response: OK, we have adjusted the STROBE checklist.

6. Kindly re-upload each figure in either TIFF or JPG format with at least 300 dpi resolution.

Response: OK, we have remade the figure, it has a resolution of 800\*800 dpi as your requirement.

7. Authors must include a statement in the methods section of the manuscript under the sub-heading 'Patient and Public Involvement'. If patients and or public were not involved please state this.

Response: Yes, we agree it's helpful. We advanced the Patient and Public Involvement in the methods section(Line 211-212).

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Mingsheng Chen Nanjing Medical University, China
<b>REVIEW RETURNED</b>	11-Apr-2018
<b>GENERAL COMMENTS</b>	I am very satisfied with the authors' comments and the revision of this manuscript.