## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Patients' preferences for health insurance coverage of new technologies for treating chronic diseases in China: a discrete choice experiment |
| **AUTHORS** | Geng, JinSong; Chen, Xiaowei; Bao, Haini; Qian, Danmin; Shao, Yuting; Yu, Hao |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Alison Pearce<br>University of Sydney, Australia |
| **REVIEW RETURNED** | 08-Apr-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for the opportunity to review this manuscript detailing a study of patient preferences for health insurance coverage of new technologies.<br><br>The paper describes a relatively straight-forward discrete choice experiment study. While the methods appear appropriate, some more detail would assist the reader in accurately interpreting the results and their implications for policy. I have made some suggestions which may strengthen the paper for future publication.<br><br>Introduction:<br>1. It would be useful for the reader to have a definition of what you considered a 'new medical technology' in the context of your study.<br><br>Methods:<br>3. It is not clear to the reader why hypertension and diabetes were selected as case studies? This hasn't been mentioned in the introduction and is not really justified in the methods. It would be good to provide some justification for why you wanted the patient experience rather than the general population, why those two patient populations were chosen, and why the scenarios provided were not specific to diabetes and hypertension.<br>4. I think you have used the systematic review to identify your attributes, and then used the focus groups about diabetes and hypertension to choose the levels for your attributes, but this is not entirely clear and I was unable to source the publication of your systematic review that you refer to (reference number 20, Chen et al 2018)<br>5. The use of an unlabelled design is not really part of selection of the attributes and levels – this paragraph (page 9, line 32) may be more appropriate under the section titled 'Experimental design and development of the questionnaire'.<br>6. I found the description and definition of the two attributes 'expected gains in health outcomes from the treatment (not as |

| | |
|---|---|
| | expected/as expected)' and 'likelihood of effective treatment (30-90%)' difficult to interpret. Some more detail about what these two attributes mean, how they differ and how they interact would be useful for the reader.<br>7. It is not clear to me how the attribute about availability of alternatives would have been interpreted by participants. It appears there was no mention of the efficacy or side effects etc of these treatments, so perhaps it makes sense that the impact of this attribute was negligible on participants preferences. Please clarify if any additional information was given about these alternatives.<br>8. Please provide more information about how the out of pocket costs attribute was described to participants. The current wording ('Out of pocket costs for new technology (if not reimbursed') could be interpreted two ways: a) that these are the costs patients pay, even if the treatment is reimbursed by social insurance (in which case the wording of 'if not reimbursed' should be changed), or b) that the cost would be $0 for the selected treatment because it would now be reimbursed, so only the non-selected (and therefore not reimbursed) treatment would incur an out of pocket costs. This is important because it has implications for your interpretation of the results. If Option B is correct, then you would expect to see a cost as a positive predictor rather than a negative predictor of choice.<br>9. Did you do any assessment of participant understanding of the DCE attributes, levels or choice task? (e.g. duplicate/dominated choice set, debriefing questions for participants, debriefing questions for interviewers?) In particular, risk and cost attributes are known to sometimes be difficult for participants to understand in a healthcare context.<br>10. Given you see heterogeneity of preferences when you analyse the results by subgroup, it would be useful to justify why you didn't do a specific analysis of this, such as a mixed logit or possibly a latent class analysis.<br>Results:<br>11. You haven't mentioned the results for the attribute 'severity of target disease' or how you interpret them.<br>12. You mention (and present in Appendix 5) the results by inpatient status subgroup, but you have not mentioned this in your methods section.<br>Discussion:<br>13. Depending on your response to Question 7 above, you may wish to reconsider your interpretation of the results around other available treatments<br>14. Depending on your response to Question 8 above, you may need to revise your interpretation of the results around the cost attribute<br>Tables & Figures:<br>15. Table 2: Classification of employment seems unusual. Is 'farmer' really the only employment option for rural residents?<br>16. Table 2: 80% of people being an inpatient seems quite high – what does this relate to?<br>17. Appendix 3, supplemental table 2: would be good to also split out the number of patients with each disease who had UEBMI without extra benefits<br>18. Appendix 4, supplemental figure 1: could patients only have one other type of comorbidity (ie no multimorbidity)? Also, I would suggest selecting a different chart type (e.g. bar graph) rather than a 3D pie chart to display this information.<br>Overall: |

| | |
|---|---|
| | 19. Needs a proofread for spelling mistakes, e.g.<br>a. Methods, line 22 – DEC should be DCE<br>b. Acknowledgements – 'out interviewers' should be 'our interviewers' |

| | |
|---|---|
| **REVIEWER** | Amarech G. Obse<br>University of Cape Town |
| **REVIEW RETURNED** | 14-May-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Title: Patients' preferences for health insurance coverage of new technologies for treating chronic diseases in China: a discrete choice experiment<br><br>Review comments<br><br>Many thanks for the opportunity to review this paper it is an interesting study. I hope the comments and suggestions are useful.<br><br>Kind regards,<br>Amarech<br><br>Comments<br><br>Background:<br>You have mentioned that there are differences in benefits between URRBMI and UEBMI members (and referred the reader to another paper by Yu for details). However, you may need to highlight the differences in health insurance benefits for members with hypertension and diabetes. This is important information since the paper provided analysis of the differences in valuation of attributes of new technologies for these diseases. Similarly, you briefly mentioned that "[e]ven among the UEBMI enrollees, insurance benefit is not equal". Again, you need to expand this a little further so that the sub-group analysis you provided will be well understood.<br><br>Page 8, hypotheses - the first hypothesis is not clear. It is stated as "new technology attributes regarding health benefits are most preferred by chronic disease patients for health insurance coverage". Do you mean all attributes of any new technology are preferred? In comparison to what? When I read this, I thought you will have attributes/levels of the status quo or opt-out (or status quo) option to support (or reject) this hypothesis, but I see you have forced choice sets. So, how can you provide evidence in support or against this hypothesis?<br><br>Attribute selection:<br>You mentioned the methods followed to select attributes and levels, but it is not clearly indicated how the final attributes were selected. Only the criteria identified from systematic review is listed. The findings from the other methods is not clear.<br><br>During the FGDs for attribute selection, you stated that there was no consensus among experts about attribute levels. Is this only about the attribute levels or both attribute and their levels? Can you provide an example of an attribute/level where there was disagreement? How does the database search help you in rectifying the differences? I understand the database search still |

| | |
|---|---|
| | results in long list of attributes/levels, how was the final decision made? Please clarify.<br><br>On Page 9 lines 21-26, you stated that "[…] we defined new medical technologies in this study as the therapeutics for hypertension and diabetes, which had been marketed in China but were not covered by the public health insurance programs in Jiangsu in 2018". If the medical technologies used in the DCE are already in market, how did you use the information from the literature reviews and the FGDs (conducted to identify attributes and levels)? Or why was there a need to go through finding attribute and levels when the medical technologies used in the final DCE already existed?<br><br>I understand the attributes for a DCE that aims to assess preferences for new technologies should be attributes describing the new technologies. I do not understand how the attribute "alternative technologies currently covered by insurance" is considered as an attribute for a new technology. Please clarify.<br><br>The attribute "severity of target disease" may be correlated with respondents and may bias valuation of attributes. For instance, an individual with severe disease my choose an alternative with a level 'severe' for this attribute without considering the other alternatives. Please clarify how you dealt with such correlation.<br><br>Please clarify why forced choice set is regarded to be suitable for this study given the limitation that it results in overstatement of the valuation of attributes.<br><br>Sample size:<br>You indicated that you have sufficient sample size (408) for overall and sub-group analyses based on the recommendation of a sample size of 150 suggested as sufficient from the literature. This would have been correct if you did not block your design. The sample size requirement for blocked design increases exponentially. You have indicated that your experimental design is partitioned into five blocks; which means you need five people to complete one design. So, your effective sample size is about 82 (408/5 = 81.6). The number of observations in Table 3, which is 4896, also indicates this. Please clarify.<br><br>Please clarify how the pilot study informed the final survey. Was there any change in the design or attribute or any other aspect of the study based on the pilot?<br><br>Data analysis and results:<br>You stated that "[…] analysis of the DCE data was based on the random utility model and assumed that the utility of two scenarios was equal". I wonder how the assumption of equal utility works. If individuals get the same utility from two alternatives; that means, they are indifferent between the two alternatives. And we may not need to ask them to make choice in the first place. This is in contradiction with the theory underpinning a DCE. Please clarify.<br><br>Conditional logit model has an assumption of the independence of irrelevant alternatives which is violated by a DCE data. With development in statistical software, more flexible models that accommodate limitations of logit model has been suggested. Please explain why conditional logit is estimated given the |

| | |
|---|---|
| | restrictive assumption and why other flexible models were not considered.<br><br>In relation to the comment above, flexible models are also used to assess unobserved heterogeneity in valuation of attributes among the respondents. Please clarify why this was not considered.<br><br>For conditional logit, alternative specific constant should be reported in the result tables.<br><br>Please clarify why dummy coding is used for categorical variables in comparison to other schemes of coding; particularly, given that the marginal utility of the reference level of the dummy coded variable is confounded with alternative specific constant.<br><br>Sub-group analysis is performed by type of disease and type of insurance. I wonder why other socio-demographic variables are not considered. In particular, gender differences in valuation of health care services is widely reported.<br><br>Page 13, line 12 … Please clarify that you have two groups in UEBMI, you mentioned that you have further analysis by UEBMI (please see the comment on background)<br><br>The sub-group analysis by type of disease is narrated as respondents with hypotension only versus diabetes only. But the analysis was done on respondents with hypertension only versus respondent with both hypertension and diabetes, as indicated in Table 3. This needs to be corrected. Why does the respondents with both diabetes and hypertension added to the groups with diabetes only and not with patients with hypertension only? Please clarify if there is any clinical or insurance related reason for this.<br><br>Discussion:<br>This section is thin, needs to be strengthened<br><br>There are suggestions of new technologies that needs to be considered for the introduction of new technologies. But, how is this compared with technologies currently in use. I think this is important information for policy.<br><br>Please clarify if the statement on page 19 line 46-48 … "However, variation in patient preferences existed and mainly depended on patients' own experiences of the disease." What do you mean by "patients' own experiences"? Is this statement based on your finding or another research?<br><br>Page 20 - You discussed about inequalities in health insurance coverage affecting preferences which is fine. But I do not understand the discussion further provided about inequalities and not sure if it directly relates to your findings. I think, the discussion is a bit stretched.<br><br>P21 line 30. Here you stated that the attribute 'alternative technologies currently covered by insurance' is not valued because URRBMI patients have lower education. Firstly, the education of the respondents by insurance type is not reported in the result section. Secondly, how can you relate less valuation of this attribute to a single sociodemographic characteristic of respondents? |

**VERSION 1 – AUTHOR RESPONSE**

**Reviewer: 1**

**Reviewer Name: Alison Pearce**

**Institution and Country: University of Sydney, Australia**

Please state any competing interests or state 'None declared': None declared

Thank you for the opportunity to review this manuscript detailing a study of patient preferences for health insurance coverage of new technologies. The paper describes a relatively straight-forward discrete choice experiment study. While the methods appear appropriate, some more detail would assist the reader in accurately interpreting the results and their implications for policy. I have made some suggestions which may strengthen the paper for future publication.

**1. Introduction: It would be useful for the reader to have a definition of what you considered a 'new medical technology' in the context of your study.**

Reply: Dear Alison, we are so grateful to you for your suggestions. Your comments are very helpful and valuable for us to improve the quality of the manuscript. We have revised the manuscript carefully according to your suggestions.

We added the definition of 'new medical technology' after Table 1 in the section of 'Methods: Identification of technology attributes and levels'. The definition is as follows:

*'New medical technologies which were mentioned in our research were already in use in clinical practice. However, they had not yet been included in the catalogs of social health insurance including URRBMI and UEBMI.'*

**2. Methods: It is not clear to the reader why hypertension and diabetes were selected as case studies? This hasn't been mentioned in the introduction and is not really justified in the methods.**

Reply:

(1) We mentioned hypertension and diabetes in the 'Introduction' section to show the importance of these diseases:

*'It was estimated that among Chinese adults aged 35 to 75 years, nearly half had hypertension. The overall prevalence of diabetes in Chinese adults was about 10.9%. Furthermore, comorbidities are highly prevalent among patients with chronic diseases, which have a negative impact on the patient's quality of life and impede the efficacy of treatment.'*

(2) The objective of our study was to inform insurance decision-making in China by investigating patients' preferences for insurance coverage of new technologies for treating chronic diseases. We took diabetes and hypertension as an example in our DCE mainly due to the high prevalence, long-term complications, and heavy economic and social burden of these diseases. To make the research purpose clearer, we added this sentence in the first paragraph of 'Methods: DCE implementation and data collection':

*'Due to the high prevalence, serious complications and heavy burden of hypertension and diabetes, we selected patients with these diseases as participants.'*

**3. It would be good to provide some justification for why you wanted the patient experience rather than the general population, why those two patient populations were chosen, and why the scenarios provided were not specific to diabetes and hypertension.**

Reply:

(1) We tried to identify attributes which were most relevant to the medical reimbursement decision-making in chronic disease, especially for diabetes and hypertension.

Department of Medical Informatics in Medical School of Nantong University is also the Center for Evidence-based Medicine in Nantong University. We have activities for evidence-based medicine (website: http://ebm.ntu.edu.cn), we also have the MOOC of Evidence-based Medicine (website: http://www.icourse163.org/course/NTU-1207007806). Therefore, we maintain good relationship with physicians and have workshops and training activities regarding evidence-based medicine and evidence-based clinical practice. During the pilot survey in one of our evidence-based clinical practice workshop, physicians found it difficult to give accurate definitions to levels of attributes. For example, if they wanted to classify the outcome measures regarding effectiveness, there were several kinds of

outcomes measures and nobody knew how to integrate them into a single composite measure. Even evidence-based guidelines never told them how to do that.

(2) We did focus group discussion and expert consultation with attributes and levels regarding reimbursement of new medical technologies. 14 experts (from 6 provinces in China) in reimbursement, health economics, healthcare service and evidence-based medicine were consulted. Results were published in another separate paper (Geng JS, Chen XW, Yu XL, *et al*. Study on the evidence-based decision-making framework for reimbursement of new technologies in view of EVIDEM. *Chinese Journal of Health Policy* 2018;11(4):50-54).

(3) We also searched the famous health technology assessment database established by Canadian Agency for Drugs and Technologies in Health (CADTH) to select potential new technologies and find reasons for recommendation of reimbursement to better define the levels of attributes. We found 68 reports regarding hypertension and diabetes which were published before March 2018. We analyzed the health technology assessment report to further determine the level of attributes. For example, serious adverse events were reported in about 11% of patients in the Riociguat 2.5 mg groups. Although serious adverse events were rare in drugs like Insulin Glulisine and Exenatide, rarely had treatment techniques for chronic disease were reported to be 0.0% serious adverse events for all patients.

(4) We analyzed the price of new technologies in China in year 2018. For example, the out-of-pocket costs for Empagliflozin was near 300 RMB per month, if not reimbursed. The out-of-pocket costs for Riociguat was almost 6000 RMB per month. The out-of-pocket costs for other technologies were between 300 RMB and 3500 RMB per month, with the median near 1500 RMB. As we know, many patients had more than one type of chronic disease. They may have diabetes, hypertension and other types of comorbidities. Therefore, the actual out-of-pocket costs per month might be even higher. China is implementing a centralized procurement method and the procurement decision was made at provincial level. China also pilots the centralized procurement mechanism, which is mainly to improve the volume procurement method in exchange for more favorable prices. In terms of these mechanisms, we finally determined the definition regarding out-of-pocket costs.

(5) As we have mentioned, some hypertension and diabetes patients have comorbidities, while patients with other chronic diseases may have hypertension and diabetes. Therefore, the selection of patients with hypertension and diabetes is meaningful.

**4. I think you have used the systematic review to identify your attributes, and then used the focus groups about diabetes and hypertension to choose the levels for your attributes, but this is not entirely clear and I was unable to source the publication of your systematic review that you refer to (reference number 20, Chen et al 2018)**

Reply:

(1) The main results of systematic review were mentioned in new reference number 28, Chen et al 2018 (Chen XW, Chen Q, Geng JS, *et al*. The application of multi-criteria decision analysis in medical insurance reimbursement: a systematic review. *Chinese Health Resources* 2018;21:218-223). The abstract of this published paper was as follows:

*To investigate the application of multi-criteria decision analysis (MCDA) in medical insurance reimbursement through a systematic review. Search terms were developed, and several Chinese and English databases were searched. The inclusion criteria were formulated and the literature was screened; quality assessment methods were established on the basis of the items of appraising qualitative research which were developed by Dixon-Woods M. Twenty-two literatures were included. The average quality score of included literatures was 9.8 points. MCDA was designed to support medicine reimbursement decision-making/establishment of medicine reimbursement catalogs, health technology assessment of various medical technologies, and disease screening. The most commonly mentioned decision-making dimensions were comparative outcomes, economic consequences and the needs of technology.*

(2) The main findings of focus group and expert consultation were summarized in another paper (Reference number 29: Geng JS, Chen XW, Yu XL, *et al*. Study on the evidence-based decision-making framework for reimbursement of new technologies in view of EVIDEM. *Chinese Journal of Health Policy* 2018;11(4):50-54). The abstract of this published paper was as follows:

*This study aimed to establish the evidence-based decision-making framework for reimbursement of new technologies on the basis of EVIDEM. Literature review, focus group discussion and qualitative interview were used to construct the preliminary decision-making framework, and expert consultation was adopted to determine the necessity and weight of the criteria. The evidence-based decision-making framework consisted of normative universal aspect and contextual aspect. The normative aspect included following criteria, needs of the technology (severity, benefit type of technology, unmet needs of reimbursed technology), comparative outcomes (effectiveness, safety/tolerability, patient-perceived/patient-reported outcomes), and economic aspects of the technology (costs and cost-*

*effectiveness). The contextual aspect considered the mission and mandate of medical insurance, population priorities and the accessibility, common goal and specific interests, political context, and affordability of medical insurance.*

(2) Results from our two studies as mentioned above helped us to determine the attributes and levels in our DCE research, since DCE was the most common method within the MCDA theory.

(3) Too many attributes in DCE will be difficult for respondents to make a choice. Best-practice guidelines for stated-preference methods suggest that there is a limit to the number of attributes respondents are able to reliably evaluate. After the literature review about DCE, we decided to select six attributes which were most relevant to reimbursement decision-making.

(4) To make it more clear, we revised the first and second paragraph in the section 'Identification of technology attributes and levels'.

**5. The use of an unlabeled design is not really part of selection of the attributes and levels – this paragraph (page 9, line 32) may be more appropriate under the section titled 'Experimental design and development of the questionnaire'.**

Reply:

We are most grateful to you for your suggestions. It is surely more appropriate to move the use of an unlabeled design to another part. Therefore, we moved the description of unlabeled design from the part of 'Methods: Identification of technology attributes and levels' to the second paragraph of 'Methods: Experimental design and development of the questionnaire'.

**6. I found the description and definition of the two attributes 'expected gains in health outcomes from the treatment (not as expected/as expected)' and 'likelihood of effective treatment (30-90%)' difficult to interpret. Some more detail about what these two attributes mean, how they differ and how they interact would be useful for the reader.**

Reply:

(1) The principles for multi-criteria decision-analysis (MCDA) include essential, non-overlapping, concise, operational, and preference-independent. DCE is an important method for eliciting weights in the MCDA approach. We classify the effectiveness into two categories to avoid overlap. An expected gain in health outcomes from the treatment was the qualitative aspect, while likelihood of effective treatment was the quantitative aspect.

(2) We give training to interviewers to let them know how to give explanations to patients. During the DCE survey, we asked interviewers to give examples to patients in terms of their medical knowledge. We also keep in touch with interviewers to ensure the validity of the survey.

(3) We provided the explanation for attributes and levels in Appendix 1. Such information had also been provided to interviewers.

**7. It is not clear to me how the attribute about availability of alternatives would have been interpreted by participants. It appears there was no mention of the efficacy or side effects etc of these treatments, so perhaps it makes sense that the impact of this attribute was negligible on participants preferences. Please clarify if any additional information was given about these alternatives.**

Reply:

(1) During the expert consultation, unmet needs of reimbursed technology was determined as a necessary decision-making criteria for reimbursement of new medical technologies. 'Alternative technologies currently covered by insurance' was used as an attribute in terms of 'unmet needs of reimbursed technology'. For chronic diseases, even new technologies developed, traditional technologies which were already reimbursed can also be used in clinical practice. We wanted to include this attribute to reflect the real-world decision-making requirement. Since this is a new and exploratory research, we did not know the attribute was non-significant until we did the final data analysis.

(2) Thanks so much for your suggestions. During the DCE survey, we explained this attribute to patients as alternative technologies already reimbursed had similar effectiveness and safety to the new technology (Level: Yes); no alternatives already reimbursed had similar effectiveness and safety to the new technology (Level: No).

**8. Please provide more information about how the out of pocket costs attribute was described to participants. The current wording ('Out of pocket costs for new technology (if not reimbursed') could be interpreted two ways: a) that these are the costs patients pay, even if the treatment is reimbursed by social insurance (in which case the wording of 'if not**

**reimbursed' should be changed), or b) that the cost would be $0 for the selected treatment because it would now be reimbursed, so only the non-selected (and therefore not reimbursed) treatment would incur an out of pocket costs. This is important because it has implications for your interpretation of the results. If Option B is correct, then you would expect to see a cost as a positive predictor rather than a negative predictor of choice.**

Reply:

(1) During our DCE survey, we listed all of the attributes and asked patients to select the optimal technology which they preferred to be reimbursed in health insurance. Therefore, there was a hypothesis that before they made the decision, the medical technology was not reimbursed. Because if the technology was already reimbursed, then it would be meaningless to select which one should be reimbursed. As a result, the cost of the technology was the out-of-pocket costs, and the hypothesis was the technology was not reimbursed before they choose.

(2) We are grateful to you for your kind suggestions. However, option A does not fully consistent with our study hypothesis. Our definition does not fully comply with option B, because the health insurance reimbursement ratio of technologies is not 100%.

(3) We realize that we should make out-of-pocket costs more explicit in the paper. Therefore, we added the definition of out-of-pocket costs in Appendix 1, and added description of out-of-pocket costs in the second paragraph of 'Discussion: Summary of the findings'.

**9. Did you do any assessment of participant understanding of the DCE attributes, levels or choice task? (e.g. duplicate/dominated choice set, debriefing questions for participants, debriefing questions for interviewers?) In particular, risk and cost attributes are known to sometimes be difficult for participants to understand in a healthcare context.**

Reply:

(1) Thanks to your valuable suggestions. We are sorry for not reporting all of the necessary details in our survey. We added evaluation of confidence in the third paragraph of 'Methods: DCE implementation and data collection', and showed the figure of evaluation in Appendix 3.

(2) We had the confidence evaluation question in the DCE survey to ask patients show how confidence they were in their choice, and asked them to choose the score (from 0 to 10). We excluded data from patients who choose the score for less than 8.

(3) We did the one-to-one, face-to-face interviews with the patients to ensure validity and quality of the investigation. Our interviewers consisted of 13 medical students, all of whom were in hospitals doing their internships during the research period. For quality assurance, the interviewers were trained before the experiment. We compiled a survey training manual, provided detailed descriptions of how to interpret each scenario, and asked each interviewer to make explanations to patients. If a patient could not understand, we required the interviewers to explain with patience. The interviewers were the interns who were medical students and familiar with these patients.

(4) We agree with your idea that attributes are sometimes difficult for some patients to understand in a healthcare context. However, since patients care about the out-of-pocket costs, it was not too difficult to be understood. In case of patients could not understand serious adverse events, we told them that serious adverse events meant life-threatening adverse events caused by the new technology, like severe hypoglycemia, severe hypersensitivity reaction, kidney or liver damage, etc. All of our efforts were to ensure the better understanding of the DCE attributes and levels for patients.

**10. Given you see heterogeneity of preferences when you analyse the results by subgroup, it would be useful to justify why you didn't do a specific analysis of this, such as a mixed logit or possibly a latent class analysis.**

Reply:

We are very grateful to you for your suggestions. The mixed logit model allows for attribute coefficients to be randomly distributed, and improves the fit of the model over conditional logit model. Mixed logit model allows for unknown heterogeneity in individual preferences and estimates both the mean preference weight and the standard deviation. Therefore, we changed the analysis method to mixed logit according to your suggestions. We assumed that all variables of the attributes, except for the constant, had a random component and that the weights of preference were normally distributed. The description of mixed logit model was stated in the second paragraph of 'Methods: Data analysis'.

**11. Results: You haven't mentioned the results for the attribute 'severity of target disease' or how you interpret them.**

Reply:

(1) We were also not sure about the potential reasons why severity of target disease was preferred only by URRBMI enrollees. Perhaps it could be associated with the disease severity of patients themselves. However, it is difficult to define the severity of patients.

(2) This reminds us that we also measured EQ-5D-5L of each patient in the DCE survey. We did not realize the importance of this data on patients' preferences before. We did subgroup analysis and found that patients with lower EQ-5D-5L index value tend to prefer the attribute 'severity of target disease'. We also found that the proportion of lower EQ-5D-5L index value was higher in the URRBMI enrollees as compared with UEBMI enrollees.

(3) We added 'Subgroup analysis by HRQoL' in the results section and listed the results of our new findings.

**12. You mention (and present in Appendix 5) the results by inpatient status subgroup, but you have not mentioned this in your methods section.**

Reply:

Thank you so much for your comments. We do not think the results of inpatients have important meanings. We just wanted to show that DCE results from inpatients were similar to all of the patients included. Therefore, we deleted this misunderstanding result from the Appendix.

**13. Discussion: Depending on your response to Question 7 above, you may wish to reconsider your interpretation of the results around other available treatments**

Reply:

Thank you very much for your kind suggestions. We revised the third paragraph of 'Discussion: Implications of the study findings':

*'We found that patients did not care about alternative technologies currently covered by insurance. However, it is an essential attribute in reimbursement decision-making. Decision-makers need to compare the new technologies with available alternative technologies and determine whether to cover new medical technologies or obsolete the alternatives.'*

**14. Depending on your response to Question 8 above, you may need to revise your interpretation of the results around the cost attribute**

Reply:

Thank you so much for you kind suggestions. Patients had out-of-pocket spending for treatment technologies even they had social health insurance, and they care about the money which they would pay for new technologies. The reimbursement level of a specific technology depends on the category of the technology in the health insurance catalog and also the type of insurance. We know what we want to investigate, and we provided explicit definitions of the out-of-pocket costs to patients.

**15. Tables & Figures: Table 2: Classification of employment seems unusual. Is 'farmer' really the only employment option for rural residents?**

Reply:

Yes, farmers are those who engaged in agriculture, raising living organisms for food or raw materials. They are all rural residents. We are very sorry for the description of the employment which could lead to misunderstanding. We checked data and revised the type of employment in table 2.

**16. Table 2: 80% of people being an inpatient seems quite high – what does this relate to?**

Reply:

Results from a separate analysis in the original manuscript focusing exclusively on inpatients were similar to the results for all patients. The priorities of health insurance reimbursement in China are the inpatients since inpatients had substantial economic burden. For example, according to the China Health and Family Planning Statistics Yearbook 2019, average cost per inpatient was CNY 13313.3 in year 2018 (about US$2029.5), while the average cost per outpatients was only CNY 322.1 (about US$49.1). The burden of inpatients was much higher than outpatients. Therefore, inpatients need the reimbursement of new technologies more urgently. In fact, they are the priority groups in the context of China's social health insurance policy.

**17. Appendix 3, supplemental table 2: would be good to also split out the number of patients with each disease who had UEBMI without extra benefits**

Reply:

We are very grateful to you for your suggestions. We divided the number of patients with each disease who had UEBMI without additional benefits according to your comments. The new tables were in Appendix 5 (Supplemental Table 3).

**18. Appendix 4, supplemental figure 1: could patients only have one other type of comorbidity (ie no multimorbidity)? Also, I would suggest selecting a different chart type (e.g. bar graph) rather than a 3D pie chart to display this information.**

Reply:

Thank you so much for your suggestions. In fact, many patients have more than one type of comorbidity. We admit that the 3D pie chart is inferior to bar graph in displaying this information. The new figure was in Appendix 6.

**19. Overall: Needs a proofread for spelling mistakes, e.g. a. Methods, line 22 – DEC should be DCE; b. Acknowledgements – 'out interviewers' should be 'our interviewers'**

Reply:

We are very grateful to you for your responsible and careful inspection. We read the manuscript carefully and revised the spelling mistakes.

**Reviewer: 2**

**Reviewer Name: Amarech G. Obse**

**Institution and Country: University of Cape Town, South Africa**

Please state any competing interests or state 'None declared': None declared

Many thanks for the opportunity to review this paper it is an interesting study. I hope the comments and suggestions are useful.

Kind regards,
Amarech

**1. Background: You have mentioned that there are differences in benefits between URRBMI and UEBMI members (and referred the reader to another paper by Yu for details). However, you may need to highlight the differences in health insurance benefits for members with hypertension and diabetes. This is important information since the paper provided analysis of the differences in valuation of attributes of new technologies for these diseases. Similarly, you briefly mentioned that "[e]ven among the UEBMI enrollees, insurance benefit is not equal". Again, you need to expand this a little further so that the sub-group analysis you provided will be well understood.**

Reply:

Dear Amarech, we are very grateful to you for your comments. Your comments are very important and helpful. We have revised the manuscript according to your comments.

(1) We added following sentences in the second paragraph of 'Introduction' to highlight the underlying reasons for investigating patients with hypertension and diabetes:

*'China's National Healthcare Security Administration is promoting the health insurance payment based on diagnosis-related groups (DRGs), a patient classification for standardizing payment in the national health insurance schemes. Accordingly, medical fees and insurance payments will be based on DRG classification, which includes chronic diseases such as diabetes and hypertension.'*

Consequently, subgroup analysis of different type of disease was reasonable and the results would be helpful to inform health insurance reimbursement decision-making.

(2) In order to better describe 'even among the UEBMI enrollees, insurance benefit is not equal', we added an example in the third paragraph of 'Introduction':

*'For example, if the medical expenditure exceeds the ceiling of health insurance reimbursement, outpatients and inpatients that enjoy civil servant subsidies could be subsidized by 70% and 80% respectively for the exceeding parts.'*

(3) We also added data about the differences between UEBMI and URRBMI in the third paragraph of 'Introduction':

*'According to the 2018 statistical bulletin issued by China's National Healthcare Security Administration, the average per capita hospitalization cost for inpatients was 11,181 CNY (about US$1704) for UEBMI enrollees, while 6577 CNY (about US$1003) for URRBMI enrollees. The average inpatient reimbursement ratio for UEBMI enrollees was 71.8% (61.5% for inpatients in tertiary hospitals), while the reimbursement ratio for URRBMI enrollees was 56.1% (49.0% for inpatients in tertiary hospitals). '*

**2. Page 8, hypotheses - the first hypothesis is not clear. It is stated as "new technology attributes regarding health benefits are most preferred by chronic disease patients for health insurance coverage". Do you mean all attributes of any new technology are preferred? In comparison to what? When I read this, I thought you will have attributes/levels of the status quo or opt-out (or status quo) option to support (or reject) this hypothesis, but I see you have forced choice sets. So, how can you provide evidence in support or against this hypothesis?**

Reply:

(1) We are very grateful to you for your comments. The first hypothesis meant that the attributes related to health benefits would be more important as compared with other attributes. We revised the statement of the first hypothesis to make it more explicit.

The revised hypothesis was as follows:

*'new technology attributes regarding health benefits are more important than other attributes for health insurance coverage'.*

(2) To explain why forced choice sets were used in our DCE survey, we added the following sentences in the second paragraph of 'Methods: Experimental design and development of the questionnaire':

*'The forced choice sets were used in our DCE because when no option had a definitive advantage, it was assumed that forced choice under preference uncertainty would favor options that were easier to justify and associated with a lower likelihood of error and regret, such as compromise and asymmetrically dominating options.'*

In other words, such options were not selected primarily because of the utility embedded in their attribute values, but rather because they could help patients to resolve difficult choices.

**3. Attribute selection: You mentioned the methods followed to select attributes and levels, but it is not clearly indicated how the final attributes were selected. Only the criteria identified from systematic review is listed. The findings from the other methods is not clear.**

Reply:

(1) Thank you so much for your comments. We did not realize this problem before and we added the main findings from other methods in the second paragraph of 'Methods: Identification of technology attributes and levels'.

(2) As a matter of fact, two articles regarding attributes selection was published by us in Chinese journals with certain reputation, one was the systematic review (Reference number 28: Chen XW, Chen Q, Geng JS, *et al*. The application of multi-criteria decision analysis in medical insurance reimbursement: a systematic review. *Chinese Health Resources* 2018;21:218-223), and the other was the focus group and expert consultation (Reference number 29: Geng JS, Chen XW, Yu XL, *et al*. Study on the evidence-based decision-making framework for reimbursement of new technologies in view of EVIDEM. *Chinese Journal of Health Policy* 2018;11:50-54).

**4. During the FGDs for attribute selection, you stated that there was no consensus among experts about attribute levels. Is this only about the attribute levels or both attribute and their levels? Can you provide an example of an attribute/level where there was disagreement? How does the database search help you in rectifying the differences? I understand the database search still results in long list of attributes/levels, how was the final decision made? Please**

**clarify.**

Reply:

(1) We are sorry for the mistakes for the word ‘experts’ in the manuscript. In fact, the experts we listed here were physicians. We have corrected such mistakes in the second paragraph of ‘Methods: Identification of technology attributes and levels’.

(2) We tried to get consensus about the level of attributes among physicians during a workshop. During our workshop of evidence-based medicine in Medical School of Nantong University, near 30 physicians attended and most of them gave us the feedback about the levels of attributes. However, there were no consensus from them and they thought it was difficult to define the levels. They believed that there lacked the evidence-based clinical practice guidelines told them about the most accurate levels of effectiveness, safety, and cost.

(3) Then, we decided to do the systematic review, to find information about attributes and levels from published paper. We also did focus group discussion and expert consultation to determine the attributes and levels.

(4) We did a pilot study that included 90 patients with diabetes or hypertension. Intelligibility, acceptability and reliability of the questionnaire were tested during the pilot. Finally, we conducted the formal survey.

**5. On Page 9 lines 21-26, you stated that “[…] we defined new medical technologies in this study as the therapeutics for hypertension and diabetes, which had been marketed in China but were not covered by the public health insurance programs in Jiangsu in 2018”. If the medical technologies used in the DCE are already in market, how did you use the information from the literature reviews and the FGDs (conducted to identify attributes and levels)? Or why was there a need to go through finding attribute and levels when the medical technologies used in the final DCE already existed?**

Reply:

(1) Each medical technology its own attributes. We needed to find out the attributes and levels of these technologies which were relevant to health insurance reimbursement decision-making. However, were not able to include all of the identified attributes in our DCE. Therefore, focus group

discussion and expert consultation were needed to help us determine the attributes and levels of DCE.

(2) The new medical technologies in our research were those already in the market, and we needed to get information about the attributes such as effectiveness and safety from health technology assessment reports. In other words, we needed to find evidence regarding each attributes for these technologies.

(3) In a word, we used the methods to determine the levels of attributes which were most suitable to solve our research problems.

**6. I understand the attributes for a DCE that aims to assess preferences for new technologies should be attributes describing the new technologies. I do not understand how the attribute "alternative technologies currently covered by insurance" is considered as an attribute for a new technology. Please clarify.**

Reply:

(1) We established the evidence-based decision-making framework for reimbursement of new technologies on the basis of EVIDEM. The paper was published in a Chinese journal (Reference number 29: Geng JS, Chen XW, Yu XL, *et al*. Study on the evidence-based decision-making framework for reimbursement of new technologies in view of EVIDEM. *Chinese Journal of Health Policy* 2018;11:50-54). Unmet needs of reimbursed technology was determined as a necessary criteria for the reimbursement decision-making. 'Alternative technologies currently covered by insurance' was used as an attribute in our DCE just with the consideration of 'unmet needs of reimbursed technology'.

(2) Although this attribute is not the attribute directly related to new medical technology, reimbursement decision makers cannot neglect the existed technologies which are already reimbursed. Decision makers need to consider whether to include new medical technologies in the health insurance catalogue, or obsolete the alternatives.

(3) During the DCE survey, we explained this attribute to patients as alternative technologies already reimbursed had similar effectiveness and safety to the new technology (Level: Yes); no alternatives already reimbursed had similar effectiveness and safety (Level: No) to the new technology. To make it clearer, we added the definition of this attribute in Appendix 1.

**7. The attribute "severity of target disease" may be correlated with respondents and may bias valuation of attributes. For instance, an individual with severe disease my choose an alternative with a level 'severe' for this attribute without considering the other alternatives. Please clarify how you dealt with such correlation.**

Reply:

(1) We are very grateful to you for such an important comment. We recognize that it is hard to determine accurately disease severity among patients with chronic diseases. As you know, it is not easy to determine which kind of comorbidity is severe and which is non-severe. However, your comments remind us that we also tested the EQ-5D-5L of each patient in the DCE survey questionnaire. We did not use all of the data in our research previously since we did not realize the problem of severity of patients, and we did not think it is necessary to analyze the data from EQ-5D-5L.

(2) EQ-5D-5L index value can help us to explore whether there was correlation between health-related quality of life and the preference for severity of target disease.

(3) We retrieved literature regarding EQ-5D-5L index value for hypertension and diabetes patients, and used 0.8 as the threshold. We found that patients with lower EQ-5D-5L index value (≤0.8) were more likely to prefer the attribute 'severity of target disease' (Appendix 7: Supplemental table 6).

(4) We also found that severity of target disease was an important attribute for URRBMI patients (P<0.01), but not for UEBMI patients (Appendix 7: Supplemental table 5).

(5)We did further analysis and results showed that the proportion of patients with lower EQ-5D-5L index value was significantly higher in URRBMI group (P<0.01) (Appendix 8).

**8. Please clarify why forced choice set is regarded to be suitable for this study given the limitation that it results in overstatement of the valuation of attributes.**

Reply:

We added the following sentences in the second paragraph of 'Methods: Experimental design and development of the questionnaire' to list the underlying reasons for using forced choice sets:

*'The forced choice sets were used in our DCE because when no option had a definitive advantage, it was assumed that forced choice under preference uncertainty would favor options that were easier to*

*justify and associated with a lower likelihood of error and regret, such as compromise and asymmetrically dominating options.'*

**9. Sample size:**

**You indicated that you have sufficient sample size (408) for overall and sub-group analyses based on the recommendation of a sample size of 150 suggested as sufficient from the literature. This would have been correct if you did not block your design. The sample size requirement for blocked design increases exponentially. You have indicated that your experimental design is partitioned into five blocks; which means you need five people to complete one design. So, your effective sample size is about 82 (408/5 = 81.6). The number of observations in Table 3, which is 4896, also indicates this. Please clarify.**

Reply:

(1) We used the blocking design due to its advantages. Blocking mainly ensures that the level balance of attributes within each block is satisfied, so that respondents are not only faced with low or high attribute levels for a certain attribute. The orthogonal template of 30 choice situations was divided into 5 blocks, so that each respondent only had to address just 6 choice situations instead of 30. We added the advantaged mentioned above in the first paragraph of 'Methods: Experimental design and development of the questionnaire'.

(2) We still unclear about why the sample size requirement for blocked design needs to increase exponentially, since the level balance of attributes within each of the blocks is also satisfied.

(3) Generally speaking, a less efficient design may also require a larger sample size, resulting in increased costs.

(4) We did not find the standardized sample size calculation method for our design. We agree with your opinion that the number of total observations is 4896, with 12 observations for each patient, and 4896/12=408.

(5) According to other researchers' empirical experience, more than 20 respondents per questionnaire version could estimate reliable DCE models (Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomic*s 2008;26(8):661-77).

(6) Previous DCE which used the block design believed that 50 per block would provide sufficient statistical power. A DCE had 116 participants with 2 blocks of choice sets (Oluboyede Y, Ternent L, Vale L, *et al*. Using a discrete-choice experiment to estimate the preferences of clinical practitioners for a novel non-invasive device for diagnosis of peripheral arterial disease in primary care. *Pharmacoecon Open* 2019;3(4):571-81).

Another DCE with good quality only had 237 respondents, a fractional blocked design of 64 scenarios (eight blocks of eight scenarios) (Shanahan M, Seddon J, Ritter A, et al. Valuing families' preferences for drug treatment: a discrete choice experiment. *Addiction* 2020;115(4):690-99).

(7) Consequently, we revised the description about sample size in the second paragraph of 'Methods: DCE implementation and data collection'.

**10. Please clarify how the pilot study informed the final survey. Was there any change in the design or attribute or any other aspect of the study based on the pilot?**

**Reply:**

(1) A good principle for design a DCE is as follows: start simple, and gradually add complexity to the design. For constructing D-optimal designs of choice sets for the pilot DCE survey, no prior parameters are used (i.e, we assume the priors are all zeros).

(2) We did the pilot and got the priors and the distribution to inform the formal Bayesian efficient design.

(3) Intelligibility, acceptability and reliability of the questionnaire including attributes and levels were tested during the pilot. The attributes and levels were not changed in formal design.

**11. Data analysis and results:**

**You stated that "[…] analysis of the DCE data was based on the random utility model and assumed that the utility of two scenarios was equal". I wonder how the assumption of equal utility works. If individuals get the same utility from two alternatives; that means, they are indifferent between the two alternatives. And we may not need to ask them to make choice in the first place. This is in contradiction with the theory underpinning a DCE. Please clarify.**

Reply:

We are sorry for our statement. We deleted wrong statement of the equal utility in the first paragraph of 'Methods: Data analysis'.

**12. Conditional logit model has an assumption of the independence of irrelevant alternatives which is violated by a DCE data. With development in statistical software, more flexible models that accommodate limitations of logit model has been suggested. Please explain why conditional logit is estimated given the restrictive assumption and why other flexible models were not considered.**

Reply:

We are very grateful to you for your comments. We used mixed logit to obtain more reliable results. Details of mixed logit model were stated in the second paragraph of 'Methods: Data analysis' as follows:

*'We implemented the above equation by estimating mixed logistic regressions using STATA 14.2 SE (StatCorp LP, College Station, Texas, USA) and was specified with 500 Halton draws. Mixed logit model allows for unknown heterogeneity in individual preferences and estimates both the mean preference weight and the standard deviation. We assumed that all variables of the attributes, except for the constant, had a random component and that the weights of preference were normally distributed.'*

**13. In relation to the comment above, flexible models are also used to assess unobserved heterogeneity in valuation of attributes among the respondents. Please clarify why this was not considered.**

Reply:

Thank you so much for your suggestions. In response to your comments, mixed logit was used in our data analysis. However, the conclusions drawn from mixed logit model regarding main effects model, subgroup analysis by hypertension and diabetes, subgroup analysis by insurance types remained unchanged when compared with the previous clogit model.

**14. For conditional logit, alternative specific constant should be reported in the result tables.**

Reply:

Alternative specific constant was also used in mixed logit model, however, we did not report it since it was not meaningful and perhaps it could lead to misunderstanding by readers.

**15. Please clarify why dummy coding is used for categorical variables in comparison to other schemes of coding; particularly, given that the marginal utility of the reference level of the dummy coded variable is confounded with alternative specific constant.**

Reply:

(1) We are sorry for not mention the reasons for dummy coding in the manuscript.

(2) Two commonly used methods for categorical coding of attribute levels are effects coding and dummy coding. According to ISPOR guideline for DCE, the decision to use effects coding or dummy coding of variables should be based on ease of interpretation of the estimates from the model, and not on the expectation that one type of coding will provide more information than the other (Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health* 2016;19:300-15).

According to our research objective, results from dummy coding would be easier for us to interpret.

(3) In order to explain the underlying reason for choosing dummy coding, we added the following sentences in the second paragraph of 'Methods: Data analysis' as follows:

*'For dummy variable coding, each model-estimated coefficient is a measure of the strength of preference of that level relative to the omitted level of that attribute.'*

**16. Sub-group analysis is performed by type of disease and type of insurance. I wonder why other socio-demographic variables are not considered. In particular, gender differences in valuation of health care services is widely reported.**

Reply:

(1) Our study aims to provide evidence to inform health insurance reimbursement decision-making. We conducted sub-group analysis by type of disease and type of insurance because these are

important factors considered by policymakers during the health insurance reimbursement policy-making process.

(2) We agree with your suggestion that gender differences in valuation of healthcare service were widely reported. Therefore, we did the subgroup analysis by gender to investigate whether gender would have some impact on the preferences. However, no important differences between subgroups were identified. We did not do the analysis by other socio-demographic characteristics, since patients were not treated deferentially by socio-demographic characteristics in health insurance reimbursement policy-making in China.

**17. Page 13, line 12 … Please clarify that you have two groups in UEBMI, you mentioned that you have further analysis by UEBMI (please see the comment on background)**

Reply: Thank you very much for your suggestions. We did further analysis by UEBMI and revised the description *'There was no statistically significant difference between hypertension and diabetes patients in terms of insurance types'* to the following:

*'There was no statistically significant difference between hypertension and diabetes patients in terms of insurance types (UEBMI vs URRBMI, P=0.392) and UEBMI benefits (UEBMI with extra benefit vs UEBMI without extra benefit, P=0.598).'*(Second paragraph of 'Results: Characteristics of patients').

**18. The sub-group analysis by type of disease is narrated as respondents with hypotension only versus diabetes only. But the analysis was done on respondents with hypertension only versus respondent with both hypertension and diabetes, as indicated in Table 3. This needs to be corrected. Why does the respondents with both diabetes and hypertension added to the groups with diabetes only and not with patients with hypertension only? Please clarify if there is any clinical or insurance related reason for this.**

Reply:

(1) We are very grateful to you for your suggestions. In response to your comments, we did the subgroup analysis on hypertension patients and diabetes patients in the revised manuscript.

(2) Results were shown in the section 'Results: Subgroup analysis by type of disease', and data were listed in Appendix 7 (Supplemental Table 4).

**19. Discussion: This section is thin, needs to be strengthened.**

Reply:

We have revised discussion according to your suggestions.

**20. There are suggestions of new technologies that needs to be considered for the introduction of new technologies. But, how is this compared with technologies currently in use. I think this is important information for policy.**

Reply:

(1) We firmly agree with you that policy-makers need to compare new technologies with technologies currently in use and appreciate your suggestions. We want to provide evidence on reimbursement decision-making from patients' perspective.

(2) To provide further information for policy-makers, we added the following sentence in the first paragraph of 'Discussion: Implications of the study findings':

*'Policy-makers are suggested to make evidence-based comparisons among technologies according to the attributes patients preferred to achieve patient-centered and evidence-informed reimbursement decision-making.'*

**21. Please clarify if the statement on page 19 line 46-48 … "However, variation in patient preferences existed and mainly depended on patients' own experiences of the disease." What do you mean by "patients' own experiences"? Is this statement based on your finding or another research?**

Reply:

We are sorry for the unclear phrase of 'patients' own experiences'. The underlying meaning was the impact of the medical technology on the improvement of symptoms and the clinical outcomes. These are the most important attributes that patients can feel by themselves. We deleted the word 'experience' to avoid misunderstanding.

**22.Page 20 - You discussed about inequalities in health insurance coverage affecting preferences which is fine. But I do not understand the discussion further provided about**

**inequalities and not sure if it directly relates to your findings. I think, the discussion is a bit stretched.**

Reply:

(1) We are so glad that you agree with our discussion about the inequalities in health insurance coverage could affect preferences.

(2) To make this paragraph even more explicit, we deleted the following sentence in the third paragraph of 'Discussion: Comparison with other studies':

*'Earlier pilot of URRBMI attempting to integrate health insurance for urban and rural residents and increase benefit coverage was not able to substantially reduce the inequality between URRBMI and UEMBI patients'.*

**23. P21 line 30. Here you stated that the attribute 'alternative technologies currently covered by insurance' is not valued because URRBMI patients have lower education. Firstly, the education of the respondents by insurance type is not reported in the result section. Secondly, how can you relate less valuation of this attribute to a single sociodemographic characteristic of respondents?**

Reply:

We are very grateful to you for your comments. We did not investigate the impact of education on the preference of medical technologies since education would not be considered in health insurance reimbursement decision-making. Therefore, we deleted the description regarding education in the third paragraph of 'Discussion: Implications of the study findings'.

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Amarech Guda Obse<br>University of Cape Town, South Africa |
| **REVIEW RETURNED** | 01-Jul-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors addressed the raised issues. Only need to improve presentation of the findings and the language. For instance, data were now estimated with mixed logit but in the presentation of finding, only the values of the parameters are changed. What does the mixed logit result indicate about heterogeneity?<br><br>Page 52 lines 45-56 - you stated that "Blocking mainly ensures that the level balance...". I am wondering if this statement is necessarily true. Please provide a source. Is this the reason why |

| | |
|---|---|
| | you blocked the design? But, level balance does not necessarily need blocking of a design. Please clarify why the design was divided into blocks.<br><br>Page 55, line 46. You described the two components of a utility function as "utility component and random component". Please improve the language. Both are components of utility; only that one is observable and the other unobservable. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer: 2
Reviewer Name: Amarech Guda Obse
Institution and Country: University of Cape Town, South Africa
Please state any competing interests or state 'None declared': None declared

1. The authors addressed the raised issues. Only need to improve presentation of the findings and the language. For instance, data were now estimated with mixed logit but in the presentation of finding, only the values of the parameters are changed. What does the mixed logit result indicate about heterogeneity?
Response:
We are very grateful to you for your valuable comments. We added the results and implications of heterogeneity in 'results' and 'discussion'.
(1) Results (Regression analysis of the DCE data):
'Unobservable preference heterogeneity as indicated by the estimated standard deviation (SD) of the mean coefficients, were identified for four variables— expected gains in health outcomes from the treatment, likelihood of effective treatment, out-of-pocket costs, and lethal disease.'
(2) Results (Subgroup analysis by type of disease):
'The SD revealed coefficient heterogeneity within both subgroups for the random parameters of three variables—likelihood of effective treatment, lethal disease, and out-of-pocket costs.'
(3) Results (Subgroup analysis by type of insurance):
'Preference heterogeneity for the lowest incidence of serious adverse events was identified.'
(4) Results (Subgroup analysis by HRQoL):
'For patients with EQ-5D-5L index value less than or equal to 0.8, expected gains in health outcomes from the treatment had more variation in heterogeneity than other attributes. However, the heterogeneity for lethal disease was most significant for patients with EQ-5D-5L index value greater than 0.8.'
(5) Discussion (Summary of the findings)
'The degree to which respondent preferences were heterogeneous was described by the estimated SD around each mean preference estimate. Heterogeneity was found mainly for four variables— expected gains in health outcomes from the treatment, likelihood of effective treatment, out-of-pocket costs, and lethal disease. Although heterogeneity existed, the preferences for new technologies with expected gains in health outcomes from the treatment, and likelihood of effective treatment remained significant in all patients and each subgroup, suggesting that these attributes were generally valued by patients. Variations in preferences over out-of-pocket costs and lethal disease had implications for the optimal design of insurance reimbursement schemes and should be analyzed in future research.'

2. Page 52 lines 45-56 - you stated that "Blocking mainly ensures that the level balance...". I am wondering if this statement is necessarily true. Please provide a source. Is this the reason why you blocked the design? But, level balance does not necessarily need blocking of a design. Please clarify why the design was divided into blocks.
Response:

We are very grateful to you for your comments. As you suggested, we provided a source by indicating that ‘Blocking design promoted response efficiency by reducing the potential cognitive burden to respondents (Ref 31. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16:3-13)’.

3. Page 55, line 46. You described the two components of a utility function as "utility component and random component". Please improve the language. Both are components of utility; only that one is observable and the other unobservable.
Response:
Thank you so much for your comments. We revised the description as ‘observable component and unobservable component’.

**VERSION 3 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Amarech Guda Obse<br>University of Cape Town<br>South Africa |
| **REVIEW RETURNED** | 07-Aug-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Dear authors,<br><br>The paper needs language editing throughout.<br>I indicated few of the editorial comments in the manuscript.<br><br>Kind regards,<br><br>The reviewer provided a marked copy with additional comments.<br>Please contact the publisher for full details. |

**VERSION 3 – AUTHOR RESPONSE**

Reviewer: 2
Reviewer Name: Amarech Guda Obse
Institution and Country: University of Cape Town, South Africa
Please state any competing interests or state ‘None declared’: None declared

1. Abstract: Results
“The technology attributes regarding expected gains in health outcomes from the treatment (P<0.01), high likelihood of effective treatment (P<0.01), and low incidence of serious adverse events (P<0.01) were significant, positive predictors of choice by the study patients.”
(1) Please improve the writing, to avoid repetition of P<0.01;
(2) ... costs were significant or ... cost was a significant
Response:
Thank you so much for your comments. We revised the mistakes as you have mentioned:
(1) ‘The technology attributes regarding expected gains in health outcomes from the treatment, high likelihood of effective treatment, and low incidence of serious adverse events were significant, positive predictors of choice by the study patients (P<0.01).’
(2) ‘The out-of-pocket cost was a significant, negative attribute ...’

2. Methods: Identification of technology attributes and levels
“We used a three-step approach to complete the preliminary stage of DCE, which aimed to identity and define the attributes and levels of new medical technologies.” Should identify instead of identity be used in this sentence?
Response:
We are sorry for the mistakes and identify should be used. We have revised the word:
“We used a three-step approach to complete the preliminary stage of DCE, which aimed to identify and define the attributes and levels of new medical technologies.”

3. Methods: Patient and public involvement
“Response from patients led to more explicit and apprehensible description of the survey questions. Patients were not involved in the recruitment of study participants or the conduct of the study. Findings of the study would be disseminated through publication and social media.
(1) Response from ‘the patients’?
(2) Please re-write to make it clear that these 'patients' were respondents of the pilot... for instance ... Patients involved in the pilot were not included in the final survey ...
(3) The final sentence is not clear ... "Findings of the study would be disseminated through publication and social media" ... Is this about the pilot survey? Which social media? Also I am not sure if you need this sentence here as your paper is already in the process of publication.
Response:
(1) We are very grateful to you for your comments. We have revised the second sentence in the paragraph:
“Response from the patients led to a more explicit and apprehensible description of the survey questions.”

(2) We have revised the third and fourth sentence in the paragraph to make it clear:
“The patients participating in the pilot were not included in the final survey. No patients were involved in the recruitment of study participants or the conduct of the study.”

(3) We deleted the final sentence in the paragraph according to your suggestions.