

## 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>	Factors influencing self-care behaviours of patients with type 2 diabetes in China based on the health belief model: A cross-sectional study
<b>AUTHORS</b>	Hu, Yue; Liu, Huijun; Wu, Jie; Fang, Guixia

## VERSION 1 – REVIEW

<b>REVIEWER</b>	Karimy, Mahmood Saveh University of Medical Sciences
<b>REVIEW RETURNED</b>	31-Oct-2020

<b>GENERAL COMMENTS</b>	<p>Abstract</p> <ul style="list-style-type: none"><li>- In the conclusion of the abstract you should link it to the findings in your results and show how these results were important to reach this conclusion.</li></ul> <p>Introduction</p> <ul style="list-style-type: none"><li>- For a better comparison, the prevalence of Self-care behavior in other country or other parts of the world is not mentioned.</li><li>- In choosing the theory, attention to psychosocial determinants related to behavior is important. But in this section, these determinants of behavior are not mentioned.</li></ul> <p>Methods</p> <ul style="list-style-type: none"><li>- The sample size justification should be presented. What formula is used to determine the sample size in this study?</li><li>- The scientific resource of design of questionnaire should be clearly stated.</li><li>- The method of scoring and handling the data of instruments should be clearly presented.</li><li>- The content validity and reliability and their key elements including qualitative and quantitative evaluation process together with related measures should be clearly stated in details.</li><li>- How much time did participants need to complete the questionnaire booklet? However, it is unclear what was done when the patients finished their questionnaire. How were the questionnaires collected?</li><li>- For multiple regression analyses: 1) report the alpha level used in the univariate analysis; 2) report whether the variables were assessed for a) collinearity and b) interaction; and 3) describe the variable selection process by which the final model.</li><li>- To each of the constructs, written a sample</li></ul> <p>Result</p> <p>Ok</p> <p>Discussion</p> <ul style="list-style-type: none"><li>- What is implication to practice based on the result of this study.</li></ul>
-------------------------	---

	- Improve discussion based on the new research in this area.
<b>REVIEWER</b>	Hall, Alix Univ Newcastle
<b>REVIEW RETURNED</b>	03-Mar-2021

<b>GENERAL COMMENTS</b>	<p>This is a cross-sectional survey study assessing factors related to diabetes self-care of patients from China, based on the health belief model. This is a large and impressive survey study. I have listed below some suggestions for improvement and clarification. I have focused mainly on the statistical analysis section of the study.</p> <p><b>Abstract:</b></p> <ul style="list-style-type: none"> <li>• It would be useful to be more descriptive in the objectives regarding the specific category of factors that were assessed, for instance demographic, socioeconomic and domains of the health belief model.</li> <li>• Please state that this is a cross-sectional survey study.</li> <li>• Under the settings heading a brief description of the complex sampling procedure used to select communities would be helpful to include.</li> <li>• Please include the response rate under the participants heading.</li> <li>• The conclusions in the abstract do not match the results presented in the abstract, please ensure these are aligned.</li> </ul> <p><b>Introduction:</b></p> <ul style="list-style-type: none"> <li>• Please ensure that the studies that are referred to in the introduction are referenced. For instance on page 4 at the end of the introduction section the authors refer to most studies in China include only patients from one city and that there are only a few studies including patients from other regions. However, no references have been provided of these studies. Please ensure that these studies are referenced.</li> </ul> <p><b>Methods:</b></p> <ul style="list-style-type: none"> <li>• Are the items used to assess the constructs of the health belief model part of a standardised scale? Or are these questions that have been developed by the authors specifically for this study? Please include this information.</li> <li>• If the HBM constructs are part of a validated scale please include information on the reliability and validity of this scale.</li> <li>• The information on the scoring and item response scales is confusing. It states that the items were calculated on a one-point scale, but then a range is provided. It is also not clear how the overall scores for these constructs are obtained. Please include the scoring information.</li> <li>• Where do the self-care behaviour items come from? Were they designed specifically for this study or is this a standardised scale. Please include this information.</li> <li>• The scoring for the self-care behaviour items seems to be just a count of behaviours, but the authors treat this variable as though it is an interval outcome, reporting means and standard deviations as well as analysing it as the dependent variable in a linear regression. Is this appropriate, given this seems to be a count variable?</li> </ul> <p><b>Statistical analysis</b></p> <ul style="list-style-type: none"> <li>• The authors used quite a complex sampling procedure and survey data in their regression analysis, were any adjustments used to account for this complex data? On a quick search I found the following paper that may be useful for analysis of such data in</li> </ul>
-------------------------	---

	<p>SPSS. doi: 10.22237/jmasm/1556670300. Unfortunately I am not very familiar with SPSS so cannot provide any specific suggestions for analysis or references in SPSS.</p> <ul style="list-style-type: none"> <li>• How did the authors select the independent variables for inclusion in their multivariable regression? Are they based on theory? Please provide details on the model building approach used.</li> <li>• Did the authors assess any of the model assumptions, in particular multicollinearity? Looking at the variables included in the regression it seems probable that some of these characteristics may be collinear (i.e. work status and monthly income). This should be assessed and if present addressed appropriately. Details on how the main assumptions of linear regression were assessed should be included.</li> <li>• Why did the authors decide to analyse the main outcome in a linear regression? Was this considered appropriate given the count nature of the outcome variable?</li> <li>• Were there any missing data? If so what was the proportion of missing, especially for the main outcome, and how was this handled?</li> <li>• The authors state that the total health belief score was analysed using t-test or analysis of variance. It is not clear what comparisons are being made with these test. Please make sure this is clear and also be specific as to what tests were used to assess what comparisons. It is also not clear what aim this analysis is addressing. Please ensure all analyses have a clear link to the aims of the paper.</li> <li>• Is there a clinically meaningful cut-point for the self-behaviour score or the HBM constructs that can be used to interpret the results? If there is please include in the analysis section and then use this as a point of reference in the discussion.</li> </ul> <p>Results</p> <ul style="list-style-type: none"> <li>• The authors present a comparison of the health belief score by a long list of patient characteristics. It is not clear why this was performed or what aim this is related to. Please make sure there is a clear link between the analysis, results and aims.</li> <li>• The diabetes self-care behaviours are reported as means and standard deviations, is this appropriate given they are a count of seven yes/no questions?</li> <li>• On page 11 the authors report "...no significant effect." As this is a cross-sectional study please refrain from using the term effect and describe the results from the regression as associations.</li> <li>• In table 4 both unstandardized and standardised regression coefficients are presented. Why? The type of regression coefficients should be reported in the analysis section and it be made clear which will be interpreted.</li> </ul> <p>Discussion</p> <ul style="list-style-type: none"> <li>• There are a number of sections in the discussion where the authors report on results from the study (i.e. average score on the health belief scale). Please refrain from introducing new results or restating results in the discussion and use the discussion to interpret the potential meaning of the findings.</li> <li>• There is discussion about results that were not clear aims of the study, for instance the associated characteristic with the level of health beliefs. Please make sure there is a clear link between aims, analysis, results and discussion. The discussion should be focused on interpreting the results relating to the aims of the study.</li> <li>• The authors describe the average self-care score as moderate. What is considered a moderate score on this scale? What does this</li> </ul>
--	---

	<p>mean? Please include in the analysis section what is clinically meaningful for this scale.</p> <ul style="list-style-type: none"> <li>• The conclusions seem to focus only on the findings between the HBM and the self-care score, however the aims and findings were broader, looking at a range of characteristics. Please relate the conclusion to the overall findings and main aim of the study.</li> </ul>
--	--

### VERSION 1 – AUTHOR RESPONSE

**Reviewer #1:**

**1. “ In the conclusion of the abstract you should link it to the findings in your results and show how these results were important to reach this conclusion.”** Thank you very much for your valuable comments. We enriched the conclusion of the abstract in the manuscript. Page 1, line 25-26 ,the statement were corrected as “The perceived severity, benefits, and barriers were key factors affecting self-care behaviors among patients with type 2 diabetes”.

**2. “For a better comparison, the prevalence of self-care behavior in other country or other parts of the world is not mentioned.”**

We are very sorry for our negligence of the prevalence of self-care behaviors in other country, we have added the prevalence of self-care behaviors in different country to the article for a better comparison.

Page3, line 19-22 “Previous studies showed diabetes self-care behaviors in other country is suboptimal. A study by Halal A indicated that diabetes self-care behaviors among Omanis have inadequate level of diabetes self-care behaviors[15]. self-care behaviors among Jordanian patients with type 2 diabetes were also suboptimal” were added.

**3. “In choosing the theory, attention to psychosocial determinants related to behavior is important. But in this section, these determinants of behavior are not mentioned.”**

Thank you very much for your valuable comments, we are very sorry for our negligence of psychosocial determinants related to behavior, it is one of the limitations of this study, we will take these determinants of behavior into consideration in the future study.

**4. “The sample size justification should be presented. What formula is used to determine the sample size in this study? ”**

Thank you very much for your valuable comments. The estimation of the survey sample is based on the number of items in the scale. The items contained in each scale need about 5-10 samples. This study has 64 items in the health belief scale for patients with type 2 diabetes. Taking into account the differences in eastern, central and western regions in China, each region selected 420 type 2 diabetes patients, and a total of 1,260 type 2 diabetes patients were selected. 1260 patients with type 2 diabetes were contacted, but 118 refused to participate in the study, out of the 1142 participants, 2 were subsequently excluded, the final number of participants included in the study was 1140.

**5. “The scientific resource of design of questionnaire should be clearly stated.”**

Thank you very much for your suggestion, we have made correction according to your comments. Page 4, line 3-4 “The questionnaire for diabetic patients was developed to collect data based on literature review, expert evaluation, pilot study, and reliability and validity evaluation.” was added.

**6. “The method of scoring and handling the data of instruments should be clearly presented.”**

Thank you very much for your valuable comments. The method of scoring have been clearly presented on Page 5, line 1-6 “ Each item was calculated on a one-point scale, ranging from 1 (Agree) to 0.25 (Disagree) and 1 (Very helpful) to 0.25 (Hardly helpful); for the perceived barriers construct, ranging from 1 (Very difficulty) to 0.25 (Hardly difficulty).Total health belief score = total score of perceived susceptibility + total score of perceived severity + total score of perceived effectiveness + total score of perceived benefits - total score of perceived barriers, hence, the maximum score is 4, the minimum score is 0.” and the method of handling the data of instruments have been clearly presented on Page 6, line 18-23.

**7. “The content validity and reliability and their key elements including qualitative and quantitative evaluation process together with related measures should be clearly stated in details.”**

We are very sorry for our negligence of content validity and reliability, the content validity and reliability were added to Page 5, line 7-16 “In order to determine reliability, the questionnaire were given to 175 individuals not part of the study, according to this pilot study, the questionnaire was slightly modified. The Cronbach’s alpha coefficient for the health belief scale was 0.895, which showed good reliability. and the Cronbach’s alpha coefficients for each HBM constructs were 0.859 (perceived susceptibility), 0.837 (perceived severity), 0.906 (perceived benefits), 0.604 (perceived barriers), and 0.873 (perceived effectiveness). The construct validity was assessed through the principal component analysis. Seventeen factors were extracted, the cumulative variance contribution rate was 0.615, which suggested the good construct validity. These results demonstrated that the questionnaire was internally consistent.”

**8. “How much time did participants need to complete the questionnaire booklet? However, it is unclear what was done when the patients finished their questionnaire. How were the questionnaires collected?”**

Thank you very much for your valuable comments. We have have made correction according to your comments. Page 6, line 18-21 “The data was collected by trained graduate students, data collection was conducted with the help of community workers. Participants were invited to complete questionnaire in 30 minutes, The trained graduate student was also available to answer questions and clarify any issues as participants filled out questionnaires.” was added.

**9. “For multiple regression analyses: 1) report the alpha level used in the univariate analysis; 2) report whether the variables were assessed for a) collinearity and b) interaction; and 3) describe the variable selection process by which the final model.”**

We are very sorry for our negligence of the alpha level used in the univariate analysis, the alpha level was 0.05,  $p < 0.05$  was considered statistically significant, it's described in page 6, line 27. Collinearity diagnosis and interaction have been assessed in this study, variance inflation factor (VIF) is an indicator to measure the severity of collinearity in a linear regression. In this study, we have adjusted the method of analysis, the VIF is less than 10 which indicating that collinearity did not exist. The method that variables come into the equation was “Enter” which all variables entered the logistic regression model. “The logistic regression analysis was used to examine the predictors of diabetes self-care behaviors based on health belief model.” was added to page 9, line 7-8.

**10. “To each of the constructs, written a sample.”**

We have made correction according to your comments. The sample of each HBM constructs was added on Page 4, line 18-29. “The HBM believes that self-care behaviors are mainly related to the following five aspects: (a) perceived susceptibility (10 items): individuals’ judgment of their own risk of diabetes (e.g. “Everyone can get diabetes”); (b) perceived severity (12 items): individuals’ awareness of serious complications associated with diabetes (e.g. “Diabetes can easily cause damage to the feet”); (c) perceived benefits (11 items): individuals’ perception of the benefits obtained after performing self-care behaviors (e.g. “Get treatment and medication guidance”); (d) perceived barriers (19 items): individuals’ awareness of difficulties in adopting diabetes self-care behaviors (e.g. “Deal with diabetes complications in time”); (e) perceived effectiveness (12 items): individuals’ awareness of the ability to control blood glucose and reduce the delay of diabetic complications after performing self-care behaviors (e.g. “Regular blood glucose measurement”).”

**11. “What is implication to practice based on the result of this study.”**

We are very sorry for our negligence of the implication to practice based on the result of this study.

Page 13, line 14-21 “This study provides support for the use of health belief model to assess adherence of diabetes patients to self-care behaviors among patients with type 2 diabetes in China. Future researches are recommended to examine how interventions can be designed to use HBM as a part of diabetes patients education. The results of this study provide an essential framework to educate patients with diabetes to increase their adherence to self-care behaviors, strategies to promote self-care behaviors for patients with type 2 diabetes are important parts of diabetes education programs.” was re-written.

**12. “Improve discussion based on the new research in this area.”**

Thank you very much for your valuable comments, we have re-written the conclusions according to your suggestion. Page 13, line 10-21 “The findings of this study indicated that some Chinese patients with type 2 diabetes carried out suboptimal self-care behaviors. Patients with diabetes might have an increased risk of developing diabetes-related complications such as nephropathy and retinopathy. There are many factors that affect self-care behaviors of these participants which will

need further investigation. This study provides support for the use of health belief model to assess adherence of diabetes patients to self-care behaviors among patients with type 2 diabetes in China. Future researches are recommended to examine how interventions can be designed to use HBM as a part of diabetes patients education. The results of this study provide an essential framework to educate patients with diabetes to increase their adherence to self-care behaviors, strategies to promote self-care behaviors for patients with type 2 diabetes are important parts of diabetes education programs.” was re-written.

Reviewer #2:

### General Comments:

**“This is a cross-sectional survey study assessing factors related to diabetes self-care of patients from China, based on the health belief model. This is a large and impressive survey study. I have listed below some suggestions for improvement and clarification. I have focused mainly on the statistical analysis section of the study.”**

We really appreciate your careful reading and evaluation for our research. According to your comments, we enriched the content of theoretical framework the Introduction part and added more explanations for the reasons why the predictors and covariates were selected.

- 1. “It would be useful to be more descriptive in the objectives regarding the specific category of factors that were assessed, for instance demographic, socioeconomic and domains of the health belief model.”**

It is really true as you suggested that it would be useful to be more descriptive in the objectives.

Page 1, line 9-11 “The aim of this study was to explore the status of self-care behaviors and self-care behaviors predictors among patients with type 2 diabetes in China based on health belief model.” was corrected.

- 2. “Please state that this is a cross-sectional survey study.”**

We are very sorry for our negligence, considering your suggestion, we have added this part in the design.

Page 1, line 12 “The cross-sectional study was conducted on 1140 patients aged  $\geq 36$  years old with type 2 diabetes who had established health records in community health service institutions. The questionnaire was designed based on the health belief model, which mainly included: perceived

susceptibility, severity, benefits, barriers, effectiveness, sociodemographic characteristics and self-care behaviors” was added.

**3. “Under the settings heading a brief description of the complex sampling procedure used to select communities would be helpful to include.”**

We are very sorry for our negligence of the complex sampling procedure used to select communities. Page 1, line 17-18 “Using a multistage sampling method randomly selected 36 villages and communities in China.” was corrected.

**4. “Please include the response rate under the participants heading.”**

We are very sorry for our negligence of the response rate, we have included under the participants heading. Page 1, line 21 “90.5% response rate” was added.

**5. “The conclusions in the abstract do not match the results presented in the abstract, please ensure these are aligned.”**

We have rewritten the conclusions in the abstract in order to match the results presented in the abstract.

Page 1, line 25-27 “The perceived severity, benefits, and barriers were key factors affecting self-care behaviors among patients with type 2 diabetes, health education for patients should be strengthened to improve the self-care level of patients with diabetes.” was corrected.

**6. “Please ensure that the studies that are referred to in the introduction are referenced. For instance on page 4 at the end of the introduction section the authors refer to most studies in China include only patients from one city and that there are only a few studies including patients from other regions. However, no references have been provided of these studies. Please ensure that these studies are referenced.”**

It is really true as you suggested that we have provided some references in the introduction part. Page 3 line 25-26, “The relationship of health beliefs and complication prevention behaviors of Chinese individuals with Type 2 Diabetes Mellitus” and “Status quo of self-management behaviors and its influencing factors among type 2 diabetes patients in ShanDong province” were added references.

**7. “Are the items used to assess the constructs of the health belief model part of a standardised scale? Or are these questions that have been developed by the authors specifically for this study? Please include this information.”**

We have made correction according to the your comments.



Page 4, line 3-4 “The questionnaire for diabetic patients was developed to collect data based on literature review, expert evaluation, pilot study, and reliability and validity evaluation.” was added. Page 4, line 17-18 “HBM constructs: This part consisted of 64 items derived from the available literature<sup>17,19</sup> .” was added.

**8. “ If the HBM constructs are part of a validated scale please include information on the reliability and validity of this scale.”**

We are very sorry for our negligence of content validity and reliability, the content validity and reliability were added to Page 5, line 7-16 “In order to determine reliability, the questionnaire were given to 175 individuals not part of the study, according to this pilot study, the questionnaire was slightly modified. The Cronbach’s alpha coefficient for the health belief scale was 0.895, which showed good reliability. and the Cronbach’s alpha coefficients for each HBM constructs were 0.859 (perceived susceptibility), 0.837 (perceived severity), 0.906 (perceived benefits), 0.604 (perceived barriers), and 0.873 (perceived effectiveness). The construct validity was assessed through the principal component analysis. Seventeen factors were extracted, the cumulative variance contribution rate was 0.615, which suggested the good construct validity. These results demonstrated that the questionnaire was internally consistent.”

**9. “ The information on the scoring and item response scales is confusing. It states that the items were calculated on a one-point scale, but then a range is provided. It is also not clear how the overall scores for these constructs are obtained. Please include the scoring information.”**

We have re-written this part according to your suggestion.

Page 5, line 1-6 “3. HBM constructs: ..... Each item was calculated on a one-point scale, ranging from 1 (Agree) to 0.25 (Disagree) and 1 (Very helpful) to 0.25 (Hardly helpful); for the perceived barriers construct, ranging from 1 (Very difficulty) to 0.25 (Hardly difficulty). Total health belief score = total score of perceived susceptibility + total score of perceived severity + total score of perceived effectiveness + total score of perceived benefits - total score of perceived barriers, hence, the maximum score is 4, the minimum score is 0.” was re-written. Page 4, line 11-16 “2. Self-care behaviors of diabetic patients: Self-care behaviors were selected based on Chinese Diabetes Management Guideline. It consisted of 7 dimensions: foot care, drinking, smoking, medication adherence, self-testing weight, self-testing blood glucose and number of follow-ups. For each dimension, a positive and beneficial choice scored 1; otherwise scored 0. The total score of self-care behaviors ranging from 0 to 7 (Table 1), higher scores represented better self-care behaviors.” was added.

**10. “Where do the self-care behaviour items come from? Were they designed specifically for this study or is this a standardised scale. Please include this information.”**

The diabetes self-care behaviors were selected based on Chinese Diabetes Management Guideline, they were designed specifically for this study.

Page 4, line 11-14 “Self-care behaviors were selected based on Chinese Diabetes Management Guideline. It consisted of 7 dimensions: foot care, drinking, smoking, medication adherence, self-testing weight, self-testing blood glucose and number of follow-ups.”

- 11. “The scoring for the self-care behaviour items seems to be just a count of behaviours, but the authors treat this variable as though it is an interval outcome, reporting means and standard deviations as well as analysing it as the dependent variable in a linear regression. Is this appropriate, given this seems to be a count variable? ”**

We are very sorry for our incorrect writing, we have made correction according to your comments. Page 8, line 4-5 the statement “The average score of self-care behaviors for all patients was 4.46 (standard deviation [SD] = 1.09).” was deleted, we also corrected Table 3. We also have adjusted the method of analysis, the total score of self-care behaviors (scores of self-care behaviors  $\leq 3$  is 0,  $> 3$  is 1) was the dependent variable in a logistic regression.

- 12. “The authors used quite a complex sampling procedure and survey data in their regression analysis, were any adjustments used to account for this complex data?”**

Thank you very much for your valuable comments. Page 6, line 18-23 “The data was collected by trained graduate students, data collection was conducted with the help of community workers. EpiData 3.1 software was used for the double-entry and consistency test to ensure data accuracy. Missing values of input data were also processed.” was re-written.

- 13. “How did the authors select the independent variables for inclusion in their multivariable regression? Are they based on theory? Please provide details on the model building approach used.”**

The independent variables mainly included sociodemographic characteristics and HBM constructs, these may influence the self-care behaviors among diabetic patients. Page 9, line 7-8 “The logistic regression analysis was used to examine the predictors of diabetes self-care behaviors based on health belief model.” The method that variables come into the equation was “Enter” which all variables entered the logistic regression model.

- 14. “Did the authors assess any of the model assumptions, in particular multicollinearity? Looking at the variables included in the regression it seems probable that some of these characteristics may be collinear (i.e. work status and monthly income). This should be assessed and if present addressed appropriately. Details on how the main assumptions of linear regression were assessed should be included.”**

Thank you very much for your valuable comments. Collinearity diagnosis have been assessed in this study, variance inflation factor (VIF) is an indicator to measure the severity of collinearity in a linear regression. In this study, the VIF is less than 10 which indicating that collinearity did not exist. We also have adjusted the method of analysis, the method that variables come into the equation was “Enter” which all variables entered the logistic regression model.

**15. “Why did the authors decide to analyse the main outcome in a linear regression? Was this considered appropriate given the count nature of the outcome variable?”**

Thank you very much for your suggestion, we have made correction according to your comments. We have adjusted the method of analysis, the logistic regression analysis was conducted to predict the variables influencing self-care behaviors on the basis of health belief model. The total score of self-care behaviors (scores of self-care behaviors  $\leq 3$  is 0,  $> 3$  is 1) was the dependent variable in the logistic regression. The method of scoring was included on page 4, line 11-16 “Self-care behaviors consisted of 7 dimensions: foot care, drinking, smoking, medication adherence, self-testing weight, self-testing blood glucose and number of follow-ups. For each dimension, a positive and beneficial choice scored 1; otherwise scored 0. The totalscore of self-care behaviors ranging from 0 to 7.”

**16. “Were there any missing data? If so what was the proportion of missing, especially for the main outcome, and how was this handled?”**

After all investigations are completed, the questionnaires were sorted and checked in time, and if key variables are missing or the data is missing up to 5%, they will be treated as invalid questionnaires. For variables with a small amount of missing data, use the mode to fill in the data.

**17. “The authors state that the total health belief score was analysed using t-test or analysis of variance. It is not clear what comparisons are being made with these test. Please make sure this is clear and also be specific as to what tests were used to assess what comparisons. It is also not clear what aim this analysis is addressing. Please ensure all analyses have a clear link to the aims of the paper.”**

We are very sorry for our incorrect writing, we have made correction according to the your comments. Page 6, line 26-28, “Differences in health belief level in different sociodemographic characteristics subgroups were examined using t-test or analysis of variance.” was added. We also have added relevant descriptions in the results part, page 7, line 12-18 “The average score of health beliefs of the participants was( $0.71 \pm 0.08$ ), the scores of each dimension of the health belief model are perceived susceptibility score( $0.68 \pm 0.16$ ), perceived severity score ( $0.80 \pm 0.16$ ), perceived benefits score( $0.81 \pm 0.13$ ), perceived barriers score( $0.42 \pm 0.09$ ), and perceived effectiveness score( $0.84 \pm 0.15$ ). Univariate analysis showed that age, marital status, educational level, account nature, work status, household size and areas were negatively correlated with health belief level (Table 2).” was added.

**18. “Is there a clinically meaningful cut-point for the self-behaviour score or the HBM constructs that can be used to interpret the results? If there is please include in the analysis section and then use this as a point of reference in the discussion.”**

Thank you very much for your valuable comments. We are very sorry that there was not a clinically meaningful cut-point for the self-behavior score or the HBM constructs. The total score of self-behaviors depends on sociodemographic characteristics and other factors, this is the first time that we

explore factors influencing self-care behaviors of patients with type 2 diabetes in China, and we will take the clinically meaningful cut-point for self-care behaviors into consideration in the future study.

**19. “The authors present a comparison of the health belief score by a long list of patient characteristics. It is not clear why this was performed or what aim this is related to. Please make sure there is a clear link between the analysis, results and aims.”**

Thank you very much for your valuable comments. It is really true as you suggested that there should be a clear link between the analysis, results and aims, so we have added relevant descriptions in the results part. Page 7, line 12-18 “The average score of health beliefs of the participants was(0.71 ± 0.08), the scores of each dimension of the health belief model are perceived susceptibility score(0.68±0.16), perceived severity score (0.80 ± 0.16), perceived benefits score(0.81 ± 0.13), perceived barriers score(0.42 ± 0.09 ), and perceived effectiveness score(0.84 ± 0.15). Univariate analysis showed that age, marital status, educational level, account nature, work status, household size and areas were negatively correlated with health belief level (Table 2).” was added.

**20. “The diabetes self-care behaviours are reported as means and standard deviations, is this appropriate given they are a count of seven yes/no questions?”**

We are very sorry for our incorrect writing, we have made correction according to your suggestion. Page 8, line 4-5 the statement “The average score of self-care behaviors for all patients was 4.46 (standard deviation [SD] = 1.09).” was deleted, we also corrected table 3.

**21. “On page 11 the authors report “...no significant effect.” As this is a cross-sectional study please refrain from using the term effect and describe the results from the regression as associations.”**

We are very sorry for our incorrect writing, it is really true as you suggested that this is a cross-sectional study, we have made correction to refrain from using the term effect and describe the results from the regression as associations. Page 9, line 9-12 the statement “significant effect “ was corrected as “significant association”, other similar statements in the manuscript have also been modified.

**22. “In table 4 both unstandardized and standardised regression coefficients are presented. Why? The type of regression coefficients should be reported in the analysis section and it be made clear which will be interpreted.”**

Thank you very much for your valuable comments. The unstandardized regression coefficient reflects the absolute effect of the change of the independent variable on the dependent variable, while the standardized regression coefficient reflects the relative effect of different independent variables on the dependent variable, which can show the importance of the influence of different independent variables on the dependent variable. We have adjusted the method of analysis, the logistic regression

analysis was conducted to predict the variables influencing self-care behaviors on the basis of health belief model. B, OR and 95% CI were presented on table 4.

- 23. “There are a number of sections in the discussion where the authors report on results from the study (i.e. average score on the health belief scale). Please refrain from introducing new results or restating results in the discussion and use the discussion to interpret the potential meaning of the findings.”**

Thank you very much for your valuable comments. It is really true as you suggested that using the discussion to interpret the potential meaning of the findings. We have made correction according to your comments. Page 10, “The average score of health beliefs of the participants was 0.71 (SD = 0.08), and the overall health belief level was relatively high. The order of scores of the five dimensions of the HBM from low to high were as follows: perceived barriers, perceived susceptibility, perceived severity, perceived benefits, and perceived effectiveness. A younger age and a higher educational level were associated with a higher level of health beliefs. Moreover, patients living in the central and urban regions and those who were married and still working had higher levels of health beliefs.” was deleted.

- 24. “There is discussion about results that were not clear aims of the study, for instance the associated characteristic with the level of health beliefs. Please make sure there is a clear link between aims, analysis, results and discussion. The discussion should be focused on interpreting the results relating to the aims of the study.”**

Thank you very much for your valuable comments, we have re-written this part according to your suggestion, the discussion has been focused on interpreting the results relating to the aims of the study. Page 10, line 8-11 “The main purpose of this study was to assess the health belief level and predictors of self-care behaviors based on health belief model and sociodemographic characteristics among patients with type 2 diabetes in China. The study findings indicated that participants had poor self-care behaviors and health belief played an important role in performing self-care behaviors.” was added.

- 25. “The authors describe the average self-care score as moderate. What is considered a moderate score on this scale? What does this mean? Please include in the analysis section what is clinically meaningful for this scale.”**

We are very sorry for our inappropriate writing of the the average self-care score, we have adjusted the expression. The statement “The average score of self-care behaviors was 4.46 (SD = 1.09). The level of self-care behaviors among the patients was moderate.” was corrected as “The study findings indicated that participants had poor self-care behaviors .” on page 10, line 10-11.

- 26. “The conclusions seem to focus only on the findings between the HBM and the self-care score, however the aims and findings were broader, looking at a range of characteristics. Please relate the conclusion to the overall findings and main aim of the study.”**

Thank you very much for your valuable comments, we have re-written the conclusions according to your suggestion, page 13, line 10-21 “The findings of this study indicated that some Chinese patients with type 2 diabetes carried out suboptimal self-care behaviors. Patients with diabetes might have an increased risk of developing diabetes-related complications such as nephropathy and retinopathy. There are many factors that affect self-care behaviors of these participants which will need further investigation. This study provides support for the use of health belief model to assess adherence of diabetes patients to self-care behaviors among patients with type 2 diabetes in China. Future researches are recommended to examine how interventions can be designed to use HBM as a part of diabetes patients education. The results of this study provide an essential framework to educate patients with diabetes to increase their adherence to self-care behaviors, strategies to promote self-care behaviors for patients with type 2 diabetes are important parts of diabetes education programs.” was re-written.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Hall, Alix Univ Newcastle
<b>REVIEW RETURNED</b>	14-Jun-2021

<b>GENERAL COMMENTS</b>	<p>Thank you to the authors for their amendments. The changes to their manuscript have greatly improved this paper. Below are some suggestions to further clarify and strengthen the paper.</p> <ol style="list-style-type: none"> <li>1. Please include the standard deviation along with the mean score of health beliefs in the abstract.</li> <li>2. In the limitations section please provide examples of what type of impact the multistage sampling method could have on the conclusions and why this is considered a limitation.</li> <li>3. In the limitations section it is not clear what the limitation relating to psychosocial determinants is. Is it that the theory that was chosen does not consider this category of determinants, which is important?</li> <li>4. An additional proofread of the paper, particularly the additional text that has been included in this updated manuscript is needed.</li> <li>5. The criteria for the sample size is still not clear. What is the criteria of 5-10 samples based on? The use of this criteria as a basis for why 35 random patients were selected from each region is not clear in the methods. A reference to this criteria for selecting an adequate sample size should be provided, and a statement as to why this is considered suitable for this study included.</li> <li>6. The description of the measurement scales seems incorrect and are stated as a one-point scale. A one-point scale indicates that only one response option is available but it seems that participants select from a number of different responses (i.e. “1 (Agree) to 0.25 (Disagree) and 1 (Very helpful).” Please amend this description so it is clear the number of response options available for these questions.</li> <li>7. In the analysis section it is still not clear whether the authors adjusted for the complex sampling and survey methods in their analysis? Usually with such complex sampling and survey methods adjustments such as clustered standard errors, bootstrapping, Jackknife or weighting procedures etc. are used.</li> <li>8. Please change the term multivariate to multivariable when referring to the regression modelling used throughout the manuscript.</li> <li>9. I appreciate that the authors have updated details on their regression model building. However, it would be good to include</li> </ol>
-------------------------	--

	<p>further detail including: specifically stating that the enter method was used for model building and that the specific predictors selected for assessment were chosen as they are hypothesised to impact on self-care behaviors among diabetic patients.</p> <p>10. Thank you for updating the regression model to ensure it is appropriate for the type of outcome variable. To further strengthen this could the authors please include a brief justification as to why scores of self-care behaviours were dichotomised based on a score of 3? Is this clinically relevant or based on previous literature for example?</p> <p>11. Please ensure the details describing how missing data were addressed are included in the analysis section of the paper.</p> <p>12. In the methods section under the heading “study design” I would include one to two sentences describing the design of the study, which is a cross-sectional survey study. For the description of the survey I would move this under a heading of “Study measures” as it is distinct from the study design.</p> <p>13. It is still not clear why differences in health belief scores by sociodemographic variables were assessed as the aims of the study are stated to be looking at association between health belief constructs and sociodemographic variables with self-care behaviour as the primary outcome. By looking at the univariate associations between the sociodemographic variables with the health belief score, the authors are now treating the health belief score as an outcome, rather than a predictor variable. However, there is no aim relating to this analysis. Please make sure that these analyses are tied to an aim, or removing these comparisons as it is not clear what question this analysis is relating to.</p> <p>14. In table 4 it seems there are p-values listed for each level of a categorical variable (e.g. there is a p-value for each level of education status). Only an omnibus test should be used to assess the association between the independent variables and outcome and an overall p-value provided, rather than a p-value for each level of a categorical variable. Please remove the category level p-values and include only the overall p-value for each variable and ensure that the conclusions about what is statistically associated with the outcome is based on the overall p-value only.</p> <p>15. In the discussion the authors state that patients had poor self-care. How was the self-care score determined to be poor? i.e. what is considered low self-care? I am assuming it is the classification of 3 or less, adding a sentence as to how this dichotomy was determined relevant will help frame this interpretation.</p>
--	--

## VERSION 2 – AUTHOR RESPONSE

### Reviewer 2:

1. **“Please include the standard deviation along with the mean score of health beliefs in the abstract.”**

We are very sorry for our negligence of the standard deviation of health beliefs, we have added the standard deviation in the abstract. Page 1, line 22 “The average score of health beliefs was 0.71(SD=0.08).” were added.

2. “ In the limitations section please provide examples of what type of impact the multistage sampling method could have on the conclusions and why this is considered a limitation.”

Thank you very much for your valuable comments, we are very sorry for our incorrect writing, we have re-written the limitations. Page 13, line 7-11, “This study had a few limitations that should be considered when interpreting the findings of this study. First, this was a cross-sectional study; therefore, a causal relationship between self-care behaviours and health beliefs could not be inferred. In addition, all data were based on the self-report of participants; therefore, the findings might be limited by a recall bias or social desirability bias.” were corrected.

3. “In the limitations section it is not clear what the limitation relating to psychosocial determinants is. Is it that the theory that was chosen does not consider this category of determinants, which is important? ”

Thank you very much for your valuable comments, we are very sorry for our incorrect writing, we have re-written the limitations. Page 13, line 7-11, “This study had a few limitations that should be considered when interpreting the findings of this study. First, this was a cross-sectional study; therefore, a causal relationship between self-care behaviours and health beliefs could not be inferred. In addition, all data were based on the self-report of participants; therefore, the findings might be limited by a recall bias or social desirability bias.” were corrected.

4. “An additional proofread of the paper, particularly the additional text that has been included in this updated manuscript is needed.”

Thank you very much for your valuable comments. We have invited a professional copyediting service to improve the quality of the English throughout the manuscript.

5. “ The criteria for the sample size is still not clear. What is the criteria of 5-10 samples based on? The use of this criteria as a basis for why 35 random patients were selected from each region is not clear in the methods. A reference to this criteria for selecting an adequate sample size should be provided, and a statement as to why this is considered suitable for this study included.”

We are very sorry for our negligence of the criteria for the sample size. The criteria for sample size is mainly based on the following two aspects.

First, the inclusion criteria of the participants in this study is the population that has established health records and accepted chronic disease management in the community, so the sample size is determined according to the index of chronic disease health management rate. According to relevant literature, the management rate of type 2 diabetes in China in 2016 was 81.8%. The sample size was estimated by the health management rate of type 2 diabetes of 82%, According to the following

formula: where  $\alpha=0.05$ ,  $Z_{\alpha}=1.96$ ,  $d=0.05p$ , and the rate of discarded questionnaires was set 10%, the estimated sample size of patients with type 2 diabetes is 370.



In addition, some scholars believe that the estimation of the survey sample is based on the number of items in the scale, and each item contains about 5-10 subjects. In this study, there are 64 items in the health belief scale for type 2 diabetes patients, and the sample size of type 2 diabetes patients is 320~640.

Taking into account the differences in eastern, central and western regions in China, each region selected 420 patients, and a total of 1,260 type 2 diabetes patients were selected. A total of 36 villages or communities were randomly selected for this study, so 35 patients with type 2 diabetes were randomly selected from each village or community.

Page 6, line 10-14, "A total of 36 villages or communities were randomly selected for this study. Subsequently, considering the differences in eastern, central, and western regions of China, 420 patients were selected from each region, and a total of 1,260 patients were selected, therefore, 35 patients with type 2 diabetes were randomly selected from each village or community." were added.

6. **"The description of the measurement scales seems incorrect and are stated as a one-point scale. A one-point scale indicates that only one response option is available but it seems that participants select from a number of different responses (i.e. "1 (Agree) to 0.25 (Disagree) and 1 (Very helpful)." Please amend this description so it is clear the number of response options available for these questions."**

Thank you very much for your suggestion, we have made correction according to your comments. Page 5, line 4-7, "All items were measured on a 4-point scale, ranging from 1 (Agree) to 0.25 (Disagree) and 1 (Very helpful) to 0.25 (Hardly helpful); for the perceived barriers construct, the scores ranged from 1 (Very difficult) to 0.25 (Hardly difficult)." were corrected.

7. **"In the analysis section it is still not clear whether the authors adjusted for the complex sampling and survey methods in their analysis? Usually with such complex sampling and survey methods adjustments such as clustered standard errors, bootstrapping, Jackknife or weighting procedures etc. are used."**

Thank you very much for your valuable comments. Before this study, we conducted statistics and analysis on the data, and the analysis showed that there was no significant difference in sociodemographic characteristics of populations in different regions. So we didn't adjust for the complex sampling and survey methods in our study.

8. **"Please change the term multivariate to multivariable when referring to the regression modelling used throughout the manuscript."**

Thank you very much for your valuable comments. We have changed the term multivariate to multivariable when referring to the regression modelling used throughout the manuscript.

9. **"I appreciate that the authors have updated details on their regression model building. However, it would be good to include further detail including: specifically stating that**

**the enter method was used for model building and that the specific predictors selected for assessment were chosen as they are hypothesised to impact on self-care behaviors among diabetic patients.”**

Thank you very much for your suggestion, we have made correction according to your comments. Page 7, line 9-14, “According to the median, the total score of self-care behaviours was dichotomised based on a score of 3, considering self-care behaviour scores (  $\leq 3$  is 0,  $> 3$  is 1) as dependent variables, and sociodemographic characteristics and HBM constructs as independent variables. Furthermore, the enter method was used for regression model building, and multivariable logistic regression analysis was conducted to predict the variables influencing self-care behaviours.” were corrected.

**10. “Thank you for updating the regression model to ensure it is appropriate for the type of outcome variable. To further strengthen this could the authors please include a brief justification as to why scores of self-care behaviours were dichotomised based on a score of 3? Is this clinically relevant or based on previous literature for example?”**

Thank you very much for your suggestion, we are very sorry for our negligence and have made correction according to your comments. The total score of self-care behaviors is 7 points, ranging from 0 to 7 (median=3.5). On the basis of previous literature, according to the median, scores of self-care behaviors are dichotomised based on a score of 3, so the total score of self-care behaviors (scores of self-care behaviors  $\leq 3$  is 0,  $> 3$  is 1) is the dependent variable in the logistic regression.

Page 7, line 9-14, “According to the median, the total score of self-care behaviours was dichotomised based on a score of 3, considering self-care behaviour scores (  $\leq 3$  is 0,  $> 3$  is 1) as dependent variables, and sociodemographic characteristics and HBM constructs as independent variables. Furthermore, the enter method was used for regression model building, and multivariable logistic regression analysis was conducted to predict the variables influencing self-care behaviours.” were corrected.

**11. “Please ensure the details describing how missing data were addressed are included in the analysis section of the paper.”**

We are very sorry for our negligence of missing data, Page 7, line 3-6, “Missing values of input data were also processed; if key variables were missing or the data were missing up to 5%, they were treated as invalid questionnaires. The mode was used to fill in the data for variables with a small amount of missing data.” were added.

**12. “ In the methods section under the heading ‘study design’ I would include one to two sentences describing the design of the study, which is a cross-sectional survey study. For the description of the survey I would move this under a heading of ‘Study measures’ as it is distinct from the study design.”**

Thank you very much for your valuable comments, we have made correction according to your comments. Page 4, line 6, “In this cross-sectional survey study,” were added; page 5, line 11, a new

heading “Study measures” was added, for the description of the survey we had moved this under a heading of “Study measures”.

13. **“It is still not clear why differences in health belief scores by sociodemographic variables were assessed as the aims of the study are stated to be looking at association between health belief constructs and sociodemographic variables with self-care behaviour as the primary outcome. By looking at the univariate associations between the sociodemographic variables with the health belief score, the authors are now treating the health belief score as an outcome, rather than a predictor variable. However, there is no aim relating to this analysis. Please make sure that these analyses are tied to an aim, or removing these comparisons as it is not clear what question this analysis is relating to.”**

Thank you very much for your suggestion, it is really true as you suggested that it would be better to remove these comparisons. We have removed these comparisons according to the your comments, we have also made some adjustments to the table 2.

Page 7, line 18-21, “Univariate analysis showed that age, marital status, educational level, account nature, work status, household size and areas were negatively correlated with health belief level.” were deleted.

14. **“In table 4 it seems there are p-values listed for each level of a categorical variable (e.g. there is a p-value for each level of education status). Only an omnibus test should be used to assess the association between the independent variables and outcome and an overall p-value provided, rather than a p-value for each level of a categorical variable. Please remove the category level p-values and include only the overall p-value for each variable and ensure that the conclusions about what is statistically associated with the outcome is based on the overall p-value only.”**

Thank you very much for your valuable comments, we are very sorry for our unclear expression in table 4. In table 4, we set dummy variables for education level, health insurance and areas respectively, so there is a p-value for each level of education level, health insurance and areas.

15. **“ In the discussion the authors state that patients had poor self-care. How was the self-care score determined to be poor? i.e. what is considered low self-care? I am assuming it is the classification of 3 or less, adding a sentence as to how this dichotomy was determined relevant will help frame this interpretation.”**

We are very sorry for our incorrect writing, we have made correction according to the your comments. In this study, according to the median, scores of self-care behaviors are dichotomised based on a score of 3 (scores of self-care behaviors  $\leq 3$  is 0,  $> 3$  is 1), approximately 16% of participants scored less than 3, page 11, line 8-10, “The study findings indicated that some participants had poor self-care behaviors and health belief played a important role in performing self-care behaviors.” were corrected.

Page 7, line 9-14, “According to the median, the total score of self-care behaviours was dichotomised based on a score of 3, considering self-care behaviour scores (  $\leq 3$  is 0,  $> 3$  is 1) as dependent

variables, and sociodemographic characteristics and HBM constructs as independent variables. Furthermore, the enter method was used for regression model building, and multivariable logistic regression analysis was conducted to predict the variables influencing self-care behaviours.” were corrected.

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Hall, Alix Univ Newcastle
<b>REVIEW RETURNED</b>	31-Oct-2021
<b>GENERAL COMMENTS</b>	The authors have done a fantastic job in addressing the reviewer comments. The paper has been significantly improved. I only have one outstanding comment that is yet to be addressed. In the results of the logistic regression p-values for the overall association between a variable and the outcome of self-care behaviours an overall omnibus p-value between each independent variable (e.g. education) and the outcome are not displayed. Instead, individual p-values for each of the individual levels of the independent variable is provided (e.g. for education there is a p-value for each of the following categories: primary school, junior high school and high school and above). When assessing the association between the independent variables (e.g. education) and self-care behaviours only an overall p-value for the independent variable (e.g. education) should be provided, and not a p-value for each individual category level. The authors should update this and only provide p-values for the overall independent variables and a conclusion of a statistically significant relationship between a variable and the outcome should only be made if the overall omnibus p-value is less than 0.05 and not determined based on the p-value of the individual category levels of the independent variable.

### VERSION 3 – AUTHOR RESPONSE

#### Reviewer 2:

1. “The authors have done a fantastic job in addressing the reviewer comments. The paper has been significantly improved. I only have one outstanding comment that is yet to be addressed. In the results of the logistic regression p-values for the overall association between a variable and the outcome of self-care behaviours an overall omnibus p-value between each independent variable (e.g. education) and the outcome are not displayed. Instead, individual p-values for each of the individual levels of the independent variable is provided (e.g. for education there is a p-value for each of the following categories: primary school, junior high school and high school and above). When assessing the association between the independent variables (e.g. education) and self-care behaviours only an overall p-value for the independent variable (e.g. education) should be provided, and not a p-value for each individual category level. The authors should update this and only provide p-values for the overall independent variables and a conclusion of a statistically significant relationship between a variable and the outcome should only be made if the overall omnibus p-value is less than 0.05

**and not determined based on the p-value of the individual category levels of the independent variable.”**

Thank you very much for your valuable comments, we are very sorry for our incorrect writing, we have made correction according to your comments. Page 10, table 4 was corrected, we only provided overall p-values for the independent variables (e.g. education). According to the overall p-values for the independent variables, we have also revised the discussion, the revised parts in the manuscript were marked in yellow.