

PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components: Findings from the International Mobility in Aging Study
AUTHORS	Pirkle, Catherine; Wu, Yan Yan; Zunzunegui, Maria-Victoria; Gomez, Fernando

VERSION 1 – REVIEW

REVIEWER	GAUTAM KUMAR GINJUPALLI MARSHALL UNIVERSITY, WEST VIRGINIA, USA
REVIEW RETURNED	06-Aug-2017

GENERAL COMMENTS	<p>General Comments:</p> <ol style="list-style-type: none">1. This is a well-planned and well written manuscript with clear detail for every piece of information2. Authors mention the strengths and limitations of the study clearly3. The study has worth presenting useful in predicting the risk clusters, though it doesn't include other important clinical parameters pertaining to MetS such as insulin resistance, GTT and ITT.4. The technique appears to be appropriate for the study in predicting risk clusters efficiently <p>Specific Comments:</p> <ol style="list-style-type: none">1. It would have been better had the authors included some discussion about some of the social and economic factors such as low income, lack of education, childhood adversity could contribute to the MetS with relevant citations.2. It would have been added information had the authors cited some other similar studies on other diseases that used this technique of Model-based recursive partitioning and compared the results across studies
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Bing Zhang National Institute for Nutrition and Health, China CDC P. R. of China
REVIEW RETURNED	18-Oct-2017

GENERAL COMMENTS	What is the significant finding in this study? it is not so clear for me. I do not see significant advantage of the statistical method or model used in the present study than ordinary methods. The case was selected from baseline data other than follow-up data. It's evidence is a little bit weak in comparison with other studies.
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Sai Yi Pan Public Health Agency of Canada, Canada
REVIEW RETURNED	09-Nov-2017

GENERAL COMMENTS	<p>This manuscript applied the model-based recursive partitioning to identify risk clusters for metabolic syndrome in older adult population from several different countries. Overall, the research question is well defined, the data are sound, and the statistical analysis method is suitable. Below are some points for improving the manuscript:</p> <ol style="list-style-type: none"> 1. Introduction, 2nd paragraph: Since this study has two sites from Canada, a few references should be included on the MetS prevalence in Canada. For example, <ol style="list-style-type: none"> a. Natalie D. Riediger, and Ian Clara. Prevalence of metabolic syndrome in the Canadian adult population. CMAJ. 2011 Oct 18; 183(15): e1127–e1134. b. Solmaz Setayeshgar, Susan J. Whiting, and Hassanali Vatanparast. Metabolic Syndrome in Canadian Adults and Adolescents: Prevalence and Associated Dietary Intake. ISRN Obes. 2012; 2012: 816846. 2. Introduction, 2nd paragraph, “In older adults, MetS varies considerably across populations”. It would be better to specify that it is the “MetS prevalence”. 3. Introduction, 2nd paragraph: The sentence on the MetS prevalence in the United States, Europe, and China: it is better to specify that the prevalence refer to older adult population for clarity. 4. Statistical analysis: Since MOB is the main statistical method for this study; it would be preferable to have a more detailed description on how to perform MOB. 5. Page 14: “The highest values of MetS were observed in clusters of women form the middle-income study sites ...”. Does “values” mean “prevalence”? 6. There are many places that the authors use “greater” or “greatest” prevalence/concentration. It is better to use “higher” or “highest”. 7. Page 15, Discussion, last two sentence, (58-68%) and (26-41%): This gives the audiences an impression of a range of values, however, the results only showed two values: 58% and 68% or 26% and 41%. Please consider revising it. 8. Page 15, 2nd paragraph, “This study corroborates previous findings that the prevalence of MetS varies according to age, sex and socioeconomic status”. This study did not show the prevalence of MetS by age although the MOB was adjusted for age. Therefore, the author cannot state that this study corroborates previous findings that the prevalence of MetS varies according to age. 9. Page 15, 2nd paragraph, “Consistent with other studies, we observed a concentration of MetS in participants of lower socioeconomic status”. Although this statement is appropriate for the overall results, but this statement does not hold true for women from the middle-income sites where women with greater educational attainment had a greater predicated prevalence of MetS. Please revise the sentence. 10. Page 17, 1st paragraph, last sentence: Please explain the inconsistent results between men and women regarding on education level and MetS prevalence. 11. Page 18, 1st paragraph, last sentence: It would be clearer to add that “Our study showed” or similar phrases. 12. Given the results suggests that the main partitioning variables are study site and sex, which are not modifiable, could the author add comments on the clinical and public health implication and importance of this study.
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: GAUTAM KUMAR GINJUPALLI

Institution and Country: MARSHALL UNIVERSITY, WEST VIRGINIA, USA

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

General Comments:

1. This is a well-planned and well written manuscript with clear detail for every piece of information

Response: Thank you.

2. Authors mention the strengths and limitations of the study clearly

Response: Thank you.

3. The study has worth presenting useful in predicting the risk clusters, though it doesn't include other important clinical parameters pertaining to MetS such as insulin resistance, GTT and ITT.

For this study, we use the Adult Panel Treatment III (ATP III) criteria to define MetS. Please see: National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106:3143-421. This is reference 21 in our paper. This is a widely accepted definition for MetS and broadly applied in observational epidemiological research including in representative national surveys such as NHANES in the United States and the Canadian Health Measures Survey in Canada.

While there are multiple definitions of MetS existant, only one of the commonly applied definition includes measures of insulin resistance (the 2003 American Association of Clinical Endocrinologists definition). Please see Samson SL, Garber AJ. Metabolic syndrome. *Endocrinol Metab Clin North Am* 2014;43:1-23. This is reference 4 in our paper.

This is likely because obesity and insulin resistance are believe to be at the core of most cases of MetS. Of note, MetS in of itself is not a disease per se, but rather, a combination of traits that are associated with greatly increased risk of diabetes and cardiovascular disease.

Finally, it should be noted that the insulin tolerance test (ITT) is potentially hazardous to patients and typically requires close medical supervision. It also requires several hours to complete. In the context of a large survey, use of the ITT is logistically impossible. To conclude, we believe our measure of MetS is consistent with the broader literature, comparable to other studies, and overall defensible for the purposes of this study, which were to apply recursive partitioning to investigate social and behaviour risk factors for the MetS.

4. The technique appears to be appropriate for the study in predicting risk clusters efficiently

Response: Thank you.

Specific Comments:

1. It would have been better had the authors included some discussion about some of the social and economic factors such as low income, lack of education, childhood adversity could contribute to the MetS with relevant citations.

Response: While the second and third paragraphs of the discussion in the previous version did discuss other appropriately-cited research relating low income, lack of education and childhood adversity to MetS, including covering which research supports and does not support our observations, we have elaborated these paragraphs some more in response to the reviewer's comment. Please see our revisions to page 17 (paragraph 2) and page 19 (first paragraph).

2. It would have been added information had the authors cited some other similar studies on other diseases that used this technique of Model-based recursive partitioning and compared the results across studies

Response: Thank you for bringing this point up. There is actually very little work that uses Model-based recursive (MOB) partitioning to examine social and behavioral risk clusters of diseases. Most applications of this technique focus on clinical risk factors and/or components of syndromes. This is a strength of our study in that it applies MOB in a relatively novel manner. We explain this on page 5 (end of page and first lines of page 6), and again on page 20, paragraph 1.

Reviewer: 2

Reviewer Name: Bing Zhang

Institution and Country: National Institute for Nutrition and Health, China CDC, P. R. of China

Competing Interests: none

What is the significant finding in this study? it is not so clear for me. I do not see significant advantage of the statistical method or model used in the present study than ordinary methods. The case was selected from baseline data other than follow-up data. It's evidence is a little bit weak in comparison with other studies.

Response: Model based recursive partitioning (MOB) method is the integration of the traditional classification and regression tree (CART) method, and regression models. The advantages are:
1. MOB trees have regression models (i.e., logistic regression for binary outcomes or linear regression for continuous outcomes, etc.) with one or more predictor (or confounding) variables controlled in each step of the partitioning. The terminal nodes have fitted models with the predictor variables, such as age. CART only provides mean estimates in each terminal node and cannot adjust for predictor variables.

2. Since MOB utilizes regression models and significance tests, it solves the overfitting problem in CART methods (CART requires cross validation, pruning or random forest method to validate the tree). Most health professionals are familiar with significance tests, which assists in the comprehensibility of the results.

3. MOB can identify complex higher-order interactions without requiring large sample size. To fit higher-order interaction effects, large sample sizes are needed for traditional regression models (e.g. "ordinary methods"). Most of the time it is unrealistic to achieve the required sample size and higher order-interactions can be difficult to interpret and report.

4. MOB provides intuitive graphical presentation of risk factors that are easy to understand for clinicians and other health practitioners.

With regards to the reviewer's comment, "The case was selected from baseline data other than follow-up data. It's evidence is a little bit weak in comparison with other studies," since our objective was to apply model based recursive partitioning to identify risk clustering across populations, use of baseline data is appropriate, as we are not looking at incident predictors of the syndrome. However, in a next step, it could be very interesting to examine if the risk clusters identified in this research predicted onset MetS more successfully than each individual risk factor; however, to move on to this next step, we first have to identify the risk clusters to test in the future. As for comparisons with other studies, there are very limited other studies with which to compare this work. As stated in the last line of the introduction, we know of no other studies applying recursive partitioning technique to investigate MetS that are informed by a social epidemiological perspective.

As we state to reviewer 1, this is a strength of our study, in that it consists of a relatively novel application of the recursive partitioning. We have made this more explicit on page 20, first paragraph.

Reviewer: 3

Reviewer Name: Sai Yi Pan

Institution and Country: Public Health Agency of Canada, Canada

Competing Interests: None declared

This manuscript applied the model-based recursive partitioning to identify risk clusters for metabolic syndrome in older adult population from several different countries. Overall, the research question is well defined, the data are sound, and the statistical analysis method is suitable.

Response: Thank you very much.

Below are some points for improving the manuscript:

1. Introduction, 2nd paragraph: Since this study has two sites from Canada, a few references should be included on the MetS prevalence in Canada. For example,

a. Natalie D. Riediger, and Ian Clara. Prevalence of metabolic syndrome in the Canadian adult population. *CMAJ*. 2011 Oct 18; 183(15): e1127–e1134.

b. Solmaz Setayeshgar, Susan J. Whiting, and Hassanali Vatanparast. Metabolic Syndrome in Canadian Adults and Adolescents: Prevalence and Associated Dietary Intake. *ISRN Obes*. 2012; 2012: 816846.

Response: Thank you for pointing this out. We have included Canadian prevalence data on older adults. Please see pages 4 (bottom of page) and 5 (top of page) for the improvements.

2. Introduction, 2nd paragraph, “In older adults, MetS varies considerably across populations”. It would be better to specify that it is the “MetS prevalence”.

Response: Yes, thank you. This has been clarified.

3. Introduction, 2nd paragraph: The sentence on the MetS prevalence in the United States, Europe, and China: it is better to specify that the prevalence refer to older adult population for clarity.

Response: Yes, thank you. This has been clarified.

4. Statistical analysis: Since MOB is the main statistical method for this study; it would be preferable to have a more detailed description on how to perform MOB.

Response. Thank you for this suggestion. We modified the statistical analysis section and provided a step-by-step explanation of MOB method (see page 10):

For instance, age is controlled in the MOB analysis of MetS using logistic regression models and the MOB algorithm cycles iteratively through the following steps: (1) fit the logistic regression with MetS as response variable and age as control variable, (2) test for parameter instability over a set of partitioning variables (socioeconomic and demographic characteristics) while controlling for age, (3) if there is some overall parameter instability, split the data set with respect to the variable associated with the highest instability (i.e. the smallest p-value), (4) repeat the procedure in each of the resulting subsamples with different risk of MetS. The process is termed recursive because each sub-population may be split a number of times until a particular stopping criterion is reached. Our stopping criteria were: 5% level of significance and minimum sample size of 100 at terminal nodes. For continuous partitioning variables, MOB tests and selects an optimal cut-off point and split subjects into two subgroups.

5. Page 14: “The highest values of MetS were observed in clusters of women from the middle-income study sites ...”. Does “values” mean “prevalence”?

Response: Estimate is probably more accurate than value. The sentence now reads, “The highest estimates of MetS prevalence were observed in clusters of women from the middle-income study sites...”

6. There are many places that the authors use “greater” or “greatest” prevalence/concentration. It is better to use “higher” or “highest”.

Response: Okay.

7. Page 15, Discussion, last two sentence, (58-68%) and (26-41%): This gives the audiences an impression of a range of values, however, the results only showed two values: 58% and 68% or 26% and 41%. Please consider revising it.

Response: Okay, it is now revised to read, "in clusters of women from middle-income sites, the predicted proportion with MetS was quite high (58 or 68% depending on the cluster). In clusters of men, the predicted proportion with MetS was lower (26, 38 or 41% depending on the cluster) and highest among men reporting childhood social adversities (41%)."

8. Page 15, 2nd paragraph, "This study corroborates previous findings that the prevalence of MetS varies according to age, sex and socioeconomic status". This study did not show the prevalence of MetS by age although the MOB was adjusted for age. Therefore, the author cannot state that this study corroborates previous findings that the prevalence of MetS varies according to age.

Response: The results actually do show variation by participant age. If one refers to Figure 1, which depicts the partitioning results for MetS, below each node is a graph that shows the estimated MetS prevalence by age. For example, if you examine node 7 you can see a notable increase in estimated MetS prevalence by increasing age. While not presented, as our objective was to demonstrate the risk clustering by behavioral and social variables, increasing age is actually statistically associated with increasing MetS prevalence in node 7 (p-value 0.008), which is consistent with the figure. On page 16, beginning of the page, we state this relationship in a more explicit way.

9. Page 15, 2nd paragraph, "Consistent with other studies, we observed a concentration of MetS in participants of lower socioeconomic status". Although this statement is appropriate for the overall results, but this statement does not hold true for women from the middle-income sites where women with greater educational attainment had a greater predicated prevalence of MetS. Please revise the sentence.

Response: Okay, the sentence now reads, "Consistent with other studies, overall, we observed a concentration of MetS in participants of lower socioeconomic status[27-29]; although, among women from the middle-income sites, MetS was more prevalent among women with post-secondary education."

10. Page 17, 1st paragraph, last sentence: Please explain the inconsistent results between men and women regarding on education level and MetS prevalence.

Response: Okay, this has been done. Please see page 17, paragraph 2.

11. Page 18, 1st paragraph, last sentence: It would be clearer to add that "Our study showed" or similar phrases.

Response: Yes, we agree. This has been fixed.

12. Given the results suggests that the main partitioning variables are study site and sex, which are not modifiable, could the author add comments on the clinical and public health implication and importance of this study.

Response: Thank you for pointing this out. We have improved the conclusion section of this paper to address your comment about the clinical and public health implications of this study.

For example, our last line reads, "Finally, with regards to both clinical practice and health promotion activities, identifying risk clusters is important for targeting purposes, as the intensity and type of programs may differ according to sub-groups". We also highlight in this section that while study site and sex are not easily modifiable for most people, the polices that affect peoples health at certain settings and/or the opportunities afforded to men versus women can be modified.