PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validation of a non-invasive method for the early detection of metabolic syndrome: a diagnostic accuracy test in a working population
AUTHORS	ROMERO-SALDAÑA, MANUEL; Tauler, Pedro; Vaquero-Abellán, Manuel; López-González, Angel-Arturo; Fuentes-Jiménez, Francisco-José; Aguiló, Antoni; Álvarez-Fernández, Carlos; Molina-Recio, Guillermo; Bennasar-Veny, Miquel

VERSION 1 – REVIEW

REVIEWER	Dr Georgia Torres Center for exercise science and sport medicine, University of the Witwatersrand, South Africa
REVIEW RETURNED	06-Dec-2017

GENERAL COMMENTS	Some grammar errors in the abstract Could we see how many developed MetS over the study years ie. did not have MetS at onset of study and then were diagnosed with MetS at the end 2016, and compare the 2 methods in been able to predict who actually developed MetS?

REVIEWER	Wei-Fen Ma Professor, School of Nursing, China Medical University, Taiwan
REVIEW RETURNED	03-Apr-2018

GENERAL COMMENTS	1.The aim of this study is to reproduce and validate the NIM-Mets method in a large representative sample of Spanish workers. Overall, this paper proved clear results and conclusion. Meanwhile, the methods are easy and convenient for the early detection of MetS in a healthy worker population. 2. Adequate sample size and analysis were presented in this paper. The definition of MetS, measure concepts are easy to understand. 3. The blood samples were taken after a 12 h overnight fast making the study valid. 4. The only issue that need to be concerned is about limitation of the study. Although non-invasive test which is easy to apply and interpret in any health care setting, some limitations may need to be addressed in this paper.

REVIEWER	Claudio Pedone Università Campus Bio-Medico di Roma
REVIEW RETURNED	24-Apr-2018

GENERAL COMMENTS	The description of the methods is confusing. The authors state that this is a "double epidemiological study", with a cross-sectional part

	carried out from 2011 to 2015 on 60,799 participants and a "diagnostic test" part carried out in 2016 on 50,687 participants. It is not clearly stated what analyses have been carried out in the second sample, as the N in all tables reported in the results is 60,799. While the individual statistical methods used are formally correct, the overall approach is hard to understand. The authors do not follow the classic training/testing population approach to validate their index, and the terminology used ("validity" vs. "reproducibility") is confusing. If by "validity" they refer to a validation study (as it seems to be the case reading the Results), then a comparison with the results of the original study should be reported, and also a formal assessment of calibration, bias etc. should be performed. In diagnostic testing, "reproducibility" usually refers to the agreement between results of tests repeated over time or by different raters. In this study, by "reproducibility" is used to indicate the agreement of the results obtained by applying the original methodology in this new population, but this is not clearly explained in the methods and the reader needs to figure out by reading the results. The underlying rationale (not clearly stated in the paper) is that an early detection of metabolic syndrome (MetS) can have a public health impact. The new method proposed (NIM-MetS) is based on one of the components of MetS (blood pressure) plus a measure of adiposity (waist to heigth ratio, WHtR) that is strongly correlated with another component of MetS (waist circumference). This notwithstanding, this method is not very useful for early detection of MetS, as it only has a positive predictive value of about 0.5 (not better than a coin toss). In the face of this, it may be rationale to seek a different cut-off for WHtR, and the authors find a cut-off that is slightly different from the original one. With this new cut-off, sensitivity and specificity of the new NIM-MetS are 0.564 and 0.945, respectively vs. 0.547 and 0.949 of the "old" NIM-MetS: given these very small differences, I would not claim that "through the present study, the NIM-MetS has been corrected" .

REVIEWER	Jing Kang School of Dentistry, Faculty of Medicine and Health, University of Leeds, Leeds, UK
REVIEW RETURNED	30-Apr-2018

GENERAL COMMENTS	This study proposed to use WHtR and BP (non-invasive method) as indicator to diagnose Metabolic syndrome, and assessed several indexes. The authors determined the cut-off values for WHtR (>0.558) and BP (128/80mmHg) based on their ROC curves, and concluded this non-invasive method is of high specificity and validity index. I have some general questions: 1. Why the authors do not use multiple logistic regression model with the two anthropometric variables (WHtR and BP) as predictors and estimate the risk of MetS? What is the advantage of using CHAID rather than logistic regression (all variables are binary in this large population study so you don't have the problem of data distribution)? 2. Number of decimal places is inconsistent throughout the paper, from abstract to results to table presentation. I understand the authors wanted to show more digits when the value ranges are very close, but this makes the paper look unprofessional and i wonder if it is necessary to display so many digits? e.g. how much difference

	does it make to clinicians or individuals when WHtR >0.558 than WHtR >0.56? Does the result (sensitivity, specificity, AUC and etc) change much when you use cut-off point of WHtR as 0.558 or 0.56 or even 0.6, when you report everything with 1 decimal place? 3. Which index is more important, sensitivity or specificity, clinically? This paper showed a very high specificity and moderate sensitivity using the non-invasive method, but I think the authors should discuss more about the clinical importance why a sensitivity of 59% is acceptable using the non-invasive method while one more step of blood test will provide accurate diagnosis. My comments regarding statistics and data presentation: 1. The sensitivity and specificity values reported in abstract was not consistent with table 4. 2. The authors reported validity index without defining how this index is calculated--is it the same as the diagnostic validity in the abstract? My understanding of diagnostic validity indexes include sensitivity, specificity, PPV, NPV, etc, so please define how is validity index calculated in this paper. 3. Page 6 line28-33: what does "accuracy was 0.23%" mean? what does "security rate of 95%" mean? 4. Page 9 Statistical Analysis section: the first three paragraphs do not make sense. Please re-write. (what is "95% mean"? why use Kolmogorove-Smirnov test to assess goodness-of-fit? What does the second paragraph mean? How come t test is bi-variate analysis? Why the authors stated so many tests for different data type while the only outcome variable is MetS diagnose result (yes/no). 5. Page 11 Result section: In the first paragraph, %BF and HDL-cholesterol interpretation in the text is different from the Table 1 presentation (text stated these two variables are not significantly different between women and men but table 1 indicate p<0.001 for these two variables). 6. Table presentation problem (for all table1-4): no definition of the values in the brackets (is it 95% CI??); display number of digit in a consistent format (e.g. all 2 decimal places) 7. Figure 2: no horizontal axis title was presented, and the intersection value of WHtR is confusing--is the vertical axis WHtR or is the horizontal axis WHtR? The axes look wrong!

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Dr Georgia Torres
Institution and Country: Center for exercise science and sport medicine, University of the Witwatersrand, South Africa
Please state any competing interests or state ‘None declared’: None declared
Please leave your comments for the authors below

- Some grammar errors in the abstract

R: English grammar in the abstract and in the entire manuscript has been revised.

- Could we see how many developed MetS over the study years ie. did not have MetS at onset of study and then were diagnosed with MetS at the end 2016, and compare the 2 methods in been able to predict who actually developed MetS?
R: Due to this and other ones from reviewers we have decided to rewrite the design and sample section in order to clarify points like this one. Due to the large sample size of the study (n=60,799), this was developed between 2012 and 2016 (the authors needed 4 years to reach this sample size). However, we would like to clarify that the study design is cross sectional and it is not longitudinal (or an incidence study), with an only sample. And no follow-up of the participants was performed. Therefore, MetS cases are prevalent and not incident. In this sense, prevalence of MetS obtained was low, about 9%.

Reviewer: 2
Reviewer Name: Wei-Fen Ma
Institution and Country: Professor, School of Nursing, China Medical University, Taiwan
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below
1. The aim of this study is to reproduce and validate the NIM-Mets method in a large representative sample of Spanish workers. Overall, this paper proved clear results and conclusion. Meanwhile, the methods are easy and convenient for the early detection of MetS in a healthy worker population.
2. Adequate sample size and analysis were presented in this paper. The definition of MetS, measure concepts are easy to understand.
3. The blood samples were taken after a 12 h overnight fast making the study valid.
4. The only issue that need to be concerned is about limitation of the study. Although non-invasive test which is easy to apply and interpret in any health care setting, some limitations may need to be addressed in this paper.
R: We are very grateful to the reviewer for the comments and the encouragement. Limitations of the study have been considered within the Strengths and Limitations section and introduced and discussed in the Discussion section.

Reviewer: 3
Reviewer Name: Claudio Pedone
Institution and Country: Università Campus Bio-Medico di Roma
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below
The description of the methods is confusing. The authors state that this is a "double epidemiological study", with a cross-sectional part carried out from 2011 to 2015 on 60,799 participants and a "diagnostic test" part carried out in 2016 on 50,687 participants. It is not clearly stated what analyses have been carried out in the second sample, as the N in all tables reported in the results is 60,799.
R: We are grateful to the reviewer for this comment. The authors have identified that this is a critical point of the article that is not clear. Actually, the "design and sample" section has been rewritten. The present investigation consists of two studies, both performed in the same sample of workers (n=60,799), collected between 2012 and 2016. First, a cross-sectional study was carried out to reproduce NIM-MetS. And second, a study of diagnostic test was carried out to validate NIM-MetS. The reviewer indicates that a sample (n=50,687) has been used to carry out the study of diagnostic test. However, 50,687 is the theoretical (minimum value needed) sample size obtained for the study of the diagnostic test and not the real sample with which the study was made. 60,799 participants was the sample size used for both studies.

While the individual statistical methods used are formally correct, the overall approach is hard to understand. The authors do not follow the classic training/testing population approach to validate their

index, and the terminology used ("validity" vs. "reproducibility") is confusing. If by "validity" they refer to a validation study (as it seems to be the case reading the Results), then a comparison with the results of the original study should be reported, and also a formal assessment of calibration, bias etc. should be performed.
R: We would like to especially thank the reviewer for this comment as it has helped us to find out that the description of the design was not clear at all and it didn't describe what had actually been done. Furthermore, we had been using confusing terms such as validity and reproducibility when what we had actually done is a study of validation. Therefore, the manuscript has been restructured and rewritten for clarification. We hope that the reviewer find an improved and proper design description and utilization of the terms.
In the original version of the manuscript terms validity and reproducibility were referred to:
-Reproducibility: the ability of a test to be reproduced by other researchers. NIM-MetS (non-invasive method for the early detection of Metabolic Syndrome) was originally obtained in a particular worker population (Córdoba, 2015). The objective of the present study was to evaluate the reproducibility of NIM-Met in a different sample (n = 60,799) of Spanish workers (Balearic Islands, 2016).
-Validity: the validity of a test represents the degree to which an instrument (method, test, etc.) measures what it really intends to measure. To obtain this information, results of the test are compared with a gold standard test. In the present study, results of NIM-Mets test are compared with results of the NCEP-ATPIII criteria (gold standard).

In diagnostic testing, "reproducibility" usually refers to the agreement between results of tests repeated over time or by different raters. In this study, by "reproducibility" is used to indicate the agreement of the results obtained by applying the original methodology in this new population, but this is not clearly explained in the methods and the reader needs to figure out by reading the results.
R: As we have explained in the answer to the previous comment, the information regarding this point has been rewritten and corrected in order to increase clarity.

The underlying rationale (not clearly stated in the paper) is that an early detection of metabolic syndrome (MetS) can have a public health impact. The new method proposed (NIM-MetS) is based on one of the components of MetS (blood pressure) plus a measure of adiposity (waist to heigth ratio, WHtR) that is strongly correlated with another component of MetS (waist circumference). This notwithstanding, this method is not very useful for early detection of MetS, as it only has a positive predictive value of about 0.5 (not better than a coin toss). In the face of this, it may be rationale to seek a different cut-off for WHtR, and the authors find a cut-off that is slightly different from the original one. With this new cut-off, sensitivity and specificity of the new NIM-MetS are 0.564 and 0.945, respectively vs. 0.547 and 0.949 of the "old" NIM-MetS: given these very small differences, I would not claim that "through the present study, the NIM-MetS has been corrected" .
R: The authors appreciate the comment of the reviewer, but they would like to clarify the following points:
- NIM-MetS is a non-invasive method for the early detection of SMet, with a very high validity index (91.2%). But the greatest advantage of NIM-MetS is that it does not require any blood variable (non-invasive), thus becoming a method with a great applicability in Public Health clinical practice.
- In spite of the predictive positive value is 51.2%, specificity is 94,9% and the diagnostic validity index (percentage of subjects properly classified) is 91,2%. These indicators establish the validity and usefulness of the NIM-MetS in the SMet screening.
- The validation of the NIM-MetS in a large sample (n = 60,799) has shown that the cut-off values of WHtR are very similar to the ones of the original method. This indicates that the NIM-MetS is a method with an excellent reproducibility. This observation has been more clearly explained in the revised discussion.

Reviewer: 4

Reviewer Name: Jing Kang
Institution and Country: School of Dentistry, Faculty of Medicine and Health, University of Leeds, Leeds, UK
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

This study proposed to use WHtR and BP (non-invasive method) as indicator to diagnose Metabolic syndrome, and assessed several indexes. The authors determined the cut-off values for WHtR (>0.558) and BP (128/80mmHg) based on their ROC curves, and concluded this non-invasive method is of high specificity and validity index.
I have some general questions:

1. Why the authors do not use multiple logistic regression model with the two anthropometric variables (WHtR and BP) as predictors and estimate the risk of MetS?
R: The authors compared the predictive capacity of the logistic regression model with the two variables (WHtR and BP) and the clinical decision tree obtained using the CHAID method (also with WHtR and BP as considered variables). The results were the following: sensitivity in the logistic regression model was 28.3%, while the sensitivity when the clinical decision tree was used was 56.4%. Therefore, authors decided to use the clinical decision tree model.

What is the advantage of using CHAID rather than logistic regression (all variables are binary in this large population study, so you don't have the problem of data distribution)?
R: Authors think that the following are the two main advantages:
1) In clinical practice it is easier to use an algorithm (decision tree or clinical classification) than a multiple logistic regression model.
2) The sensitivity obtained with the clinical decision tree (CHAID method) is higher than the one observed when the logistic regression model was applied.

2. Number of decimal places is inconsistent throughout the paper, from abstract to results to table presentation. I understand the authors wanted to show more digits when the value ranges are very close, but this makes the paper look unprofessional and i wonder if it is necessary to display so many digits? e.g. how much difference does it make to clinicians or individuals when WHtR >0.558 than WHtR >0.56? Does the result (sensitivity, specificity, AUC and etc) change much when you use cut-off point of WHtR as 0.558 or 0.56 or even 0.6, when you report everything with 1 decimal place?
R: Authors are in agreement with this comment. A consistent format for decimal values has been introduced, with no more than 2 decimal places.

3. Which index is more important, sensitivity or specificity, clinically? This paper showed a very high specificity and moderate sensitivity using the non-invasive method, but I think the authors should discuss more about the clinical importance why a sensitivity of 59% is acceptable using the non-invasive method while one more step of blood test will provide accurate diagnosis.
R: Lo ideal es que la pruebas tengan una sensibilidad y especificidad del 100%, pero en la práctica esto no suele ocurrir. Depende de lo que deseemos nos interesa una alta especificidad o una alta sensibilidad, por ejemplo, en el caso de las enfermedades infecciosas nos conviene que tengan una alta sensibilidad. Ahora bien el caso que nos ocupa es una enfermedad crónica y lo que garantizamos con una alta especificidad es que si el sujeto esta sano será clasificado adecuadamente. Ello, no provocará que no se trate como enfermos a personas que realmente están sanas. En general, las pruebas confirmatorias del diagnóstico deben ser de alta especificidad, para evitar falsos positivos.
Depending on which is the main aim a high specificity or a high sensitivity could be preferred. For example, in the case of infectious diseases we should have high sensitivity. But in the method considered in the present study, which is related to a chronic disease, a high specificity ensures that if

the subject is healthy it will be classified properly. This will not cause that people who are really healthy are treated as sick people. In general, the confirmatory tests of the diagnosis must be of high specificity, to prevent false positives.
Some lines in the discussion section have been included regarding the moderate sensitivity and the general characteristics of the method (page 13, last paragraph). In this sense, we feel that the “next step” indicated by the reviewer is a very significant step mainly when large populations, such as in the public health field, populations with a lack of resources, or undeveloped and/or remote places are considered. Thus, taking these observations, and some more indicated in the manuscript, into account we think that the method could be acceptable in spite of the moderate sensibility.

My comments regarding statistics and data presentation:
1. The sensitivity and specificity values reported in abstract was not consistent with table 4.
R: Sensibility value reported in the abstract has been corrected and now it is consistent with values shown in table 2, which are the ones characterising the method. On the other hand, values shown in table 4 are referred to the nodes in the decision trees.

2. The authors reported validity index without defining how this index is calculated--is it the same as the diagnostic validity in the abstract? My understanding of diagnostic validity indexes include sensitivity, specificity, PPV, NPV, etc, so please define how is validity index calculated in this paper.
R: The authors are in agreement with this comment. A description of the validity index determination has been included in the manuscript (Page 9, Statistical analysis section). The validity index is also known as the diagnostic validity index and represents the percentage of subjects properly classified by the test.

3. Page 6 line28-33: what does "accuracy was 0.23%" mean? what does "security rate of 95%" mean?
R: Thus, "accuracy was 0.23%" means that the precision of the cross-sectional study was 0.23% with the 60,799-worker sample size, an expected prevalence of SMet of 10% and a 95% confidence level. The lower this value, the smaller the sampling error and the higher the accuracy of the study results. The term "security rate" should be changed to "95% confidence level". This modification has been included in the revised version of the manuscript (page 5, Design section).

4. Page 9 Statistical Analysis section: the first three paragraphs do not make sense. Please re-write.
R: We are in agreement with the reviewer. Statistical analysis section has been rewritten in order to increase its clarity.

(what is "95% mean"?
R: This was a mistake, and it has been corrected in the revised version of the manuscript. It should be related to the interval confidence rather than to the mean.

why use Kolmogorove-Smirnov test to assess goodness-of-fit?
R: The authors consider that the Kolmogorov-Smirnov test is the most appropriate for checking the normal distribution of this large samples (n>50) (Frank J. Massey Jr. (2012) The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, 46:253, 68-78, DOI: 10.1080/01621459.1951.10500769).

What does the second paragraph mean?
R: This paragraph has been deleted as the essential information included is yet shown in the “Data collection and definition of variables” (Prevalence of MetS was determined). The other information included was not relevant because it was not included in the final version of the manuscript.

How come t test is bi-variate analysis? Why the authors stated so many tests for different data type while the only outcome variable is MetS diagnose result (yes/no).
R: This has been modified, corrected and clarified in the rewritten Statistical Analysis section.

5. Page 11 Result section: In the first paragraph, %BF and HDL-cholesterol interpretation in the text is different from the Table 1 presentation (text stated these two variables are not significantly different between women and men but table 1 indicate p<0.001 for these two variables).
R: This has been corrected in the results section. Briefly, among blood and anthropometrical parameters shown in Table 1, women showed significant higher values only for %BF and HDL-Cholesterol. Men showed significant higher values for the rest of parameters included in the table.

6. Table presentation problem (for all table1-4): no definition of the values in the brackets (is it 95% CI??); display number of digit in a consistent format (e.g. all 2 decimal places)
R: This information has been clarified and the text 95% CI has been included in tables using these values. A consistent format for decimal values has been used, with no more than 2 decimal places.

7. Figure 2: no horizontal axis title was presented, and the intersection value of WHtR is confusing--is the vertical axis WHtR or is the horizontal axis WHtR? The axes look wrong!
R: We would like to thank the reviewer for this particular comment. The vertical axis tittle has been corrected and the horizontal one has been included.

VERSION 2 – REVIEW

REVIEWER	Wei-Fen Ma China Medical University, Taiwan
REVIEW RETURNED	22-Jun-2018

GENERAL COMMENTS	Adequate limitations are discussed in the revised paper. If update references can be made, the paper will be better.

REVIEWER	Claudio Pedone Università Campus Bio-Medico di Roma
REVIEW RETURNED	13-Jun-2018

GENERAL COMMENTS	The authors have addressed all my queries. I have no further comments.

REVIEWER	Jing Kang University of Leeds, UK
REVIEW RETURNED	26-Jun-2018

GENERAL COMMENTS	This manuscript improve significantly after revision. I am happy with the majority of the responses and the revised manuscript regarding the statistical issues. Just one minor comment on the decimal places: the digit format is still not consistent throughout the paper, and please avoid displaying some number as whole number, some with 1 digit, some with 2 digits. E.g. if the whole manuscript display two decimal places, write number 8 as 8.00.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer's and Editorial Comments

Reviewer: 3

Please leave your comments for the authors below: The authors have addressed all my queries. I have no further comments.

R: Thank you very much for your comments.

Reviewer: 2

Please leave your comments for the authors below: Adequate limitations are discussed in the revised paper. If update references can be made, the paper will be better.

R: Thank you ever so much for your comments. Some references (9, 19 and 21) have been updated.

Reviewer: 4

Please leave your comments for the authors below: This manuscript improve significantly after revision. I am happy with the majority of the responses and the revised manuscript regarding the statistical issues.

Just one minor comment on the decimal places: the digit format is still not consistent throughout the paper, and please avoid displaying some number as whole number, some with 1 digit, some with 2 digits. E.g. if the whole manuscript display two decimal places, write number 8 as 8.00.

R: We are very grateful to the reviewer for the comments and the encouragement.

Authors are partially in agreement with this comment. A more consistent digit format has been introduced. In the revised version of the manuscript no number is expressed as a whole number. However, we consider that using the same format for all the parameters considered in the manuscript is not the most adequate. In this sense, absolute values for parameters are shown using 2 decimal places as it is commonly done, percentages are shown with 1 decimal, and only values obtained after applying formulas (such as the ABSI and the cut-off points) are shown using three decimal places.