Author’s response to reviews

Title: Choosing the most appropriate existing type 2 diabetes risk assessment tool for use in the Philippines: A case-control study with an urban Filipino population

Authors:

Gina Agarwal (gina.agarwal@gmail.com)
Monserrat Guingona (monsgona@gmail.com)
Jessica Gaber (jgaber@mcmaster.ca)
Ricardo Angeles (angelesric@gmail.com)
Suhasini Rao (raosuh@gmail.com)
Fortunato Cristobal (khryss2448@yahoo.com)

Version: 2 Date: 03 Jul 2019

Author’s response to reviews:

June 26, 2019

RE: PUBH-D-19-00354R1

We appreciate the time spent on our manuscript "Choosing the most appropriate existing type 2 diabetes risk assessment tool for use in the Philippines: A case-control study with an urban Filipino population" to ensure our submission fits all necessary criteria. We have responded to the points raised by reviewers below. Changes to the manuscript are included in the table below, and indicated in the manuscript text through tracked changes.
Reviewer Comment: GENERAL COMMENTS: This is an interesting paper comparing the validity and accuracy of diabetes risk scores for a Filipino population. It uses ROC and sensitivity and specificity analysis as the means of demonstrating appropriateness. This is a nice paper but I don't think the analysis is sufficiently comprehensive. Suggestions are provided below. It is generally well written but would benefit from a native English speaker reading it. Somewhat surprisingly, the FINDRISC score performance the best on the Filipino population, in terms of specificity and sensitivity and discrimination.

Revision/Response: Thank you for your input. We have addressed the feedback around our analysis in the boxes below. The paper had already been partially written and edited by the authors who are native English speakers, but those authors have also reviewed again for any final language changes

Location: N/A – any minor changes to the language are tracked in the manuscript.

Reviewer Comment: There are literally hundreds of diabetes risk scores published, many on Asian populations. The authors choose 6 scores to compare in terms of sensitivity and specificity. There is no rationale given for their choice. I would have thought a score from an Asian population would have been more appropriate. A rationale for choosing these score should have been given.

Revision/Response: We inserted a rationale in selecting the 6 tools in our introduction. This was our rationale: “These tools were selected based on one or more of the following criteria: (1) tested for validity to predict diabetes; (2) widely used clinically; (3) tested in an Asian population (preferably in a Filipino population). Though previous literature suggests that similarity in ethnicity does not necessarily affect the accuracy of the tools [Tanamas SK, et al], using a similar methodology of applying these tools and comparing their results can elucidate which one will perform better in this population.”

Location: Background section, page 5, lines 109-114
Reviewer Comment: There is also a paper published which has done a similar task testing a larger range of score but comes to the conclusion that differences in diabetes scores and their application rest on the score themselves. I think the authors need to read this paper. Tanamas SK, et al. The performance of diabetes risk prediction models in new populations: the role of ethnicity of the development cohort. Acta Diabetol 2015;52(1):91-101.

Revision/Response: Thank you for pointing us to this paper. We operated on the assumption that there is a possibility that tools that were developed in an ethnically similar population might have an advantage (Indonesian, Filipino, Indian tools). However, we also considered the possibility that the more robustly tested tools would have the edge (FINDRISC, CANRISK, ADA). The population of the Philippines in the provinces, where the tools were applied to, is mono-ethnic. Therefore this agrees with the conclusion of the paper of Tanamas SK, et al. that similarity in ethnicity does not predict the performance of a tool. Since we used a similar method in the application of all the tools (which is an emphasis of the Tanamas SK article), we are confident in our findings that FINDRISC performed the best in that setting.

Location: N/A, though the Tanamas et al. article is now cited in the manuscript.

Reviewer Comment: The choice of measures to assess suitability of the scores are incomplete. The authors could also include a measure of calibration or goodness of fit.

Revision/Response: We calibrated the FINDRISC by adjusting the scores and calculating the sensitivity and specificity for each score (Table 5). We revised our method of calibration in this paper by using the Receiver Operating Characteristic (ROC) curves. We found the same result, i.e., a score of 11 was optimal for the FINDRISC.

Location: Results section, page 12-13, lines 254-260. See also: Discussion section, page 15, lines 313-317.

Reviewer Comment: Have the authors thought about re-calibrating the scores to their population? They should present some calibration curves pre and post calibration.

Revision/Response: We revised our calibration as specified above.

Location: See above.
Reviewer Comment: The other flaw with this work is that the two groups of non-diabetes and diabetes people used to validate the score are very small. Scores should be tested on larger populations. Further these people are recruited from public and private clinics so selection bias is also an issue here. These people with diabetes would more severe cases of the disease. Also the population is mainly female and this is also problematic. These issues should be recognised as limitations.

Revision/Response: We have recognized these limitations and have already written it in our “Scope and Limitations” sections. We have these paragraphs explaining these limitations: “The current study population may not be representative the population distribution in the Philippines. Convenience sampling was used for this study. The majority of the participants were females; and most were housewives, government employees, or teachers, which were the sampling pools available to the researcher. Therefore, results in the prevalence of undiagnosed dysglycemia may not be fully representative of the Philippine population.” “Although traditionally, cross-sectional analytical designs have been used to assess the accuracy of diabetes risk screening tools, the cost and time needed to acquire results can be prohibitive. A large sample of the population is required to get enough participants with diabetes for sufficient analytic power. Given the limitations of budget and time, the use of a case-control method was the best solution to ensure enough individuals diagnosed with diabetes were recruited in order to obtain a scientific answer to the question of which tool would be most appropriate to use in this situation. Further studies may be required to confirm the accuracy of the FINDRISC and CANRISK, but other tools with lower accuracy need not be tested.”

Location: Discussion section, Scope and Limitations sub-section, page 15, lines 318-323 and 335-342.
Reviewer Comment: The authors comment that a ROC greater than 0.5 is good diagnostic accuracy. This is not so true. The ROC is a measure of the tool's ability to rank people in terms of risk-discrimination, rather than accuracy. An ROC of 0.5 tells that the ability of the score to rank people in terms of risk of diabetes is 50%, i.e. the chance that the score would order two individuals in correct order in terms of diabetes risk is 50%. We really need a diabetes score to have a ROC of at least 0.7. This sentence should be modified.

Revision/Response: We agree with this comment. We reviewed the paragraph where this was mentioned and we understand how the sentence we wrote could be taken out of context. In the introduction of the paragraph we stated that: “An AUC of 0.9-1.0 is considered excellent, 0.8-0.9 very good, 0.7-0.8 good, 0.6-0.7 sufficient, 0.5-0.6 bad, and less than 0.5 considered not useful.

However in the middle of the paragraph, we said that the “performance of FINDRISC in both studies is more than 0.5, indicating showed good diagnostic accuracy”. We were referring to the fact that FINDRISC had an AUC of 0.63 which was sufficient but not good. We have revised this sentence.

Location: Discussion section, page 14, line 299