AZRT-D-21-00379  
Development and Validation of a Novel Dementia Risk Score in the UK Biobank Cohort  
Melis Anatürk; Raihaan Patel; Georgios Georgiopoulos; Danielle Newby; Anya Topiwala; Ann-Marie G. de Lange; James H. Cole; Michelle G. Jansen; Klaus P. Ebmeier; Archana Singh-Manoux; Mika Kivimäki; Sana Suri  
Alzheimer's Research & Therapy  
  
Dear Dr Anatürk,  
  
Thank you for considering Alzheimer's Research & Therapy.  
  
Peer review of your manuscript is now complete and we regret to inform you that your manuscript cannot be accepted for publication in Alzheimer's Research & Therapy.  
  
Please find the reviewers' reports at the end of this email. We hope that these will provide useful guidance for the future direction of your work. Please also take a moment to check our website at https://www.editorialmanager.com/azrt/ for any additional comments that were saved as attachments.  
  
We wish you every success with your research and we hope that you will consider us again for other manuscripts in the future however please note that we will not accept resubmissions.  
  
  
Best wishes,  
  
Douglas Galasko  
Alzheimer's Research & Therapy  
https://alzres.biomedcentral.com/  
  
Review reports:  
Reviewer #1: The authors developed and 'validated' a dementia risk tool on the UK Biobank. There are several significant limitations with this paper both conceptually and methodologically. The authors also misrepresent the significance of their findings by comparing their tool with other tools that were not designed to predict the same outcome or to be used for the same ages or settings. This means their conclusions are not justified.   
  
Specific issues are:  
1. There seems to be an overall lack of theoretical rationale underlying this paper. What was the purpose of developing this tool? The usual reason for developing a risk assessment tool is to identify risk factors that are modifiable in order to guide risk reduction interventions. The totally data driven approach has resulted in authors identifying mostly non-modifiable risk factors and omitting the key modifiable risk factors for dementia that are reported by the Lancet Commission and the WHO Guidelines. Hence it is unclear what the benefit of this tool would be, particularly when there are other tools that do assess the established risk factors for dementia. The use of the tool in primary care (which is the goal of the authors) would result in a major missed opportunity for GPs to focus on the established, modifiable risk factors for dementia. Also informing patients of their APOE status in primary care needs to be accompanied by genetic counselling and raises many ethical issues.  
  
2. Descriptions of the variables are provided in supplementary material but the thresholds for risk that were used in the scale or evaluated in the modelling are not reported. This information needs to be provided for all risk measures that were candidates to enable proper evaluation of the analysis i.e. what level of physical activity was classified as increasing risk? How were these thresholds determined? Reporting this information is essential for research quality and reproducibility.   
'  
3. What was the methodology or rationale for the selection of comparison risk scores? Why wasn't the LIBRA score considered?  
  
4. The authors say in the supplementary material "The ANU-ADRI did not provide sufficient information to compute predicted risk, hence, only the weighted sum provided was used. As a consequence, calibration was not evaluated for the ANU-ADRI [1]." What Is meant here by 'calibration? Is this Information provided In the validation articles on the ANU-ADRI where Its predictive validity Is reported for three datasets? (Anstey et al, 2014, PLOSEOne)? If this Is lacking, then how and why was the Instrument used In this study and how can the comparison have been valid?  
  
5. The authors do not provide the details of how they calculated the comparison tools ie the syntax used or algorithms for combining of variables, of which variables were included in each scale for each dataset. They do mention that they lacked variables for the ANU-ADRI and that they included different variables in the scale on each dataset - this begs the question of whether the use of this scale was appropriate in the study and whether the comparisons made were appropriate. To evaluate results and for reproducibility, the methodology for scoring each of the comparison scales (for each dataset) and the variables included in them needs to be reported.  
  
6. The most concerning aspect of this study is that the authors set up a 'straw man' comparison with other dementia risk tools which they admit were missing components and not designed for the age-groups used in the study. The comparison scales were designed for use in different settings and for different outcomes than the UK biobank scale. Therefore the conclusions that the authors make about their scale being superior are based on a flawed research design. The CAIDE scale was designed to use mid-life risk factors to predict late-life risk of dementia over a far longer follow-up period than that of the UK biobank and was not designed to include APOE. APOE was added to the CAIDE scale and of course improved prediction.  
  
The authors also compared their scale to the Framingham. This scale was designed to predict cardiovascular disease and not dementia. The ANU-ADRI was designed to be self-report, and assesses risk of AD and was designed to be accessible in low resource settings and does not include APOE for this reason. The authors however, compare the tools on the outcome of dementia, in a different age-group to that for which it was designed. The authors include APOE In the biobank scale but not in their computation of the ANU-ADRI scales and then of course find better prediction. This is an invalid comparison. A comparison to the ANU-ADRI ie as a scale would use AD as the outcome measure and make the two comparison scales equivalent on APOE (either include in both or neither).   
  
7. Another major limitation of the study is the use of registry data for the outcome variables, rather than clinical diagnoses. How do the authors think their findings will apply in real life and in predicting clinical diagnoses? What biases would the use of the registry data introduce. This needs to be discussed and the implications and biases associated with the registry data explained.  
  
8. The authors note that their scale needs to be evaluated in different settings but the inclusion of APOE would limit its applicability for any quick administration or use in low income settings. Interestingly, the article by Stephan that evaluated the dementia risk scores on cohorts from low income countries which was cited by the authors, found the ANU-ADRI was more predictive than other tools designed for higher income settings.   
  
  
  
  
  
  
  
Reviewer #2: The purpose of this study was to develop and validate a novel dementia risk prediction model that could be used in everyday clinical practice. The authors leveraged the UK Biobank dataset to develop their risk score and validated it in the Whitehall II cohort. The authors compared the predictive performance of their novel risk score to that of other published risk scores. Strengths of this study include the large sample size of the UK Biobank cohort. However, there are several weaknesses and concerns, as outlined below. These weaknesses limit enthusiasm and applicability of this novel risk score.  
  
\*A major limitation of this study is the lack of rigorous adjudication for dementia diagnosis in both the UK Biobank and the Whitehall II cohorts. This limits the applicability of their novel dementia risk score. Can the authors test their new score in another dataset with adjudicated dementia diagnoses?  
\*Do the authors have information on the accuracy of dementia diagnoses using their current definition of dementia in the UK Biobank and Whitehall II cohorts?  
\*Another limitation is the low number of dementia cases in the Whitehall II cohort (n=69)   
\*The authors should determine whether their risk score is more predictive in men or women  
\*The racial and ethnic makeup of the testing and validation sets is not clear. The authors should clarify this.   
\*Can the authors validate their score in a dataset with high racial and ethnic diversity?  
\*Do the authors have any information on plasma AD biomarkers to compare the performance of their score to these biomarkers? These novel biomarkers (e.g., p-tau181) have been found to predict AD with high accuracy and can be applied in clinical practice; thus, it's not clear what their new score would add beyond these novel plasma AD biomarkers  
\*Did the authors consider that the other dementia risk scores may have had poorer predictive accuracy because UK Biobank and Whitehall II did not have information on adjudicated dementia diagnoses?  
\*The authors should validate their dementia risk score in an older cohort, given that their validation and test sets were aged on average 56 years