**Author’s response to reviews**

**Title:** Evaluation and refinement of the PRESTARt tool for identifying 12-14 year olds at high lifetime risk of developing type 2 diabetes compared to a clinicians assessment of risk: a cross-sectional study

**Authors:**

Laura Gray (lg48@le.ac.uk)
Emer Brady (emer.brady@uhl-tr.nhs.uk)
Olatz Albania (oalbaina@kronikgune.org)
Charlotte Edwardson (ce95@leicester.ac.uk)
Deirdre Harrington (dh204@leicester.ac.uk)
Kamlesh Khunti (kk22@le.ac.uk)
Joanne Miksza (jkm42@leicester.ac.uk)
Joao Raposo (filipe.raposo@sapo.pt)
Ellesha Smith (eas24@leicester.ac.uk)
Andriani Vazeou (agerasim@gmail.com)
Itziar Vergara (itziar.vergaramitxeltorena@osakidetza.eus)
Susann Weihrauch-Blüher (Susann.Blueher@medizin.uni-leipzig.de)
Melanie Davies (melanie.davies@uhl-tr.nhs.uk)

**Version:** 2  **Date:** 12 Mar 2019

**Author’s response to reviews:**

We thank the reviewers for their thorough review of our work, we have amended the paper accordingly and feel these changes have strengthen the manuscript.
Reviewer 1

Comment: Title: The authors must modify the title to reflect the essence of the study. For example, include the name of the studied tool.

Response: We have amended the title to take into account this comment and the comments made by reviewer 2 – “Evaluation and refinement of the PRESTARt tool for identifying 12-14 year olds at high lifetime risk of developing type 2 diabetes compared to a clinicians assessment of risk: a cross-sectional study”.

Comment: Background: Some information in this section: please, see lines 12 to 26 and 33 to 43 are part of the methods. The authors could include information regarding to similar tools published in the scientific literature, but not describe how their study was performed, it must be placed in the appropriated section.

Response: We have rearranged the text as suggested by the reviewer.

Comment: Methods: The authors indicated that they collected data from 636 adolescents aged 12-14 years from five European countries in the abstract, but they did not say it in the methods section. The authors only said: "The target sample size was 500 adolescents (100 per country)…" how was calculated the sample size to be representative for the interest group, in this case for adolescents?

Response: The sample size was calculated based on the number needed to ensure sufficient high risk adolescents being included to validate the tool. The representativeness of the sample was not taken into account. We have reordered the sample size section to make this clearer. “The target sample size was 500 adolescents (100 per country). This minimum sample size was chosen as methodological studies have suggested that 100-200 cases and 100-200 non-cases should be included for the external validation of risk prediction models [20]. Using the proposed sampling frame, we estimated that at least 100 of the 500 adolescents recruited should be at high risk. The final sample size recruited was 636 adolescents”. We have also emphasised in the limitations of the study that the sample recruited within each country may not be representative – “Also we cannot guarantee the representativeness of the included samples within each country as the study was not designed to recruit representative samples”.

Comment: The authors stated in the manuscript: "they planned to sample across the BMI distribution with over sampling at the higher BMI percentiles to ensure they recruited high risk
participants." What was the rationale to plan it in this way, however, the findings shown in the study did not support this observation?

Response: We have clarified in the text of the methods section why we initially planned to over-sample in those with high BMI – “We planned to sample across the BMI distribution with over sampling at the higher BMI percentiles to ensure we recruited sufficient numbers of participants with risk factors for developing type 2 diabetes in order to be able assess the tool”. We have also stated in the results “Although we aimed to quota recruit by BMI this was not possible in all countries”.

Comment: The authors must clarify accurate the inclusion, exclusion and elimination criterions used in the performance of the study.

Response: We have clarified in the methods section the inclusion/exclusion criteria for the study – “The inclusion and exclusion criteria were purposely broad. Only 12–14 year olds inclusive were included (as stipulated by the European Union tender) and those who were willing and able to give written informed assent (after obtained written informed parental/guardian consent). Individuals were ineligible if did not meet the inclusion criteria and/or had an existing diagnosis of type 1 or type 2 diabetes mellitus”.

Comment: In the Biological maturity status: the authors said that the questionnaire to evaluate this status was not administered at the Portuguese site, and then the authors must explain how this data were management in comparison with the countries where this status was measured?

Response: We were unable to gain ethical approval to collect this data in Portugal. The biological maturity data was collected as part of the extensive phenotyping conducted. Where this data were available they were shared with the clinical reviewers when conducting their adjudication. If these data were not available this part of the database was blank and the clinical adjudicators were not able to use it in their review. We have added the following text to clarify this in the methods section –“Missing data were shown as blank responses and therefore the reviewers could not use this in their assessment”.

Comment: Results: It is highly recommendable to elaborate a flow chart to show the experimental design of the study. It will help too much to understand how the selection of the participants was and all features evaluated along of the study.

Response: We have added a sentence on how participants were recruited to the methods section – “A variety of recruitment settings were used. In Spain, Greece and Germany potential
participants were identified in clinical settings, whereas schools were used in Portugal and the UK”. We have not added a flow diagram as participants were only attended one visit for data collection as this was a cross-sectional study. The methods section lists all the data collected at this session.

Comment: To paste the prestart form used in the study in the figure 1 without edit it is not convenient for the presentation of the data of the study. The authors need to modify the form to present it adequately in the manuscript.
Response: This has been updated as requested.

Comment: Why there were many missing values in the analysis please see table 1? For example, in the Ethnicity category for Portugal and Greece the 100% of the sample is missing.
Response: In the methods we state that we were unable to collect data on ethnicity in Portugal and Greece due to the within country ethical requirements. We have reiterated this in the results section – “The majority of participants were Caucasian from the sites in Germany and Spain. In the UK site 54% were of non-white ethnicity, reflecting the ethnic diversity of the area where recruitment took place. We were unable to collect ethnicity data in Greece and Portugal”.

Comment: The title of all tables must be rewritten, titles are very brief. In the table 3: describe PPV and NPV
Response: We have added more detail to all of the tables and ensured that all abbreviations are explained.

Reviewer 2
Comment: There is no clear definition of developing T2DM given. The lifetime incidence of T2DM is high and increasing rapidly. Is the tool trying to predict the development of T2DM during participants' lifetime? And I wonder what timeframe the clinicians used and if there were all consistent among the clinicians in the study.
Response: We thank the reviewer for this important point, we have now clarified throughout the manuscript that we are identifying those at high risk of development of T2DM during participants' lifetime. We do not know if the timeframe used was consistent across clinicians.
Comment: Without longitudinal assessments, it is practically impossible to properly evaluate the performance of a predictive tool for T2DM development. As the authors noted, the approach they took is not a typical way to evaluate a predictive assessment tool. Clinicians' judgement on the risk of developing T2DM for the study cohort cannot be objective gold-standard in such a prediction model development/evaluation. To me, the study examined how well the questionnaire-based tool could recapture what human experts, clinicians, judged on the risk of developing T2DM. Therefore, using 'validation' in this study should be avoided and the overall theme needs to be changed to the revising the tool and the examination of how well the tool's results agreed with clinicians' judgement.

Response: We agree with the reviewer, we have used a nonstandard way of assessing our tool. We had tried to capture the limitations of our validation in the text, but we agree that the use of the term validation is maybe too strong given the methodology used. Throughout we have tried to make it clearer that we are comparing the outcome of the tool to a clinician’s assessment of diabetes risk. We have used the word evaluation instead of validation throughout. For example, in the abstract we have rephrased our aim, which now states – “We aimed to evaluate whether a tool developed for community use to identify adolescents at risk of developing T2DM during their lifetime agreed with a risk assessment conducted by a clinician using data collected from five European countries. We also assessed whether the tool could be simplified”. We have also updated the title.

Comment: I wonder if clinicians reported the rationales of their judgements for each participant. Compiling these rationales into rules would probably lead to a better or more accurate or representative of the clinicians' risk assessment.

Response: We thank the reviewer for this suggestion, unfortunately we did not capture how the clinicians made their judgement. We have added this as a limitation of the work to the discussion – “In terms of the clinical review, a strength of this is the use of two independent reviewers for each individual. Unfortunately we did not collect data on how the clinicians made their decisions and which data they used to form these. These data could have informed which risk factors to include in the refined tool. Even though this was not conducted the final tool still maintain a high level of performance”.

Comment: I wonder how well the clinicians agreed or used the same judgement criteria. As the assignment of clinicians was done in a country-based cluster, there could have been region-based biases. Did the authors examine how well all the clinicians in the study generally agree? This could be assessed by having all the clinicians evaluate the same subset of participants, if not all.
Response: Before starting the actual adjudications all clinicians took part in some training, they were provided with data from 20 participants and all asked to review their risk status. This process made sure they were comfortable with using the system and gave them an opportunity to feedback any comments on improving the system. Unfortunately the data from these were not captured. Therefore we are unable to assess between county differences in adjudication using the same data. We have presented data on the within country comparison in Table 2. We have added the following to the discussion – “Adjudication was also performed within countries, i.e. the clinicians reviewing each participant all came from the same country and therefore we cannot assess between country differences in adjudication”.

Comment: The refinement of the tool improved the overall performance in the current evaluation scheme. However, it is not clear why the core risk factors were separated, and the authors 'sequentially' removed risk factors for evaluation (Table 4). Did the authors also check the tool using only the overweight/obese (single factor)?

Response: The core risk factors were separated after feedback from our stakeholders and the PRESTARt collaborative, also the ADA guidance for the screening of children and young people also used only a measure of weight as the core risk factor. The sequential removal of risk factors was also based on feedback from stakeholders and our PRESTARt collaborative. We have now assessed using weight alone and this has been added to the discussion on this point – “One could argue that weight alone should therefore be used to assess risk and indeed doing this does not hamper the performance of the tool (ROC 0.80, 95% CI 0.77, 0.82), however we believe a more holistic assessment of risk is important for a number of reasons”.

Comment: The % of the participants required the third assessment needs to be disclosed.

Response: This is reported in the results section – “For 76% of cases the two clinical experts agreed, therefore 24% required a third party to reach a final decision on the risk status”.

Comment: It would be useful to include a table, like Table 1, showing the numbers of participants for each risk factor in the tool.

Response: We have included this information in Table 5.

Comment: Was there any participant who was not obese/overweight but determined to be at high risk by clinicians?
Response: This has now been added to the results – “Eighteen participants (6.41%) were deemed to be high risk by the clinicians were not overweight/obese”.

Comment: A table like Table 2 would be needed for the refined tool.
Response: We have added this data to Table 2 rather than adding an additional table.

Comment: A CONSORT chart would be useful.
Response: We have not included a CONSORT chart, these were designed for randomised controlled trials with follow up of participants. This was a cross-sectional study so participants were seen on one occasion.

Comment: The final model is Figure 1 should be indicated in Table 4.
Response: Tool 20 is the final tool, this has been added to Table 4.