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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Patient-reported outcome and experience measures for diabetes: development of scale models, differences between patient groups and relationships with cardiovascular and diabetes complication risk factors, in a combined registry and survey study in Sweden
AUTHORS	Borg, Sixten; Eeg-Olofsson, Katarina; Palaszewski, Bo; Svedbo Engström, Maria; Gerdtham, Ulf-G; Gudbjörnsdottir, Soffia

VERSION 1 – REVIEW

REVIEWER	M. Shawn McFarland Department of Veterans Affairs PBM Clinical Pharmacy Practice Office
REVIEW RETURNED	16-Jul-2018

GENERAL COMMENTS	Overall a very well written and very interesting research topic on the use of PROM/PREM for diabetes care. Although the correlation to clinical measures of outcome remain questionable, overall this is an evolving area of interest. After review I have no major methodological concerns or questions. Please comment if this study was reviewed by an appropriate IRB. Pg-17 line 9 consider rewording the use of "severe" to poorly controlled given there is no severity in the diagnosis of diabetes alas more of question of control. Very well done.

REVIEWER	Sangchoon Jeon Yale School of Nursing, United States
REVIEW RETURNED	23-Jul-2018

GENERAL COMMENTS	This study was very well organized with appropriate steps and conditions of developing instruments using IRT. The tests were performed with large sample size and possible biases by diabetes type and age were well addressed. However, there are some recommendations to present statistics. - Need to provide statistics for test-retest reliability. For example, it was stated that ICC was calculated but not presented in the

	manuscript. It would be interested in seeing ICC separately by diabetes type. - On page 12, It was stated that … (Table 2) showed no sign of local dependency or non-monotonicity, and furthermore, all were strong, …. Table 2 does not present any statistics for them. It would be good to make a table for summarizing those necessary conditions (maybe statistics). Alternatively, Authors could state more details for diagnostics of local dependency and non-monotonicity. - Does DIFs were significant for ManD and NLBS only? Figure 2 shows substantive difference for other scales as well. Instead of Figure2, it would be more informative to show Tables for DIF results in all subscales. - For tests for association between risk factors and IRT scores, Table 4 is confused. What kind of test statistics was used to determine significant high/low in groups with IRT score below 10th percentile? Need more explanations how to decide “higher” and “lower” in Table 4 (a)&(b). - Is it possible that the different associations with risk factors between Type 1 and 2 groups was influenced by age difference? Risk factors and some IRT scores might be associated with age which could be a confounding factor.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Comments to Author:
3. After review I have no major methodological concerns or questions. Please comment if this study was reviewed by an appropriate IRB.
Response: Our study was indeed reviewed by the Regional Ethical Review Board in Gothenburg, mentioned on page 2. We added the registration number of our application.

Please note, the National Diabetes Registry, located in Gothenburg, was the formal applicant and primary investigator, hence the regional ethical review board in Gothenburg was the appropriate one to review our study.

Ethical review in Sweden is carried out in an organization of independent authorities consisting of six regional boards, and a central vetting board (to which cases may be submitted to e.g. re-examine a decision). In our case, the ethical review was completed in the regional board of Gothenburg.

4. Pg-17 line 9 consider rewording the use of "severe" to poorly controlled given there is no severity in the diagnosis of diabetes alas more of question of control.

Response: Thank you for pointing this out. We reworded the sentence as suggested.

Reviewer 2: Comments to Author:
This study was very well organized with appropriate steps and conditions of developing instruments using IRT. The tests were performed with large sample size and possible biases by diabetes type and age were well addressed. However, there are some recommendations to present statistics.

5. Need to provide statistics for test-retest reliability. For example, it was stated that ICC was calculated but not presented in the manuscript. It would be interested in seeing ICC separately by diabetes type.
Response: We have added a supplementary table S3, which shows ICC for each scale, by diabetes type, in a main test-retest analysis of respondents reporting no change, and in a sensitivity analysis that included responders reporting small improvement, no change, or small worsening. This table is mentioned on page 12. We suggest this table is provided as supplementary material, to avoid a too lengthy main manuscript.

6. On page 12, It was stated that … (Table 2) showed no sign of local dependency or non-monotonicity, and furthermore, all were strong, …. Table 2 does not present any statistics for them. It would be good to make a table for summarizing those necessary conditions (maybe statistics). Alternatively, Authors could state more details for diagnostics of local dependency and non-monotonicity.
Response: We chose the second suggestion and added more details on our procedure for checking for local dependency and non-monotonicity in the Methods section, on page 8. A complete account of the statistics and graphs reviewed in our procedure, would require a large amount of space.

7. Does DIFs were significant for ManD and NLBS only? Figure 2 shows substantive difference for other scales as well. Instead of Figure2, it would be more informative to show Tables for DIF results in all subscales.
Response: We detected DIF with regards to diabetes type in ManD, NLBS and MDMT (among the oldest respondents).

Please note that Figure 2 shows mean IRT scores by diabetes type and age groups, whereas DIF refers to *items functioning differently in different subgroups*. Furthermore, Figure 2 omits Manage your diabetes (ManD) and Not Limited by Blood Sugar (NLBS), since these scales are diabetes type specific and cannot be compared across diabetes types.

We argue that Figure 2 presents interesting data and we suggest it remains in the manuscript.

Our testing of DIF was initially carried out in IRT Pro using the test mentioned. However, due to our large sample leading to high statistical power, many if not most items were flagged for DIF. We therefore proceeded by fitting (a) a common model, and (b) group specific models, with regards to the group variable under study with regards to DIF. Using (a) and (b), we examined the difference in estimated IRT score for respondents using the common versus the group-specific model. Based on these differences, we decided whether the DIF, if detected, had any meaningful impact. Hereby we reviewed graphs of the differences, looking for systematic differences, and we examined the magnitude of the differences. We added text on page 9 to clarify this. The specific results from the initial Wald tests are unfortunately no longer available due to technical computer problems. For this reason, we removed the mentioning of the Wald test in the methods text (page 9). A presentation of the material mentioned above, i.e. graphs and difference estimates, would require quite some additional space. Please advise if this would still be preferable.

8. For tests for association between risk factors and IRT scores, Table 4 is confused. What kind of test statistics was used to determine significant high/low in groups with IRT score below 10th percentile? Need more explanations how to decide “higher” and “lower” in Table 4 (a)&(b).
Response: Thank you for pointing out this omission. We have now added which test method we used, in the Methods section on page 10 as well as in Table 4. Furthermore, we reorganized the (b) part of Table 4 so it would be more similar to the (a) part. Now, in both (a) and (b), the subgroups appear as rows, and the tested variables appear as columns. We divided (b) vertically by diabetes type, although we divided (a) horizontally. Doing this consequently, e.g. horizontally in both (a) and (b), we believe would require too much space. Please advice if you think we should revise.

9. Is it possible that the different associations with risk factors between Type 1 and 2 groups was influenced by age difference? Risk factors and some IRT scores might be associated with age which could be a confounding factor.
Response: We detected differences between type 1 and 2, in some IRT scales including Not Limited by Diabetes (NLD), after controlling for age. According to Table 4, NLD in the subgroup with high HbA1c differs from NLD in type 1 overall. No such deviation is seen in type 2. A careful study of the joint distribution of NLD and HbA1c in the upper range of HbA1c might clarify why this is the case. We will explore such relationships further in a future publication.

However, our idea behind Table 4 was simply to demonstrate that when looking at subgroups with low (bad) IRT scores, or subgroups with high risk factors, there are only scattered relationships between risk factors and IRT scores (although some appear in both diabetes types). Based on this, we argue that neither risk factors nor IRT scores are superfluous, and that they rather complement each other, in forming a complete description of a person's situation.

10. You will receive a proof if your article is accepted, but you will be unable to make substantial changes to your manuscript, please take this opportunity to check the revised submission carefully.
Response: We took the opportunity to make a number of minor changes to the manuscript and in the supplementary material (indicated with Track Changes).

VERSION 2 – REVIEW

REVIEWER	Sangchoon Jeon Yale School of Nursing, U.S.
REVIEW RETURNED	29-Sep-2018

GENERAL COMMENTS	Your revision and responses are pretty satisfied.