PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	New diabetes questionnaire to add patients' perspectives to diabetes care for adults with type 1 and type 2 diabetes – Nationwide cross-sectional study of construct validity assessing associations with generic health-related quality of life and clinical variables
AUTHORS	Svedbo Engström, Maria; Leksell, Janeth; Johansson, Unn-Britt; Borg, Sixten; Palaszewski, Bo; Franzén, Stefan; Gudbjörnsdottir, Soffia; Eeg-Olofsson, Katarina

VERSION 1 – REVIEW

REVIEWER	Shaun Wen Huey Lee Monash University Malaysia
REVIEW RETURNED	21-Apr-2020

GENERAL COMMENTS	This is an interesting study conducted to examine the correlation between a new diabetes questionnaire and how these correlate with generic health related QoL and clinical variables. 1. Overall the study is rather technical in nature , so i am not sure whether the authors would be able to either simplify this to suit most of the readers of this journal 2. In the methods, the authors suggest and performed several regression analysis to support the relationship between the tool developed and those from Sf36. This i believe is a good attempt but would the authors be able to perform additional analyses including structural equation modellling i believe that this would add more value to the paper rather than it currently stands as its mostly a methods paper describing the analyses and justification for further use which is important but if the authors can make a bigger leap forward, this would make the paper much more useful 3. A short papragraph or 2 on how applicable this results to other settings or how other individuals who would be attempting to use this questionnaire in the future and how it can be adapted to their setting would be great.

REVIEWER	Thaddäus Tönnies German Diabetes Center, Düsseldorf, Germany
REVIEW RETURNED	02-Jun-2020

GENERAL COMMENTS	Engström et al. report on a cross-sectional survey conducted in Sweden, which was used to evaluate the construct validity of the

	Diabetes Questionnaire - an instrument that intends to measure a broad range of aspects important to people with diabetes (e.g. well-being, impact on daily life, support from diabetes care). Overall, the paper is well written. One strength of the study is the large and representative sample based on the National Diabetes Register. However, there are some points that might be revised in my opinion. Major points: - The authors should provide more guidance for readers not familiar with psychometrics and psychological testing. For instance, the introduction should give some background on construct validity in distinction to other types of validity (e.g. content validity). Furthermore, the concepts of convergent and discriminant construct validity should be introduced. In the current version, the authors vaguely refer to these concepts in the discussion only, without direct reference to methods and results. Against this background, the authors should state why pre-specified assumptions are needed to evaluate construct validity and which methods were used to test each of these assumptions (perhaps with regard to convergent and discriminant construct validity). Also, I suggest to structure the results according to the assumptions. - The authors should give a brief rationale for each pre-specified assumption. - The statistical methods need more elaboration and justification. For instance it is unclear to me, why the authors first used Spearman's rank correlation, random forests for non-linear association in second step and multiple regression to study group level associations in a third step (seemingly ignoring non-linearity again). In my opinion, for two continuous variables, Pearson correlation seems preferable to evaluate linear relationships. If the authors think random forests are neccesary to model non-linear relationships, it should be used for all subsequent analyses (including the group level analyses). The more complex models (random forests, multiple regression) need some justification in terms of what information is gained with regard to evaluate construct validity. - Similarly, I am not sure about the rationale for the group level analysis. HbA1c was already investigated in correlation analysis and in the random forest analysis. I suggest, the authors explain what information is added in terms of construct validity by analyzing HbA1c additionally as a categorical variable in multiple regression analysis. Minor points: Abstract - As for the main text, I suggest to provide more specific information on which results were produced by which method in order to inform which aspect of construct validity. Introduction - The second paragraph starts with 'The Diabetes Questionnaire has a sound basis and ...'. I suggest to delete 'has a sound basis and'. Methods - It is stated that no formal sample size calculation was performed because there was lack of data on the standard deviations for the Diabetes Questionnaire. I think this is a rather weak justification for completely refrain from sample size calculation, since at least

	reasonable assumptions on the standard deviation could have been used (e.g. from similar questionnaires). Instead, the sample size was estimated to enable subgroup analysis. Please state how this estimation was performed and what the desired sample size was. - Please state that multiple imputation was used before describing how imputed data sets were combined. Discussion - It is stated that 'The evaluation of construct validity is a work of putting the pieces together'. I suggest to elaborate a little bit further on this statement, because as it stands, it does not help the reader to understand the complexity of construct validity. Maybe, this aspect should already be mentioned in the introduction in order to clarify why pre-specified assumptions and several analysis strategies are helpful to evaluate construct validity.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Reviewer Name Shaun Wen Huey Lee Institution and Country Monash University Malaysia	
This is an interesting study conducted to examine the correlation between a new diabetes questionnaire and how these correlate with generic health related QoL and clinical variables.	Thank you for taking the time to review our manuscript and we are glad to hear that you found it interesting. Please find our responses to raised points below.
1. Overall the study is rather technical in nature , so i am not sure whether the authors would be able to either simplify this to suit most of the readers of this journal	We hope that our revisions have made the manuscript easier to read.
2. In the methods, the authors suggest and performed several regression analysis to support the relationship between the tool developed and those from Sf36. This i believe is a good attempt but would the authors be able to perform additional analyses including structural equation modellling i believe that this would add more value to the paper rather than it currently stands as its mostly a methods paper describing the analyses and justification for further use which is important but if the authors can make a bigger leap forward, this would make the paper much more useful	Thank you for the input. The suggested additional analyses would be tempting and motivated, however not possible to squeeze into this work. As pointed out in the previous comment, the manuscript is quite technical. We prefer not to make it even more technical.
3. A short papragraph or 2 on how applicable this results to other settings or how other individuals who would be attempting to use this questionnaire in the future and how it can be adapted to their setting would be great.	The Diabetes Questionnaire is currently only available in Swedish, a limitation which has been clarified in the Discussion section/Findings and implications, page 16. The use within other settings needs to be preceded by a translation and cultural adaption. However, these results add to the evidence for validity, thereby strengthening the relevance for others who might be interested in initiating the process

	of translation.

Reviewer: 2 Reviewer Name Thaddäus Tönnies Institution and Country German Diabetes Center, Düsseldorf, Germany	
Engström et al. report on a cross-sectional survey conducted in Sweden, which was used to evaluate the construct validity of the Diabetes Questionnaire - an instrument that intends to measure a broad range of aspects important to people with diabetes (e.g. well-being, impact on daily life, support from diabetes care). Overall, the paper is well written. One strength of the study is the large and representative sample based on the National Diabetes Register. However, there are some points that might be revised in my opinion.	Thank you for taking the time to review our manuscript and for these supportive and constructive comments. We have responded to raised points below.
Major points:	
- The authors should provide more guidance for readers not familiar with psychometrics and psychological testing. For instance, the introduction should give some background on construct validity in distinction to other types of validity (e.g. content validity). Furthermore, the concepts of convergent and discriminant construct validity should be introduced. In the current version, the authors vaguely refer to these concepts in the discussion only, without direct reference to methods and results. Against this background, the authors should state why pre-specified assumptions are needed to evaluate construct validity and which methods were used to test each of these assumptions (perhaps with regard to convergent and discriminant construct validity). Also, I suggest to structure the results according to the assumptions.	In the background section, pages 4-5, we have elaborated on the concept of construct validity including the concepts of convergent and divergent (discriminant) construct validity. These concepts have also, where relevant, been added in the Methods section/Pre-specified assumptions (pages 6-7), in the Results section (pages 12-13), and in the Discussion section (pages 15-16). In the Methods section/Pre-specified assumptions (pages 6-7), we state that the pre-specified assumptions are related to the correlations. To further clarify this, we have added a few words in Methods section/Statistical analysis, page 7. We have carefully considered the suggestion to restructure the results. We chose to keep the current structure but the headings in the Results section have been somewhat amended to clearer relate the results to the pre-specified assumptions and to clarify which aspect of construct validity it refers to (pages 12-14).

- The authors should give a brief rationale for each pre-specified assumption.	The rationale for the pre-specified assumptions have been added in the paragraph presenting the pre-specified assumptions in the methods section, pages 6-7.
- The statistical methods need more elaboration and	The statistical methods employed

justification. For instance it is unclear to me, why the authors first used Spearman's rank correlation, random forests for non-linear association in second step and multiple regression to study group level associations in a third step (seemingly ignoring non-linearity again). In my opinion, for two continuous variables, Pearson correlation seems preferable to evaluate linear relationships. If the authors think random forests are neccesary to model non-linear relationships, it should be used for all subsequent analyses (including the group level analyses). The more complex models (random forests, multiple regression) need some justification in terms of what information is gained with regard to evaluate construct validity.	are very carefully selected to address different aspects of the study. One part of the analysis is targeted at the relation between variables on an individual level which makes correlations and random forest useful, and one part is target at group level comparisons between persons with different level of diabetes control as defined by HbA1c. On an individual level we wanted to allow for nonlinear associations which leads to the preference of Spearman's correlation over Pearson's. We also prefer Pearson since it is much less sensitive to outliers. Random forest is used to capture the relative importance of the independent variable and of the main alternative here would be to calculate R^2 for each variable in a parametric model but that would only capture the linear part of the association and underestimate the importance of variable. While machine learning methods such as random forest are both flexible and very useful they are somewhat harder to interpret and communicate. In the group level analysis, we prioritized getting parameter estimates to illustrate that level of glycaemic control is to some extent manifested in the Diabetes Questionnaire scales on a group level. While the associations between the Diabetes Questionnaire scales and glycaemic control on a group level could have been illustrated by partial dependence plots from random forest models they are harder to communicate and would make an already somewhat technical paper harder to read. Some clarifications have been added in the Methods section/Statistical Analysis (pages 7-8) regarding the machine learning analyses and the multiple regression analyses.
- Similarly, I am not sure about the rationale for the group level analysis. HbA1c was already investigated in correlation analysis and in the random forest analysis. I suggest, the authors explain what information is added in terms of construct validity by analyzing HbA1c additionally as a categorical variable in multiple regression analysis.	As previously mentioned, we have elaborated on the concept of construct validity in the background section (pages 4-5) and added a clarification regarding the group-level associations in the methods section/Statistical Analysis (page 8).

	We believe this to clarify that the analysis of HbA1c as a categorical variable adds information about sensitivity to clinically relevant subgroups, which adds to the supporting evidence for construct validity. We have also altered the heading for the related parts in the Results section (page 14). Groups of glycaemic control, i.e. HbA1c, are clinically relevant as HbA1c is the most central clinical outcome in diabetes care on both individual and group level. The chosen intervals are related to international and Swedish guidelines, as mentioned in the Methods section/Statistical Analysis, page 8. We also added two references to this statement. The regression analysis comparing HbA1c groups is meant to illustrate the impact of HbA1c on the Diabetes Questionnaire scales on a group level. While this can be achieved using partial dependence plots from a random forest we decided to use a regression model to get parameter estimates of each HbA1c group as that in our experience is easier for most readers to follow.
Minor points:	
Abstract - As for the main text, I suggest to provide more specific information on which results were produced by which method in order to inform which aspect of construct validity.	We have made some alterations in the abstract (page 2) to clarify this. Consequently, we omitted some other details to keep the abstract within the word limit.

Introduction - The second paragraph starts with 'The Diabetes Questionnaire has a sound basis and ...'. I suggest to delete 'has a sound basis and'.	Revised as suggested (page 4).
Methods - It is stated that no formal sample size calculation was performed because there was lack of data on the standard deviations for the Diabetes Questionnaire. I think this is a rather weak justification for completely refrain from sample size calculation, since at least reasonable assumptions on the standard deviation could have been used (e.g. from similar questionnaires). Instead, the sample size was estimated to enable subgroup analysis. Please state how this estimation was performed and what the desired sample size was.	The sample size was determined to allow for scale development in for the Diabetes Questionnaire scales and by subgroups we mean that we wanted to have headroom for this in at least some subgroups. Scale development is not a standard situation for a sample size calculation, so we relied on rules of thumb and recommendations from literature findings (Cella & Chang, 2000; Edelen & Reeve, 2007) to come up with the sample size. Their

	recommendations indicated that a sample size of about 1 000 respondents was required for fitting scale models and estimating individual patient scores for operational use. We wished to retain headroom for developing separate scales in subgroups, e.g. separately for type 1 and type 2 diabetes. That being said, the sample size was not determined for the current analysis and we have rephrased the first paragraph in the Methods section/Sample and data-collection, page 5. Cella, D., Chang, C.H. A discussion of item response theory and its applications in health status assessment. Med. Care 2000, 38, II66-II72. Edelen, M.O., Reeve, B.B. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual. Life Res. 2007, 16, 5-18.
- Please state that multiple imputation was used before describing how imputed data sets were combined.	Revised as suggested.
Discussion - It is stated that 'The evaluation of construct validity is a work of putting the pieces together'. I suggest to elaborate a little bit further on this statement, because as it stands, it does not help the reader to understand the complexity of construct validity. Maybe, this aspect should already be mentioned in the introduction in order to clarify why pre-specified assumptions and several analysis strategies are helpful to evaluate construct validity.	As previously mentioned, we have elaborated on the concept of construct validity in the background section, pages 4- 5. With this amendment, we believe this to be clearer.

VERSION 2 – REVIEW

REVIEWER	Thaddäus Tönnies German Diabetes Center
REVIEW RETURNED	14-Jul-2020

GENERAL COMMENTS	All points have been addressed and I have no further comments.