

## 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.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	The BDS checklist as measure of illness severity: A cross-sectional cohort study in the Danish general population, primary care and specialized setting
<b>AUTHORS</b>	Petersen, Marie; Rosendal, Marianne; Ørnbøl, Eva; Fink, Per; Jørgensen, Torben; Dantoft, Thomas; Schroeder, Andreas

### VERSION 1 – REVIEW

<b>REVIEWER</b>	David Herdman King's College London, Institute of Psychiatry, Psychology & Neuroscience, UK
<b>REVIEW RETURNED</b>	27-Jul-2020

<b>GENERAL COMMENTS</b>	The authors provide a timely and well-written study of the structural and convergent validity of the BDS total score as a measure of symptom burden and illness severity. They provide a balanced assessment of their findings and the clinical and research implications. Although they provide references to the three different groups, it would be helpful to know the response rate for each cohort to indicate external validity. The paper would benefit from specialist statistical review.
-------------------------	---

<b>REVIEWER</b>	Sandra Dewar University of California at Los Angeles (UCLA)
<b>REVIEW RETURNED</b>	01-Aug-2020

<b>GENERAL COMMENTS</b>	<p>This paper is well written, the methodology is novel and the well orchestrated. The BDS checklist affords as a reasonable and useful way to evaluate symptom burden and screening in functional somatic disorders (FSD). Authors are encouraged to complete next steps including: validation of the BDS sum score, testing of the instrument in subpopulations, and testing sensitivity to change.</p> <p>Please check the Table numbering. There seems to be two Table 2's.</p> <p>For consideration: The tool is presented as a measure of illness severity, yet to me these are components of severity that may or may not correlate with how individuals rate their personal illness severity. This raises an important conceptual difference as perceptions of overall illness severity may impact when and what patients do about their symptoms.</p>
-------------------------	--

<b>REVIEWER</b>	Win Sen Kuan
-----------------	--------------

	National University Hospital and National University of Singapore, Singapore
<b>REVIEW RETURNED</b>	17-Sep-2020

<b>GENERAL COMMENTS</b>	<p>Thank you for the opportunity to review the statistical methods and analyses used in the manuscript entitled "The BDS checklist as measure of illness severity".</p> <p>The methods employed to assess the structural validity and convergent validity of the Bodily Distress Syndrome checklist total sum score as a measure of physical symptom burden and functional somatic disorders illness severity are both appropriate and robust.</p> <p>Minor comment: There should be an explicit explanation on why "data on the primary care cohort did not allow us to investigate convergent validity to the aggregate score" (Methods, Page 6 Line 34) in the Results section.</p>
-------------------------	--

<b>REVIEWER</b>	Chandrashekhar T Sreeramareddy International Medical University, Nepal
<b>REVIEW RETURNED</b>	29-Sep-2020

<b>GENERAL COMMENTS</b>	<p>The questionnaire BDS has been validated using THREE large DUTCH Cohorts which are perhaps, socio-demographically as well with respect to health status very distinct from each other. Authors however did not measured convergent validity of dutch version BDS with other proxy questionnaire based measures. Even authors admitted this a one important limitation as was the self-report. The subjectivity and a lack of objective validation make the results much weaker than the authors admittance. They had an opportunity validate by physicians report in primary care and specialist care setting.</p> <p>Methods are another weak section. Authors explained the statistical analyses very well indeed very candidly. However, the different statistics presented needs to be explained either in methods or as parenthesis under results about acceptable values for interpretations of measures such as validity, alpha, correlation coefficients etc. Perhaps all results including the confidence intervals of most results that are presented may not be needed under results. Readers can conveniently refer to the tables. How were the participants in different settings recruited, who, how and when the instruments were administered is an essential information under methods. Factor loading values are needed to be added in the appendix figures. The values inserted in the CFA figures will make the interpretations easy not having to refer to the tables once again. I recommended statistical review as i did not seem well qualified with limited experience doing such analyses to assess if the appropriate statistics were used.</p>
-------------------------	--

<b>REVIEWER</b>	Yue-Fang Chang University of Pittsburgh, USA
<b>REVIEW RETURNED</b>	29-Sep-2020

<b>GENERAL COMMENTS</b>	<p>1. What were included for the calculation of 'physical functioning'? The authors mentioned it was measured with an aggregate score of four items from the SF-36 subscales 'physical function', 'bodily</p>
-------------------------	---

	<p>pain', and 'vitality' (page 7). What was the 4th item (only 3 subscales were listed)?</p> <p>2. Four confirmatory factor analysis models were examined to assess the structural validity in the study. On page 10 it stated, 'These models revealed correlations between the four BDS symptom clusters and loadings from an underlying BDS factor to the four BDS symptom clusters may imply a bi-factor model'. Please explain why correlated 4 symptom clusters and loadings from the underlying factor may suggest a bi-factor model.</p> <p>3. Convergent validity of BDS sum score was evaluated by studying the association between BDS sum score and other measures using both non-parametric and parametric methods (i.e. Spearman correlation and linear regression). The distributions of BDS sum score were skewed in the general population cohort as well as in the primary care patient cohort. Was there any transformation applied to the BDS sum score prior to the linear regression analyses?</p> <p>4. There were 72% of symptoms with higher loadings for the general BDS factor than for the 4-symptom clusters in the general population cohort; however, this number dropped to 52% in the specialized settings cohort. Please discuss why the specialized setting cohort have much lower %?</p> <p>5. Please present median and inter-quartile range for the BDS sum score as well since the distributions were skewed for the general population and primary care cohort. Were the BDS sum scores of the specialized clinical setting statistically different from those of the general population?</p> <p>6. 'Error' was omitted from the statement 'A Root Mean Square of Approximation (RMSEA)' (page 8).</p> <p>7. Please define 'item-rest correlation' as some readers may be unfamiliar with this term.</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: David Herdman

Institution and Country: King's College London, Institute of Psychiatry, Psychology & Neuroscience, UK Please state any competing interests or state 'None declared': None declared

The authors provide a timely and well-written study of the structural and convergent validity of the BDS total score as a measure of symptom burden and illness severity. They provide a balanced assessment of their findings and the clinical and research implications. Although they provide references to the three different groups, it would be helpful to know the response rate for each cohort to indicate external validity. The paper would benefit from specialist statistical review.

Author's reply: We agree with the reviewer that this information may be valuable and have now displayed response rates for each cohort in the 'Population' section (p. 5). Response rates indicate responses of filling in the BDS checklist. For more information about external validity, we refer to the original BDS checklist studies.

Reviewer: 2

Reviewer Name: Sandra Dewar

Institution and Country: University of California at Los Angeles (UCLA), USA Please state any competing interests or state 'None declared': None

This paper is well written, the methodology is novel and the well orchestrated. The BDS checklist affords as a reasonable and useful way to evaluate symptom burden and screening in functional somatic disorders (FSD). Authors are encouraged to complete next steps including: validation of the BDS sum score, testing of the instrument in subpopulations, and testing sensitivity to change.  
Author's reply: We want to thank the reviewer for the encouragement for doing further research and validation studies on the BDS checklist.

Please check the Table numbering. There seems to be two Table 2's.

Author's reply: The table numbering has been checked, but it seems to be in order.

For consideration: The tool is presented as a measure of illness severity, yet to me these are components of severity that may or may not correlate with how individuals rate their personal illness severity. This raises an important conceptual difference as perceptions of overall illness severity may impact when and what patients do about their symptoms.

Author's reply: Thank you for this important food of thought. We highly agree that the included measures might not correlate with an individual's subjective experience of illness severity and that these indeed may constitute two different conceptual things. We have already included a global item of overall health in our test of convergent validity as an attempt to include the participant's perception of own health. Furthermore, we have mentioned in the discussion that symptom burden may not be the only important domain of illness severity, and that other domains such as impairment and mental morbidity also constitute important domains (p. 17).

Reviewer: 3

Reviewer Name: Win Sen Kuan

Institution and Country: National University Hospital and National University of Singapore, Singapore  
Please state any competing interests or state 'None declared': None declared

Thank you for the opportunity to review the statistical methods and analyses used in the manuscript entitled "The BDS checklist as measure of illness severity".

The methods employed to assess the structural validity and convergent validity of the Bodily Distress Syndrome checklist total sum score as a measure of physical symptom burden and functional somatic disorders illness severity are both appropriate and robust.

Minor comment:

There should be an explicit explanation on why "data on the primary care cohort did not allow us to investigate convergent validity to the aggregate score" (Methods, Page 6 Line 34) in the Results section.

Author's reply: We agree with the reviewer that this statement was not fully comprehensible. We have now elaborated on this issue in the 'Measures' section (p. 6).

Reviewer: 4

Reviewer Name: Chandrashekhar T Sreeramreddy Institution and Country: International Medical University, Nepal Please state any competing interests or state 'None declared': 'None declared'.

The questionnaire BDS has been validated using THREE large DUTCH Cohorts which are perhaps, socio-demographically as well with respect to health status very distinct from each other. Authors

however did not measure convergent validity of Dutch version BDS with other proxy questionnaire based measures. Even authors admitted this as one important limitation as was the self-report. The subjectivity and a lack of objective validation make the results much weaker than the authors' admittance. They had an opportunity to validate by physicians' report in primary care and specialist care setting.

Author's reply: We acknowledge that the lack of testing the convergent validity of the BDS checklist to other symptom questionnaires is a limitation of our study, as mentioned in the discussion section. However, as explained in the paper, we wanted to include equal measures for all three cohorts, and the measures and data were not completely consistent across cohorts. Hence, data did not include information on the most evident symptom questionnaires, e.g. PHQ-15 or SSS-8. We did not have access to the physician's report in primary care and specialist setting neither. It was therefore not possible for us to do this kind of test. We have elaborated more in this issue in the 'Strengths and weaknesses of the study' section (p. 16).

Methods are another weak section. Authors explained the statistical analyses very well indeed very candidly. However, the different statistics presented need to be explained either in methods or as parenthesis under results about acceptable values for interpretations of measures such as validity, alpha, correlation coefficients etc.

Author's reply: We respectfully believe that much of this has been addressed in the 'Validation procedure and statistical analyses' section: Here, fit indices for the CFA models are presented (p. 7) and our hypotheses, based on previous literature, are listed with values of acceptable convergent validity (p. 8). The item total correlation corrected for overlap indicates to which degree an item is different from the other items, but it does not come with a reference value. We acknowledge some lack of interpretation for the Cronbach's alpha and this is now added in the same section (p. 8).

Perhaps all results including the confidence intervals of most results that are presented may not be needed under results. Readers can conveniently refer to the tables.

Author's reply: We believe that the reviewer refers to the results section of 'Convergent validity' (p. 10) when proposing that some of the results may not be needed under the results section. We agree that this section is rather heavy. Unfortunately, the journal's guidelines on maximum allowed number of tables and figures (5) do not allow us to put these results into a separate table. A combination/rearrangement of tables would not be reader friendly, why we choose to present the results as main text.

How were the participants in different settings recruited, who, how and when the instruments were administered is an essential information under methods.

Author's reply: This study included already obtained data from three different cohorts including participants that were originally recruited for other studies (DanFunD, KOS, and the STreSS trials). We therefore, respectfully, believe that it is sufficient to explain this and refer to the original studies for further explanation about the recruitment and administration of the instruments, as has been done in the 'Population' section (p. 5).

Factor loading values are needed to be added in the appendix figures. The values inserted in the CFA figures will make the interpretations easy not having to refer to the tables once again. I recommended statistical review as I did not seem well qualified with limited experience doing such analyses to assess if the appropriate statistics were used.

Author's reply: The four CFA figures in the appendix do not constitute results from this paper. They are mainly theoretical graphical explanations of the four CFA models that were tested in the study. We included this appendix only as a help for readers in understanding the models. We have now changed the heading in the appendix so this gets clearer.

Reviewer: 5

Reviewer Name: Yue-Fang Chang

Institution and Country: University of Pittsburgh, USA Please state any competing interests or state 'None declared': None

1. What were included for the calculation of 'physical functioning'? The authors mentioned it was measured with an aggregate score of four items from the SF-36 subscales 'physical function', 'bodily pain', and 'vitality' (page 7). What was the 4th item (only 3 subscales were listed)?

Author's reply: The aggregate score was developed from 2 items from the 'physical function' subscale, 1 item from the 'bodily pain' subscale, and 1 item from the 'vitality' subscale. We understand if the reviewer found the description insufficient and it is now elaborated on in the 'Measures' section (p. 6).

2. Four confirmatory factor analysis models were examined to assess the structural validity in the study. On page 10 it stated, 'These models revealed correlations between the four BDS symptom clusters and loadings from an underlying BDS factor to the four BDS symptom clusters may imply a bi-factor model'. Please explain why correlated 4 symptom clusters and loadings from the underlying factor may suggest a bi-factor model.

Author's reply: In the original manuscript, we did elaborate on this in the discussions section (p. 15), where we list the indicators of the bi-factor model. We believe that this is what the reviewer is asking for, but if this is incorrect, please let us know. We are of course willing to move this section to the results or methods sections if wished by the reviewer.

3. Convergent validity of BDS sum score was evaluated by studying the association between BDS sum score and other measures using both non-parametric and parametric methods (i.e. Spearman correlation and linear regression). The distributions of BDS sum score were skewed in the general population cohort as well as in the primary care patient cohort. Was there any transformation applied to the BDS sum score prior to the linear regression analyses?

Author's reply: We want to thank the reviewer for this important question. As for any other linear regression analyses, normality of the outcome is not important but normality of the residuals is an important assumption. Testing the models, we found that the residuals were not normal distributed. However, the purpose with the linear regression analyses were to investigate (or quantify) the difference on the BDS checklist when a certain change was seen on the other scales and hold it together with our stated hypotheses as a kind of help for clinicians. Therefore, we did not perform any transformation, e.g. log-transformation, prior to the regression analyses as the interpretations of the results would not be clear or provide us with the desired answer. We do acknowledge, that this quantification should be interpreted with some caution due to the non-normality of the residuals. We have now mentioned this in the 'Strengths and weaknesses of the study' section (p. 16).

4. There were 72% of symptoms with higher loadings for the general BDS factor than for the 4-symptom clusters in the general population cohort; however, this number dropped to 52% in the specialized settings cohort. Please discuss why the specialized setting cohort have much lower %?

Author's reply: We agree that this aspect deserves some more discussion and have now elaborated more on this in the 'Principal findings' section (p. 15).

5. Please present median and inter-quartile range for the BDS sum score as well since the distributions were skewed for the general population and primary care cohort. Were the BDS sum scores of the specialized clinical setting statistically different from those of the general population?

Author's reply: Our Table 2 already contains the median and inter quartile ranges (25% perceptile, 50% perceptile, and 75% perceptile). We have now added text to the table and the table text for clarification (p.12-13).

We have also performed Wilcoxon Rank Sum tests, testing the difference between the BDS score in the three samples. These tests showed significant difference between scores in all three cohorts ( $p < 0.0001$ ). We did, however, not include these analyses to the paper as we believe this aspect already appears in the Table 2 BDS means, medians, and percentiles.

6. 'Error' was omitted from the statement 'A Root Mean Square of Approximation (RMSEA)' (page 8).  
Author's reply: The error has now been corrected (p. 7).

7. Please define 'item-rest correlation' as some readers may be unfamiliar with this term.  
Author's reply: We agree that this term might not be the most correct to use and have now replaced it with the more proper 'item total correlation, corrected for overlap' (p. 8, p. 11, and p. 12).

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Yue-Fang Chang University of Pittsburgh, USA
<b>REVIEW RETURNED</b>	04-Nov-2020
<b>GENERAL COMMENTS</b>	All my comments have been addressed satisfactorily.

#### VERSION 2 – AUTHOR RESPONSE

Reviewer Comments to Author:

Reviewer: 5

Reviewer Name: Yue-Fang Chang

Institution and Country: University of Pittsburgh, USA Please state any competing interests or state 'None declared': None declared.

Comments to the Author

All my comments have been addressed satisfactorily.

Author's response: We are pleased that we have addressed the reviewer's comments satisfactorily.