Title: The Four-Dimensional Symptom Questionnaire (4DSQ). A validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization.

Authors:

Berend Terluin (b.terluin@vumc.nl)
Harm WJ van Marwijk (hwj.vanmarwijk@vumc.nl)
Herman J Ader (hjader@keyaccess.nl)
Henrica CW de Vet (hcw.devet@vumc.nl)
Brenda WJH Penninx (brendap@ggzba.nl)
Marleen LM Hermens (mlm.ermens@tiscali.nl)
Christine A van Boeijen (c.van.boeijen@degelderseroos.nl)
Anton JLM van Balkom (tonb@ggzba.nl)
Jac JL van der Klink (j.vanderklink@nsph.nl)
Wim AB Stalman (w.stalman@vumc.nl)

Version: 2 Date: 18 July 2006

Author's response to reviews: see over
Dear Dr. Hodgkinson-Barrett, dear Editorial team,

We would like to thank BMC Psychiatry for the opportunity to revise our paper. The first referee, Dr. Müller, did not have any further comments. However, the second referee, Dr. Ferdinand, has put forward a number of specific comments. We would like to thank him for the valuable suggestions, that we have dealt with as explained below.

1. Dr. Ferdinand asked about the significance of the differences between the AUCs. We have performed the significance tests where appropriate and reported them in the paper (pp. 8, 13-14).

2. Dr. Ferdinand questioned the criterion of “any psychosocial diagnosis/reason for encounter” as a valid outcome criterion for distress because of its doubtful reliability. We have clarified the use of this criterion by adding the following text on p. 8. “Since a true valid and reliable gold standard for distress is not available, we adopted “any psychosocial diagnosis/reason for encounter” as established by the GP as the “criterion” for the 4DSQ Distress score. This criterion should represent a reasonable indicator of distress, with relevance for primary care. Given the unknown reliability of the criterion, any relationship with the 4DSQ Distress score can be interpreted as supporting validity of the 4DSQ Distress scale.”

3. Dr. Ferdinand signalled that the Depression and Distress score performed equally well in predicting major depression. We have discussed this finding in the Discussion section on p. 18: “…These findings may indicate that the 4DSQ Depression scale measures more severe depressive disorders, whereas the Distress scale measures milder depressive disorders. Indeed, we found that the 4DSQ Depression and Distress scores performed almost equally well in detecting the whole range (from mild to severe) of DSM-IV major depressive disorders. However, the Distress score did not add any predictive power to the Depression score in the logistic regression analysis. Therefore, the 4DSQ Depression score seems to be sufficient to detect depressive disorders.”
4. Dr. Ferdinand doubted the validity of the Anxiety scale, given its low AUC. We have discussed this finding in more detail in the Discussion section p. 19: “…It is this “abnormal” anxiety, which is characteristic of anxiety disorders, that is measured by the 4DSQ Anxiety scale. Yet, we failed to demonstrate unequivocal criterion validity of the Anxiety scale with respect to standardised DSM-IV anxiety disorder diagnoses. In retrospect, we suspect that study E was not ideal for investigating the criterion validity of the Anxiety scale because the patients were highly selected (for anxiety), leaving little contrast between patients with and without an anxiety disorder. More research, with more heterogeneous study samples, is needed to establish the criterion validity of the 4DSQ Anxiety scale.

5. Dr. Ferdinand also doubted the appropriateness of “GP’s suspicion of somatization” as a criterion for somatization. Further, Dr. Ferdinand questioned the validity of the Somatization scale given its low AUC.

We have clarified the use of this criterion by adding the following text on p. 9. “This criterion is not meant to be a true gold standard for somatization, but rather a reasonable indicator of somatization (i.e. the presentation of somatic complaints unexplained by physical illness), relevant for primary care. Given the unknown reliability of the criterion, any relationship with the 4DSQ Somatization score can be interpreted as supporting validity of the 4DSQ Somatization scale.”

We have discussed the implications of the low AUC in the Discussion section (p. 19): “Although the relatively low AUC (0.65) indicates that the Somatization score is not a perfect predictor of the suspicion of somatization by GPs, the association between the Somatization score in patients and the judgement by GPs is interesting. Apparently, GPs recognise something that is expressed through elevated Somatization scores by patients. It should be realised at this point that the GPs and patients probably were not able to discuss any psychosocial issues, because otherwise the GPs would have established a psychosocial diagnosis. Thus, the GP’s assessment of the presence of a psychosocial background in patients with somatic symptoms must have been the result of a rather subjective process with questionable reliability. Assuming a relatively low reliability of our “criterion” for somatization, and considering that validity can never surpass reliability, an AUC-value of 0.65 may not be too bad at all.”

6. Dr. Ferdinand commented on the high correlations between the 4DSQ scales in Table 13, stating that the findings specifically cast doubts on the usefulness of the Distress scale.
Furthermore, dr. Ferdinand wondered how it can be explained that elevated Distress scores were not always accompanied by high Anxiety and Depression scores. Indeed, the correlations between the Distress score, on the one hand, and the Depression, Anxiety and Somatization scores, on the other hand, are high. The scatterplots, shown in Figure 5, show that a low Distress score is highly predictive of the Depression and Anxiety scores (which must be low), but a high Distress score is practically not predictive of Depression and Anxiety scores (which can be anywhere between low and high). To a large extent this peculiar relationship between the 4DSQ scores is responsible for the high correlations between Distress on the one hand and Depression, Anxiety and Somatization on the other hand. (This can be tested by computing two random uncorrelated variables and selecting those cases in which the values of one (Y-)variable do not exceed the other (X-)variable. The correlation between X and Y will be about 0.50. The correlation will even be higher when through tranformations the number of observations in the lower-left corner is increased, mimicking our data.)

We have added more explanation on this issue on page 20.

7. Dr. Ferdinand discovered that the Figures were numbered the wrong way. We apologise for this, and we have corrected the numbers. There were only 6 Figures.

8. With respect to Appendix 1, dr. Ferdinand asked why not the loadings of all items on all factors were presented. The reason for this is that these figures were produced by confirmatory factor analysis, using the programme EQS. Such an analysis only produces factor loadings for item-factor relationships included in the model. The evaluation of the model relies on the goodness-of-fit measure (i.e. the comparative fit index, CFI).

9. Because we concluded (in retrospect) that study E may not have been ideal for studying the criterion validity of the Anxiety scale, dr. Ferdinand suggested to leave it out. However, we had not anticipated this problem in study E, and it was only after the results became known to us, that we figured that the special selection of study E made it less ideal (or practically unsuitable) for our purpose. Leaving out this part of the study because of the disappointing results would be a form of selective publishing, something we with dr. Ferdinand would not want to support.
10. We followed dr. Ferdinand’s suggestion to expand the discussion of the validity of the 4DSQ scales (pp. 17-20). Although this has lead to more differentiated conclusions, at the end we still conclude that the 4DSQ scales are valid.

11. Finally, considering the critical remarks of dr. Ferdinand about the practical applications section, we have decided to omit this section. Some practical applications are mentioned in the Discussion section where the data provide sufficient support for these applications.