Author’s Reply

Statistics and pain-related fear measures in acute low back pain

We would like to thank Dr. Nijs and Dr. Truyen for their interest in and careful reading of our paper entitled “Psychometric properties of the Tampa Scale for Kinesiophobia and the Fear-Avoidance Beliefs Questionnaire” (Swinkels-Meewisse et al., 2003). Dr. Nijs and Dr. Truyen expressed concerns about the statistical analyses that have been used in the study. In this response, we would like to comment on these concerns in a point-by-point reply.

The first concern is that items of the TSK-DV and FABQ are scored on a Likert scale and individual items as well as total scores should be considered ordinal data. Consequently, the Spearman rank correlation coefficient should have been used to analyze concurrent validity of both measures, rather than the Pearson product–moment correlation, of which the application is only restricted to interval and ratio scales. We agree that individual items of the TSK-DV and the FABQ should be considered ordinal data. Measurement on an ordinal scale involves rank-ordered categories. An interval scale also has this rank-order characteristic but also assumes equal distances or intervals between units of measurement (Portney and Watkins, 2000). For an individual item, it would perhaps be difficult to assume that a difference between 1 and 2 is the same as the difference between 3 and 4. However, none of the analyses have relied on scores on individual items, but all have been conducted on total scores (TSK) or total scores of subscales (FABQ) of these measures. It is not possible to statistically test whether total scores could be considered as interval data, but we can only assume that each item is equally important in the composition of the total score and subsequently assume that we may treat total scores of the TSK-DV and FABQ as continuous variables measured at an interval level. In reanalyzing our data, we obtained an ICC for the TSK-DV of 0.78 (95% confidence interval: 0.71–0.83). For two total scores on the FABQ/w and FABQ/pa subscales we found an ICC of 0.80 (95% confidence interval: 0.74–0.85) and 0.64 (95% confidence interval: 0.55–0.72), respectively. The ICC estimates are almost identical to the Pearson product moment correlation coefficient reported in our study.

The fourth concern relates to the internal consistency of the TSK-DV and the FABQ. We agree that measurement instruments with a Cronbach’s alpha below 0.80 are considered less reliable. As correctly pointed out by Dr. Nijs and Dr. Truyen, Cronbach’s alpha is affected by the number of items in a scale. Future studies could extend these measures with appropriate items.

In the final concern, Dr. Nijs and Dr. Truyen stress the importance of reporting the 95% confidence intervals of TSK-DV and FABQ scores for evaluating interventions as they may be useful for determining whether a change in the scores on these measures can be
considered clinically relevant. We agree with their notion but a greater time interval between both test occasions than used in our study, should be applied to address this concern. Future studies may focus on the responsiveness of the TSK-DV and the subscales of the FABQ. Responsiveness refers to the ability of an instrument to detect change over time, which is important in both clinical practice and research (Sim and Waterfield, 1997). To assess the responsiveness of an instrument an external criterion is necessary to determine whether real change has occurred or not (De Vet et al., 2001). Responsiveness can be quantified by either comparing change scores with the corresponding smallest real difference (SRD) or by receiver operating characteristics curves (ROC curves) (see De Vet et al., 2001 for more details). We hope that this response sufficiently clears up the concerns raised by Dr. Nijs and Dr. Truyen.

References


E.J.C.M. Swinkels-Meewisse
Ulenpas 80 5655 JD Eindhoven,
The Netherlands
E-mail address: swinky@xs4all.nl

J. Roelofs

R.A.H.M Swinkels

A.L.M. Verbeek

J.W.S. Vlaeyen

R.A.B. Oostendorp