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

Swinkels-Meewisse et al. (2003) examined some of the psychometric properties of both the Dutch Tampa Scale for Kinesiophobia (TSK-DV) and the Fear-Avoidance Beliefs Questionnaire (FABQ) in acute low back pain patients. We congratulate the authors for their work; early identification of acute (low back) pain patients at risk of becoming chronic may be of prime importance to reduce both economic costs and patients’ disability. In this view, the usage of a reliable and valid measure to assess pain-related fear (of movement) should be encouraged. However, we are concerned about the statistical analysis.

Since the items on both the TSK-DV and the FABQ are scored on a Likert scale, they generate ordinal data (even the total scores are ordinal data). A Pearson correlation analysis is therefore inappropriate for both the analysis of the test–retest data and the examination of the concurrent validity of both measures. The Pearson product-moment correlation coefficient is the most commonly reported measure of correlation, but its area of application is restricted to the interval and ratio scales (Portney and Watkins, 2000a). The Spearman rank correlation coefficient should be used with ordinal data (Portney and Watkins, 2000a), and would therefore have been appropriate to analyse the concurrent validity of both measures.

Second, the descriptive data of the total scores obtained with the TSK-DV and the FABQ were not reported adequately: the median and the interquartile range are the appropriate descriptive statistics for ordinal data, but Swinkels-Meewisse et al. (2003) reported only the means, medians and standard deviations of the total scores of both the TSK-DV and the FABQ (on p. 32 and 33). The authors were not the first to do so, previous research using the TSK(-DV) consistently reported mean and standard deviations for the total scores on the TSK(-DV) (Vlaeyen et al., 1995; Goubert et al., 2000; Silver et al., 2002).

Furthermore, neither a Spearman correlation, nor a Pearson correlation analysis, would have been appropriate for the analysis of the test–retest data. Correlation tells us how the scores vary together, but it cannot tell us the extent of agreement between the two sets of measurement (Portney and Watkins, 2000b). If a test is reliable, then the performance of the test on different occasions should yield (nearly) identical results (Fritz and Wainer, 2001), and not just results that vary together. Two sets of measurement can show perfect correlation, while at the same time have poor reliability. Assuming that Swinkels-Meewisse et al. (2003) aimed at examining the agreement between the two sets of measurement (as is most often the essence of reliability in clinical and research settings), the reported Pearson correlation coefficients may have generated irrelevant data (that is, irrelevant to the aims of their study). Rehabilitation medicine has its own measurement ‘bible’ (Dijkers et al., 2002), which clearly indicates that for the analysis of test–retest data of ordinal data, the kappa statistic (κ) and the intraclass correlation coefficient are the appropriate statistics (Johnston et al 1992). Moreover, these measurement standards indicate that if a Pearson correlation is used to describe reliability, then central tendency differences also need to be examined (Johnston et al 1992). No such examination of central tendency differences is reported in the manuscript. The intraclass correlation coefficient, reflecting both correlation and agreement (Portney and Watkins 2000b), should have been used to examine the test–retest reliability.

Fourth, one can question the internal consistency of the items of both the TSK-DV (0.70 and 0.76) and the fear-avoidance beliefs about physical activities subscale of the FABQ (0.70 and 0.72). An internal consistency below the often-recommended cut-off of Cronbach’s alpha of 0.80 is considered to be indicative of poor reliability (Dijkers et al., 2002). The fact that these Cronbach’s Alpha values are similar to those reported by Vlaeyen et al. (1995), as indicated by the authors, does not change anything and is somewhat misleading. Stating that short questionnaires are less reliable would have provided the reader with more objective information.

Fifth, the 95% limits of agreement between the two data sets (test and retest) (as reported in the Silver et al., 2002 trial) would have been of interest to the manual therapist who is trying to reduce pain-related fear in acute low back pain patients. The authors themselves indicate that the manual therapist, in acute low back pain patients exhibiting high pain-related fear, should try to focus the intervention on the fear rather than on impairments of anatomical structures. In this view, exposure in vivo to a set of individually tailored and fear eliciting physical movements has shown promising
preliminary results in patients with chronic low back pain (Vlaeyen et al., 2002; de Jong et al., 2002). In order the track possible changes in the scores obtained with the TSK-DV or the FABQ and relate these changes to the intervention, the 95% confidence intervals may assist the practitioner in assessing the efficacy of the intervention. Assuming that we have obtained both a baseline (initial visit) and post treatment score for the TSK-DV, we subtract the post treatment from the pretreatment score to obtain a change in score. The observed change in TSK-DV score should exceed the 95% confidence intervals to be considered clinically significant (for more details: Selfe et al., 2001).

References


Jo Nijs
Department of Human Physiology, Faculty of Physical Education and Physical Therapy Vrije Universiteit Brussel (V.U.B.), Brussel 1090, Belgium
Division of Occupational and Physical Therapy, Hogeschool Antwerpen, Belgium
E-mail address: jo.nijs@vub.ac.be

Steven Truyen
Department of Health Sciences, Division of Occupational and Physical Therapy, Hogeschool Antwerpen, Belgium