



CORRESPONDENCE

## Response to “The neurogenic bowel dysfunction score in patients with spinal cord injury: methodological issues in reliability and validity”

Didem Erdem<sup>1</sup> · Selmin Gülbahar<sup>1</sup> · Pembe Keskinoglu<sup>2</sup>

Received: 3 October 2017 / Accepted: 12 October 2017 / Published online: 18 January 2018  
© International Spinal Cord Society 2018

We appreciate Dr. Sabour’s interest in our article “Reliability, validity and sensitivity to change of neurogenic bowel dysfunction score in patients with spinal cord injury”. The present study aimed to investigate the validity, reliability and sensitivity to change of the Turkish version of the Neurogenic Bowel Dysfunction (NBD) score. According to the Correspondence in *Spinal Cord* by Dr. Sabour, there are limitations of the statistical and methodological approach of this study.

Sabour suggested that we correctly applied Cronbach’s alpha coefficient to determine internal consistency, but assessing the test-retest reliability with Pearson’s correlation coefficient was not appropriate because Pearson correlation coefficient assesses the linearity and gives a linear correlation. Although the Pearson’s correlation coefficient has been the most common technique for assessing reliability [1, 2], using the Pearson’s correlation coefficient for assessing test-retest reliability is typically discouraged [3]. The primary limitation in using the Pearson’s correlation coefficient is that it cannot detect the existence of systematic error [3, 4]. Second, it is difficult to determine the test-to-test variation when multiple tests are administered [4]. However, this recommendation for Pearson correlation coefficient is not universal. Rousson et al. [5] suggested that the usual notion of Pearson correlation is well adapted in a test-retest situation, whereas intraclass correlation coefficient should be used for intra-rater and interrater reliability. The critical difference between these two approaches is the

treatment of systematic error, which is often due to a learning effect for test-retest data.

Furthermore, Sabour referred two weaknesses of kappa. Kappa can be used to compute reliability for categorical data or data involving dichotomous decisions [6]. Several statistical tools are available to examine agreement that refers to interchangeability among raters or assessment instruments, including kappa and weighted kappa and various chance-corrected measures of proportion or percent agreement [1]. According to these data, in this study, three main statistics were computed in examining reliability: Pearson correlation coefficient, Kappa coefficient, and Cronbach’s alpha coefficient. We assessed the reliability of NBD score by test-retest reliability and internal consistency, and also the consistency of frequency distribution of all answers for each item in the test-retest application by kappa statistics. There are other studies evaluating reliability with the Pearson correlation coefficient, as well as the Kappa coefficient [7, 8].

Sabour has reported that another methodological issue in our study was the difference between clinical importance and statistical significance. We have already emphasised the strength of the correlation instead of statistical significance. All of the  $r$ -values and  $p$ -values of the correlations between NBD score and SF-36 subscales have been reported in the results section. The correlations of NBD score with physical functioning subscale and physical role functioning subscale were very low, and  $p$ -values were also not significant. Based on the results, we noted that NBD score had a moderate significant negative correlation with mental subgroups of SF-36 and no significant correlation with physical subgroups of SF-36.

Sabour suggested that our conclusion of the Turkish version of the NBD score is a valid and reliable instrument might be a misinterpretation of the results. We do not agree with his suggestion.

We examined internal consistency in addition to test-retest reliability to determine reliability. NBD score is not a Likert-type scale but the additivity of the total score

✉ Didem Erdem  
didem\_blue86@hotmail.com

<sup>1</sup> Department of Physical Medicine and Rehabilitation, Dokuz Eylül University Faculty of Medicine, Izmir, Turkey

<sup>2</sup> Department of Biostatistics and Medical Informatics, Dokuz Eylül University Faculty of Medicine, Izmir, Turkey

was tested with Tukey's nonadditivity test, and the additivity characteristic was demonstrated. Due to the presence of additivity characteristic of the test and ordinal characteristic of the answers, Cronbach's alpha coefficient was calculated to determine the internal consistency. Cronbach's alpha coefficient for internal consistency was 0.547. This result indicates that the components of the score do not have a high degree of internal homogeneity. In the discussion section, we suggested that the reasons behind this result could be the fact that NBD score evaluates both incontinence and constipation, does not contain subscales, and has a small number of questions. If the items in a test are correlated to each other, the value of alpha is increased. However, a high Cronbach's alpha coefficient does not always mean a high degree of internal consistency because alpha coefficient is also affected by the length of the test. If the test length is too short, the value of alpha is reduced [9].

Cronbach's alpha coefficient was not high, but the test-retest reliability of the Turkish version of NBD score was high. According to these results, we suggested that *it revealed acceptable reliability*.

For validation analysis, no gold standard test is available to be used to determine the criterion validity of NBD score and also the data structure of the test is not appropriate for factor analysis. Thus, construct validity of NBD score was assessed with the correlations between the score and SF-36, Physician Global Assessment and patient assessment of the impact on the quality of life (QoL). The NBD score did not correlate with all subscales of SF-36. The possible causes of this result were discussed, and all the limitations were reported in the discussion section. On the other hand, the NBD score correlated with patient assessment of the impact on QoL and Physician Global Assessment. These results support the construct validity of the score.

In conclusion, we suggest that the Turkish version of the NBD score is a valid and reliable instrument in patients with spinal cord injury according to the measurements in this sample, but it might be verified by other studies. Once again we thank Dr. Sabour for his interest in our study and his view on the limitations of our publication in Spinal Cord. We also thank the editor for the opportunity to write this response.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

1. Ottenbacher KJ. An examination of reliability in developmental research. *J Dev Behav Pediatr.* 1995;16:177–82.
2. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26:217–38.
3. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19:231–40.
4. Yen M, Lo LH. Examining test-retest reliability: an intra-class correlation approach. *Nurs Res.* 2002;51:59–62.
5. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater, and test-retest reliability of continuous measurements. *Stat Med.* 2002;21:3431–46.
6. Ottenbacher KJ, Tomchek SD. Reliability analysis in therapeutic research: practice and procedures. *Am J Occup Ther.* 1993;47:10–6.
7. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. *J Psychosom Res.* 2002;53:737–40.
8. Silverman WK, Saavedra LM, Pina AA. Test-retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for DSM-IV: child and parent versions. *J Am Acad Child Adolesc Psychiatry.* 2001;40:937–44.
9. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53–5.