



CORRESPONDENCE

The neurogenic bowel dysfunction score in patients with spinal cord injury: methodological issues in reliability and validity

Siamak Sabour^{1,2}

Received: 11 September 2017 / Accepted: 10 October 2017 / Published online: 18 January 2018
© International Spinal Cord Society 2018

I was interested to read the paper by Erdem et al. [1] published in *Spinal Cord*, July 2017. The purpose of the authors was to investigate the validity and reliability of neurogenic bowel dysfunction (NBD) score. The study included 42 patients with spinal cord injury. Although they correctly applied Cronbach's alpha coefficient to determine internal consistency, in test–retest reliability analysis, they reported a high correlations between test–retest total NBD score and also test–retest answers of each question ($r = 1.000$, $P < 0.001$). Pearson r just assesses the linearity and we can have a linear correlation with no reliability at all. Any shift in location and scale of the regression line, which indicates no reliability cannot be detected by Pearson r . Moreover, in reliability analysis, our approach should be individual based instead of global average and Pearson r cannot cover this approach [2, 3].

They also reported that the consistency of frequency distribution of all answers for each item was analyzed by kappa statistics and very high consistency was found ($\kappa = 1.000$, $P < 0.001$). It revealed acceptable reliability [1]. It is important to know that two important weaknesses of kappa are as follows. First, it depends on the prevalence in each category, which means it can be possible to have different k values having the same percentage for both concordant and discordant cells. Table 1 shows that in both (a) and (b) situations, the prevalence of concordant cells is 90% and of discordant cells, 10%; however, we get different kappa values (0.44 as moderate and 0.80 as very good, respectively). Kappa value also depends on the number of categories [2, 3].

Table 1 Limitation of kappa for comparison of two observers' diagnoses with different prevalence in the two categories

Situation (a)		Observer 1		
		Positive result	Negative result	Total (%)
Observer 2	Positive result	85	5	90
	Negative result	5	5	10
	Total	90	10	100
Situation (b)		Observer 1		
		Positive result	Negative result	Total (%)
Observer 2	Positive result	45	5	50
	Negative result	5	45	50
	Total	50	50	100

$K = 0.44$

$K = 0.80$

Another methodological issue is differentiation between clinical importance and statistical significant. They reported that a statistically significant negative correlation was detected between bodily pain ($r = -0.3$, $P = 0.01$), general health ($r = -0.5$, $P < 0.001$), vitality ($r = -0.6$, $P < 0.001$), social role functioning (SF) ($r = -0.7$, $P < 0.001$), emotional role functioning ($r = -0.6$, $P < 0.001$), and mental health ($r = -0.6$, $P < 0.001$) subscales, whereas no significant correlation was found with physical functioning (PF) ($r = -0.2$, $P = 0.13$) and physical role functioning (RP) ($r = 0.06$, $P = 0.67$) subscales of SF-36 [1]. In both positive and negative association, we should emphasize on the strength of the association instead of statistical significant because the second issue dramatically depends on the sample size.

Finally they reported that Cronbach alpha coefficient for internal consistency of NBD score was 0.54. However, test–retest reliability of the Turkish version of NBD score was high. They concluded that the Turkish version of the NBD score is a valid and reliable instrument. Such conclusion may be a misinterpretation of the results [2, 3].

✉ Siamak Sabour
s.sabour@sbmu.ac.ir

¹ Department of Clinical Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

² Safety Promotions and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

In conclusion, to assess reliability and validity, appropriate test as well as correct interpretation should be considered. In this letter, I discuss the limitations of their statistical and methodological approach and the lack of evidence for such a sweeping conclusion. Their conclusion should be supported by the above mentioned statistical and methodological issues. Otherwise, misdiagnosis and mismanagement of the patients cannot be avoided.

Compliance with ethical standards

Conflict of interest The author declares that he has no competing interests.

References

1. Erdem D, Hava D, Keskinoglu P, Bircan Ç, Peker Ö, Krogh K et al. Reliability, validity and sensitivity to change of neurogenic bowel dysfunction score in patients with spinal cord injury. *Spinal Cord*. <https://doi.org/10.1038/sc.2017.82>. (2017).
2. Sabour S. Reproducibility of dynamic Scheimpflug-based pneumotonometer and its correlation with a dynamic bidirectional pneumotonometer device: methodological issues. *Cornea*. 2015;34: e14–5.
3. Sabour S. Spinal instability neoplastic scale: methodologic issues to avoid misinterpretation. *AJR Am J Roentgenol*. 2015;204: W493.