CORRESPONDENCE

Reliability of the International Spinal Cord Injury Musculoskeletal Basic Data Set; methodological and statistical issue to avoid misinterpretation

We were interested to read the paper by Baunsgaard CB and colleagues[1] published in the May 2016 issue of *Spinal Cord*. The authors aimed to determine the intra- and inter-rater reliability of the International Spinal Cord Injury (SCI) Musculoskeletal Basic Data Set (ISCIMSBDS). Kappa statistics (ranged from $\kappa = 0.62$ to 1.00) was used to measure reliability.[1] Reliability (precision) is an important methodological issue. For qualitative variables, using simple kappa is among common mistakes in reliability analysis. Regarding reliability (precision, repeatability or reproducibility) for qualitative variables, weighted kappa should be used with caution because kappa has its own limitation.[2–8] Two important weaknesses of *k*-value to assess the agreement of a qualitative variable are as follows: it depends upon the prevalence in each category, which means it is possible to have different kappa value having the same percentage for both concordant and discordant cells! Figure 1 shows that in both (a) and (b) situations, the prevalence of concordant cells are 80% and discordant cells are 20%, however, we get different kappa value (0.38 as fair and 0.60 as moderate-good) respectively. Kappa value also depends upon the number of categories, which means that higher the categories lower the kappa value.[2–8] Therefore, reporting weighted kappa can be highly recommended.

They reported that the crude agreement ranged from 75 to 100% for each of the variables on the ISCIMSBDS.[1] Regarding reliability, it is crucial to know that an individual-based approach instead of group-based (crude agreement) should be considered.[2–9] The reason is in reliability assessment; we should consider individual results and not global average. In other words, possibility of getting exactly the same crude agreement of a variable between methods with no reliability at all is high.[2,3,9]

As the authors pointed out in their conclusion, overall, the ISCIMSBDS is reliable. Such conclusions may be a misleading message due to inappropriate use of statistical tests. In conclusion, for reliability analysis, appropriate tests as well as correct interpretation should be applied. Otherwise, misdiagnosis and mismanagement of the patients cannot be avoided.

COMPETING INTERESTS

The authors declare no conflict of interest.

Siamak Sabour[1,2] and Fariba Ghassemi[3]
[1]Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran;
[2]Department of Clinical Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran and
[3]Department of Retina & Ocular Oncology, School of medicine, Farabi Eye hospital, Tehran University of Medical sciences, Tehran, Islamic Republic of Iran
Correspondence: S Sabour (s.sabour@sbmu.ac.ir)

REFERENCES

1 Baunsgaard CB, Chhabra HS, Harvey LA, Savic G, Sisto SA, Qureshi F et al. Reliability of the International Spinal Cord Injury Musculoskeletal Basic Data Set. *Spinal Cord.*, e-pub ahead of print 3 May 2016; doi: 10.1038/sc.2016.

2 Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 4th edition, Baltimore, United States: Lippincott Williams & Wilkins, 2010.

3 Szklo M, Nieto FJ. *Epidemiology Beyond The Basics*. 2nd edition, Jones and Bartlett Publisher: Manhattan, New York, NY, USA, 2007.

4 Sabour S. Reliability of a new modified tear breakup time method: methodological and statistical issues. *Graefes Arch Clin Exp Ophthalmol* 2016; **254**: 595–596.

5 Sabour S. Spinal instability neoplastic scale: methodologic issues to avoid misinterpretation. *AJR Am J Roentgenol* 2015; **204**: W493.

6 Sabour S. Reproducibility of dynamic Scheimpflug-based pneumotonometer and its correlation with a dynamic bidirectional pneumotonometry device: methodological issues. *Cornea* 2015; **34**: e14–e15.

7 Sabour S. Reliability of automatic vibratory equipment for ultrasonic strain measurement of the median nerve: common mistake. *Ultrasound Med Biol* 2015; **41**: 1119–1120.

8 Sabour S. Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder? *Clin Radiol* 2015; **70**: 333–334.

9 Sabour S. Validity and reliability of the 13C-methionine breath test for the detection of moderate hyperhomocysteinemia in Mexican adults; statistical issues in validity and reliability analysis. *Clin Chem Lab Med* 2014; **52**: e295–e296.

a

Observer 2		Observer 1		
		Positive	Negative	Total
Observer 2	Positive	70	10	80
	Negative	10	10	20
	Total	80	20	100

K=0.38

b

Observer 2		Observer 1		
		Positive	Negative	Total
Observer 2	Positive	40	10	50
	Negative	10	40	50
	Total	50	50	100

K=0.60

Figure 1. Comparison of two observers' diagnosis with different prevalence in the two categories.