Supportive Care in Cancer (2019) 27:2339–2340
https://doi.org/10.1007/s00520-019-4641-7

LETTER TO THE EDITOR

Inter-rater reliability of the Oral Assessment Guide for oral cancer patients between nurses and dental hygienists: methodological issue

Soodeh Shahsavari[1] · Mehdi Naderi[2]

Received: 27 August 2018 / Accepted: 7 January 2019 / Published online: 16 January 2019

Dear Editor,

We were interested to read an article that was recently published by Aoki T and colleagues in the Aug 2018 issue of Support Care Cancer [1]. The purpose of the authors was to investigate the OAG's inter-rater reliability between nurses and dental hygienists to identify any limitations in performing objective oral assessments [1]. The inter-rater reliability was assessed with Cohen's k coefficient. The eight assessment categories of OAG tool (voice, swallowing, lips, tongue, saliva, mucous membrane, gingiva, and teeth/dentures) were investigated by a nurse and a dental hygienist. They reported that oral health status of all patients was poor, and none were given the worst score in the mucous membrane or gingiva categories. Also, the tongue, saliva, mucous membrane, gingiva, and teeth/denture categories had low κ coefficients, indicating poor nurse–dental hygienist inter-rater reliability. In contrast, the κ coefficients and agreement rates for voice and swallowing were high. Dental hygienists' scores were significantly higher for the tongue, gingiva, and teeth/denture categories than were nurses' scores [1].

To assess agreement of a qualitative variable, using kappa coefficient is not always appropriate and is not reliable in some circumstances. First, kappa value depends on the prevalence in each category. Second, kappa value also depends on the number of categories [2–6]. We should mention that when a variable with more than two categories or an ordinal scale is used (with three or more ordered categories), then the weighted kappa would be a good choice of association rather than agreement measure. Finally, the third important flaw is when the two raters have unequal marginal distributions of their responses [2–6]. Table 1 shows the agreement by applying kappa (0.43 as moderate) and weighted kappa (0.63 as good) which has different values and consequently different interpretations. In this table, the marginal distribution in the first category (grade 1) is different from the other categories and also, the number of categories is more than two categories.

Authors concluded that this study showed low nurse–dental hygienist inter-rater reliability for the OAG and highlighted the difficulties in objectively assessing patients' symptoms and oral health conditions. Accordingly, rather than only relying on an objective assessment of symptoms by a clinician, assessments should also include patients' subjective reporting of symptoms.

In this letter, we discussed two important limitations of the kappa value to assess reliability [2–6]. Any conclusion in reliability analyses needs to be supported by the methodological and statistical issues mentioned above.

✉ Mehdi Naderi
m.naderi51@yahoo.com

[1] Department of Health Information Management, School of Paramedical, Kermanshah University of Medical Sciences, Kermanshah, IR, Iran

[2] Department of Operating Room, School of Paramedical, Kermanshah University of Medical Sciences, Doulat Abad. Blvd, Kermanshah, IR 671985135, Iran

Table 1 The kappa and weighted kappa values for calculating agreement between 2 observers for more than 2 categories and depend on prevalence

Observer 1					Sum
	Grade	1	2	3	
Observer 2	1	60	20	1	81
	2	2	12	4	18
		3	11	11	25
Sum	3	65	43	16	124
		Estimate	Lower	Upper	
Kappa		0.43	0.31	0.55	
Weighted kappa		0.63	0.52	0.74	

Author contributions SS and MN conceived and designed the study protocol. SS and MN wrote the first draft of the manuscript. SS and MN reviewed and revised the manuscript and produced the final version.

Data Availability None.

Compliance with ethical standards

Ethics approval and consent to participate Not applicable.

Competing interests The authors declare that they have no competing interests.

References

1. Aoki T, Kudo M, Endo M, Nakayama Y, Amano A, Naito M, Ota Y (2018) Inter-rater reliability of the Oral Assessment Guide for oral cancer patients between nurses and dental hygienists: the difficulties in objectively assessing oral health. Support Care Cancer. https://doi.org/10.1007/s00520-018-4412-x
2. Szklo M, Nieto FJ (2014) Epidemiology beyond the basics, 3nd edn. Jones and Bartlett Publisher, Manhattan
3. Jalalvandi F, Naderi M (2018) Methodological issue on mixed-methods study of reported clinical cases of undesirable events, medical errors, and near misses in health care. J Eval Clin Pract 24:784. https://doi.org/10.1111/jep.13006
4. Naderi M (2018) Letter to the editor on "Is Regular Knee Radiograph Reliable Enough to Assess the Knee Prosthesis Position?". J Arthroplast 33:3591. https://doi.org/10.1016/j.arth.2018.07.006
5. Naderi M, Sabour S (2018) Reproducibility of diagnostic criteria associated with atypical breast cytology: a methodological issue. Cytopathology 29:396. https://doi.org/10.1111/cyt.12560
6. Naderi M, Jalalvandi F, Fatahi S (2018) Observer reproducibility of breast strain elastography in data acquisition and interpretation: a methodological issue. Eur J Radiol 106:210. https://doi.org/10.1016/j.ejrad.2018.05.001