American Cleft Palate–Craniofacial Association

Perspectives/Letters to the Editor/Editorials

The Cleft Palate-Craniofacial Journal
1-2

DOI: 10.1177/1055665620943429
journals.sagepub.com/home/cpc
SAGE

Nasalance-Based Preclassification of Oral–Nasal Balance Disorders Results in Higher Agreement of Expert Listeners: Methodological Issue

Mehdi Naderi, MSc[1], **and Siamak Sabour, MD, PhD**[2,3]

We were interested to read an article that recently published by de Boer G and colleagues in the April 2020 issue of *Cleft Palate Craniofac J* (de Boer et al., 2020). Evaluation of nasalance-based preclassification of oral–nasal balance disorders for improve listener agreement was the aim of the authors' study. Fifty-four records of patients with repaired unilateral cleft lip and palate (UCLP) assessed by speech-language pathologists participated as expert listeners. Two listening experiments were based on nasalance scores and audio recordings of speakers with repaired UCLP. The speakers were preclassified as normal, hypernasal, hyponasal, or mixed based on their nasalance scores. Initially, the listeners determined the diagnostic category of the oral–nasal balance for 62 audio recordings (8 repeats). The kappa statistic (κ) was applied to assess intralistener, interlistener, and agreement between listener categories. The authors reported that the agreement for the diagnostic category (54%, $\kappa = 0.37$) was between listeners 1 and 3. Listener 1 and listener 2 agreed for 26% ($\kappa = 0.10$) of the recordings, and listener 2 and listener 3 agreed for 30% ($\kappa = 0.10$) of the recordings. In terms of inter listener agreement, listeners 1 and 2 agreed on 31 (96.9%) of the 32 nasality categories ($\kappa = 0.96$), listeners 1 and 3 agreed on 25 (78.1%, $\kappa = 0.70$), and listeners 2 and 3 agreed on 26 (81.3%, $\kappa = 0.75$) of the categories. (de Boer et al., 2020).

There are methodological issues in the reliability assessment that can affect the outcome or main message of the study. One of the drawbacks is the use of κ, which in certain circumstances can affect the results of the study for the following reasons. First, the amount of κ depends on the prevalence in each group. Second, it also depends on the number of categories (Sabour and Dastjerdi, 2013; Szklo and Nieto, 2014; Sabour, 2015; Naderi and Sabour, 2018, 2019). It should be noted that when using a variable with more than 2 categories or an ordinal scale (arranged in 3 or more categories), then a weighted κ will be a good choice. Finally, the third problem is when 2 voters have uneven marginal distributions in their responses (Sabour and Dastjerdi, 2013; Szklo and Nieto, 2014; Sabour, 2015; Naderi and Sabour, 2018, 2019). Table 1 shows agreement with different interpretations and conclusions based on κ (0.43 as moderate) and weighted κ values (0.63 as good). In this table, the number of the category (more than 2) and the marginal distribution of the first category (grade 1) are different from the other categories.

Table 1. The κ and Weighted κ Values for Calculating Agreement Between 2 Listeners for More than 2 Categories and Depend on Prevalence.

		Listener 1			
Estimates	Grade	1	2	3	Sum
Listener 2	1	60	20	1	81
	2	2	12	4	18
	3	3	11	11	25
Sum		65	43	16	124
		Estimate			
κ		0.43			
Weighted κ		0.63			

Authors concluded that preclassification of oral–nasal balance disorders based on nasalance scores may help listeners achieve better diagnostic accuracy and higher agreement. In this letter, we discussed important limitations of applying Cohen κ coefficient to assess reliability (Sabour and Dastjerdi,

[1] Clinical Research Development Centre, Taleghani and Imam Ali Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran
[2] Department of Clinical Epidemiology, School of Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3] Safety Promotions and Injury Prevention Research Centre, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Corresponding Author:
Siamak Sabour, Department of Clinical Epidemiology, School of Health and Safety, Shahid Beheshti University of Medical Sciences, Chamran Highway, Velenjak, Daneshjoo Blvd, Tehran, Iran.
Email: s.sabour@sbmu.ac.ir

2013; Szklo and Nieto, 2014; Sabour, 2015; Naderi and Sabour, 2018, 2019). Any conclusion in reliability analyses needs to be supported by the methodological and statistical issues mentioned above. Otherwise, misinterpretation cannot be avoided.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Siamak Sabour, MD, PhD https://orcid.org/0000-0002-1928-992X

References

de Boer G, Marino VCC, Dutka JCR, Pegoraro-Krook MI, Bressmann T. Nasalance-based preclassification of oral-nasal balance disorders results in higher agreement of expert listeners' auditory-perceptual assessments: results of a retrospective listening study. *Cleft Palate Craniofac J*. 2020;57(3):448-457. doi:10.1177/1055665619873506

Naderi M, Sabour S. Reproducibility of diagnostic criteria associated with atypical breast cytology: a methodological issue. *Cytopathology*. 2018; 29(2):396. doi:10.1111/cyt.12560

Naderi M, Sabour S. Reproducibility of the Bethesda system for reporting thyroid cytopathology: a methodological issue. *J Cytol*. 2019;36(3):185-186. doi:10.4103/JOC.JOC_44_18

Sabour S. Reliability of the ASA physical status scale in clinical practice: methodological issues. *Br J Anaesth*. 2015;114(1):162-163. doi:10.1093/bja/aeu423

Sabour S, Dastjerdi EV. Reliability of four different computerized cephalometric analysis programs: a methodological error. *Eur J Orthod*. 2013;35(6):848. doi:10.1093/ejo/cjs074

Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*, 3rd ed. Jones and Bartlett Publisher; 2014.