Reliability of Smartphone-based Instant Messaging Application for Diagnosis, Classification, and Decision Making in Pediatric Orthopedic Trauma: Methodological Issue

To the Editors:

We were interested to read an article that recently published by Stahl I and colleagues in the Jun 2019 issue of Pediatr Emerg Care.[1] The purpose of the authors was to assess the reliability of diagnosis, classification, and formulation of treatment plans for various pediatric fractures captured and sent via smartphones using instant messaging applications, in comparison with the traditional method of viewing x-ray images over picture archiving and communication system.[1] Intraobserver agreement was calculated by Cohen kappa (κ) statistics for overall agreement and per fracture site. Based on the authors' results, there is "near perfect agreement" between interpretations of the radiographs on smartphones compared with computer-based picture archiving and communication system, with κ of 0.84, 0.82, and 0.89 for diagnosis, classification, and treatment planning, respectively. Also, we looked at the results per fracture site and found substantial to near perfect agreement.[1]

There are several points to calculate reliability using κ value which should be considered. These points are as follows: First, κ value depends on the prevalence in each category. Table 1 shows that in both (a) and (b) situations, the prevalence of concordant cells is 90% and discordant cells is 10%; however, we get different κ value (0.44 as moderate and 0.80 as very good), respectively. Second, κ value also depends on the number of categories.[2-6] Finally, the third important flaw is when the 2 raters have unequal marginal distributions of their responses. Therefore, the weighted κ would be a good choice to investigate intraobserver reliability.[2-6]

TABLE 1. Limitation of κ to Assess Interpretation Reliability of 2 Surgeons With Different Prevalence in the Two Categories

Surgeon 2		Positive Result	Surgeon 1: Negative Result	Total (%)
Situation (a)	Positive result	85	5	90
	Negative result	5	5	10
κ = 0.44	Total	90	10	100
Situation (b)	Positive result	45	5	50
	Negative result	5	45	50
κ = 0.80	Total	50	50	100

Authors concluded that smartphone-based instant messaging applications are reliable for evaluation of a wide range of pediatric limb fractures. This method of obtaining an expert opinion from the off-site specialist is immediately accessible and inexpensive, making smartphones a powerful tool for doctors in the emergency department, primary care clinics, or remote medical centers, enabling timely and appropriate treatment for the injured child. Their conclusion should be supported by the abovementioned methodological issues. Otherwise, misleading message cannot be avoided.

In this letter, we discussed important limitations of applying κ coefficient to assess reliability.[2-6] Any conclusion in reliability analyses needs to be supported by the methodological and statistical issues mentioned above. Otherwise, misinterpretation cannot be avoided.

Mehdi Naderi, MSc
Clinical Research Development Centre
Taleghani and Imam Ali Hospital
Kermanshah University of Medical Sciences
Kermanshah, I.R. Iran

Siamak Sabour, MD, PhD
Department of Clinical Epidemiology
School of Health and Safety
Shahid Beheshti University
of Medical Sciences
Tehran, I.R. Iran
Safety Promotions and Injury
Prevention Research Centre

Shahid Beheshti University of
Medical Sciences
Tehran, I.R. Iran
s.sabour@sbmu.ac.ir

Disclosure: The authors declare no conflict of interest.

REFERENCES

1. Stahl I, Katsman A, Zaidman M, et al. Reliability of smartphone-based instant messaging application for diagnosis, classification, and decision-making in pediatric orthopedic trauma. *Pediatr Emerg Care*. 2019;35:403–406.
2. Szklo M, Nieto FJ. *Epidemiology Beyond the Basics*. 3rd ed. Manhattan, New York: Jones and Bartlett Publisher; 2014.
3. Naderi M, Sabour S. Reproducibility of diagnostic criteria associated with atypical breast cytology: a methodological issue. *Cytopathology*. 2018;29:396.
4. Sabour S, Dastjerdi EV. Reliability of four different computerized cephalometric analysis programs: a methodological error. *Eur J Orthod*. 2013;35:848.
5. Sabour S. Reliability of immunocytochemistry and fluorescence in situ hybridization on fine-needle aspiration cytology samples of breast cancers: methodological issues. *Diagn Cytopathol*. 2016;44:1128–1129.
6. Naderi M, Sabour S. Reproducibility of malignant pleural mesothelioma histopathologic subtyping: methodologic issue. *Arch Pathol Lab Med*. 2018;142:1311.