

LETTER TO THE EDITOR



Response from the authors

Dear Editor

Naderi et al. have shown interest in the methodological approach of our paper entitled "Low reproducibility between oral radiologists and general dentists with regards to radiographic diagnosis of caries". They claim that *"there is no value of kappa that can be regarded universally as indication good agreement"*. Our response is that Cohen's Kappa originally is used for measuring level of agreement between two or more observers when there are at least three categories involved [1]. In this study, we have three groups of observers: radiologist 1, radiologist 2 and 42 general dentists clustered in one group as one observer. Observers have assessed tooth surfaces with respect to three variables. With such background Cohen's kappa statistic is the most appropriate for finding level of agreement between our observers.

Cohen's kappa value ranges from -1 to $+1$ interpreted and represented as different level of agreements universally accepted and used world widely [2].

In addition, our paper contains a statistic as an index of inter-rater agreement among a set of raters in Appendix 1. This procedure is based on a statistic not affected by paradoxes of Kappa [3,4].

Secondly, Table 1 presented by Naderi et al. is not clear to us with respect to prevalence in each category and we do not understand where these numbers originate from. Cohen's kappa values presented in our paper in Table 2 are calculated based on the number of the intact or decayed surfaces of the teeth over which oral radiologists and general dentists had either agreed or disagreed.

In conclusion, we agree with Naderi et al. that Cohen's kappa should be interpreted with care. However, our opinion

is that kappa, despite its' weaknesses, is still useful measure for description of interrater agreement.

References

- [1] Armitage P, Berry G, Mathews JNS. Statistical methods in medical research Blackwell, Science Ltd.; 2008.
- [2] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. (Zagreb). 2012;22:276282.
- [3] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76:378382.
- [4] Falotico R, Quatto P. On avoiding paradoxes in assessing inter-rater agreement. Ital J Appl Stat. 2010;22:151–160.

Elmira Pakbaznejad Esmaeili
Department of Oral and Maxillofacial Diseases, Clinicum,
University of Helsinki, Helsinki, Finland
 elmira.pakbaznejad@helsinki.fi

Tuomas Pakkala
Department of Social Services and Health Care, City of Helsinki,
Helsinki, Finland

Jari Haukka
Department of Public Health, Clinicum, University of Helsinki,
Helsinki, Finland

Päivi Siukosaari
Department of Oral and Maxillofacial Diseases, Clinicum,
University of Helsinki, Helsinki, Finland
Helsinki University Hospital, Helsinki, Finland

Received 22 April 2018; accepted 6 June 2018

© 2018 Acta Odontologica Scandinavica Society