LETTER TO THE EDITOR

Shoulder & Elbow

Shoulder & Elbow
2018, Vol. 10(4) 308

DOI: 10.1177/1758573218791813
journals.sagepub.com/home/sel
SAGE

The kappa paradox

Rens Bexkens, Femke MAP Claessen, Izaak F Kodde, Luke S Oh, Denise Eygendaal and Michel PJ van den Bekerom

Dear Editor,

With interest we have read the Letter to the Editor in response to our article. We agree with the authors of the letter that using Cohen's kappa statistic to assess observer agreement of a qualitative variable has its limitations. However, the advantage of the kappa coefficient is its correction for the amount of agreement that can be expected to occur by chance alone.[1–3] This feature of the kappa statistic has made it one of the most commonly used measures in agreement studies.[4]

Interestingly, a study may report a high absolute percentage of observer agreement (i.e. percentage of observers that agree on the matter, which is independent of the answer as long as they agree) and at the same time report a low kappa value, which is counter-intuitive. The reason for this statistical phenomenon, which is called the first kappa paradox, is the effect that prevalence of the subject under study in a data set has on marginal values.[2,3,5,6] Because of this feature, an imbalance in case distribution will render lower kappa values. This paradox is not a limitation, rather a logical consequence of its purpose; to correctly interpret agreement adjusted for agreement by chance alone.[5,6]

We agree with the authors that one should critically review the study design when interpreting the results of interobserver studies. More specifically, one should look at the case distribution in case of low kappa values. With regard to our article, kappa values may have been higher when using more cases that demonstrated obvious abnormalities (i.e. different case distribution), as mentioned in the discussion of the paper. However, not many patients have radiographic abnormalities after radial head arthroplasty and therefore this study, and thus the kappa value, more closely resembles reality.

In short, we agree that it is important not to neglect the kappa paradox but, taking the absolute number of radiographic abnormalities in daily practice into account, we stand by our original conclusion that one should be cautious when interpreting radiographs after radial head arthroplasty.

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

Rens Bexkens http://orcid.org/0000-0001-6993-0941

References

1. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
2. Doornberg J, Lindenhovius A, Kloen P, et al. Two and three-dimensional computed tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. *J Bone Joint Surg* 2006; 88-A: 1795–1801.
3. Lindenhovius A, Karanicolas PJ, Bhandari M, et al. Radiographic arthrosis after elbow trauma: interobserver reliability. *J Hand Surg* 2012; 37: 755–759.
4. Claessen FM, van den Ende KI, Doornberg JN, et al. Osteochondritis dissecans of the humeral capitellum: reliability of four classification systems using radiographs and computed tomography. *J Shoulder Elbow Surg* 2015; 24: 1613–1618.
5. Cicchetti DV and Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551–558.
6. Feinstein AR and Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543–549.

Amphia Hospital, Breda, Noord-Brabant, Netherlands

Corresponding author:
Rens Bexkens, AMC, Meibergdreef 9, Noord-Holland, 1100 DD Amsterdam, Netherlands.
Email: rensbexkens@gmail.com