

Letters From Our Readers

To: Editor, The Angle Orthodontist

Re: Response to: Reliability of subjective, linear, ratio, and area cephalometric measurements in assessing adenoid hypertrophy among different age groups. By Marcelo Quiroga Souki, Bernardo Quiroga Souki, Letícia Paiva Franco, Helena Maria Gonçalves Becker and Eustáquio Afonso Araújo. *ANGLE ORTHOD.* 2012 E-Pub, doi: <http://dx.doi.org/10.2319/010612-13.1>

We would like to thank Dr. Sabour for the interest in our research work. The concerns by Dr. Sabour probably derive from his clinical epidemiology expertise on the subject, and all raised points were considered by the authors. Our research group is committed to follow the methodological and ethical principles of evidence-based dentistry and work hard to improve the scientific base of our specialty, which is an ongoing challenge. Therefore, all suggestions that may contribute to upgrade our research are sincerely welcome.

The paragraph posted below, extracted from the article abstract written by Gjørup,¹ reflects our thoughts on the questions raised by Dr. Sabour.

"The reliability of a diagnostic test depends on the accuracy and reproducibility of the test results. The accuracy is defined by comparing the test results with a final true diagnosis. The predictive values are here the most important clinical measures. Since it may be impossible to establish a final true diagnosis the reliability must in some cases be measured by a determination of reproducibility. The reproducibility is measured by comparing results of repeated examinations of the same patient. The reproducibility is measured by the use of the kappa coefficient which adjusts the observed agreement for expected chance agreement. A study of reliability of a diagnostic test should fulfill the same methodological requirements as other clinical studies. Both the predictive values and the kappa coefficient are supposed to depend on the prevalence and this should be noticed when results of different studies are compared. Reliability of diagnostic tests is often poor and scientific development of how to improve clinicians' diagnostic practice is much needed".

The measurement of reproducibility is relatively straightforward. The average correlation among two items can be used to obtain an accurate estimate of reproducibility.² We opted to measure the association between measurements and to test for its statistical dependence using the Kendall coefficient and for assessing the reliability of agreement it was employed the weighted kappa score. **It is important to emphasize that such methods were used to study reproducibility, not validity**, as mentioned by Dr. Sabour. It is not the issue here to discuss if weighted Kappa is a recommended indices of diagnostic accuracy for evidence-based practice.³ We recognized that improvements in the investigations methods may give us alternative formulas to test reproducibility and they are very much needed.

In relation to the question why we did not used the likelihood ratio positive/negative and odds ratio, which we agree "are among the best tests to evaluate the validity (accuracy) of a single test compared to a gold standard", we chose not to present them since previous investigations suggested that clinicians rarely make these calculations in practice⁴ and when they do, they often make errors.⁵ However, with the published data, those interested on such ratios can calculate them.

Due to the necessity of being concise, because space limitation, other interesting informations could not be explored, such the great one suggested by Dr. Sabour (to study the ROC curve to assess diagnostic validity of combining linear and ratio LCR measurements), but our group intend to work on such data and present them in a future publication.

Finally, the concern expressed in the commentary by Dr. Sabour "It is crucial to know that statistics cannot provide a simple substitute for clinical judgment" is totally in agreement with our thoughts. In the discussion section it is clear our position on this topic when "we suggest that such numbers should be used only as a reference guide", and that if the "LCR image does not show that the size of the adenoid tissue is increased, but the patient has other clinical features of mouth breathing, referral for a thorough ENT assessment should be recommended". The reader is left to evaluate our article on the basis of its own merits and limitations.

Marcelo Q. Souki, souki@uol.com.br

Bernardo Q. Souki
Letícia P. Franco

Helena M.G. Becker
Eustáquio A. Araújo

REFERENCES

1. Gjørup T. Reliability of diagnostic tests. *Acta Obstet Gynecol Scand Suppl.* 1997;166:9–14.
2. Nunnally JC. *Psychometric Theory*. New York, NY: McGraw-Hill, 1978.
3. Gilchrist JM. Weighted 2×2 kappa coefficients: recommended indices of diagnostic accuracy for evidence-based practice. *J Clin Epidemiol.* 2009;62:1045–1053.
4. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy". *Am J Med.* 1998;104:374–380.
5. Steurer J, et al. *BMJ.* 2002;324:824–826.