

Letters From Our Readers

To: Editor, The Angle Orthodontist

Re: Reliability of subjective, linear, ratio, and area cephalometric measurements in assessing adenoid hypertrophy among different age groups. By Marcelo Quiroga Souki, Bernardo Quiroga Souki, Leticia 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>

Reliability and validity assessment in clinical researches; common mistakes. I was interested to read the paper by Souki MQ and colleagues published in the *Angle Orthod.* The authors assessed the validity of four different types of lateral cephalometric radiograph (LCR) measurements as a diagnostic test of adenoid hypertrophy in different age groups of mouth-breathing children. They reported that Kendall correlation coefficients for agreement between tests were ≥ 0.67 and kappa scores were substantial (≥ 0.64).¹ Such a correlation has nothing to do with validity analysis and actually is one of the common mistakes.

Reliability and validity are two completely different methodological issues in researches. They report the sensitivity of LCR varied from 71% (ratio) to 84% (linear). The specificity varied from 83% (linear) to 97% (ratio). The positive predictive value (PPV) varied from 88% (linear) to 97% (ratio). The negative predictive value (NPV) varied from 70% (ratio) to 78% (linear).¹ Why did the authors not used likelihood ratio positive (true positive/false negative) and likelihood ratio negative (false positive/ true negative) as well as odds ratio (true results/false results – preferably more than 50) which are among the best tests to evaluate the **validity (accuracy)** of a **single test** compared to a gold standard.^{2–5}

As the authors point out in their conclusion, the combination of linear and ratio LCR measurements is a reliable screening tool to determine the need for an ear, nose, and throat evaluation. Reliability (repeatability, inter-changeability or reproducibility) is being

assessed by different statistical tests.⁵ Briefly, for quantitative variable Intra Class Correlation Coefficient ICC and for qualitative variables weighted kappa should be used with caution because kappa has its own limitation too. Regarding reliability or agreement, it is good to know that for computing kappa value, just concordant cells are being considered, whereas discordant cells should also be taking into account in order to reach a correct estimation of agreement (Weighted kappa).^{2–4} It is crucial to know that statistics cannot provide a simple substitute for clinical judgment.^{2–5} As a rule of thumb in clinical researches, clinical importance should be considered a priority instead of statistically significant. The *P* value can easily be changed from significant to non significant due to small sample size, the amount of mean difference and more important factor which is standard deviation of the variable in the study population.^{2–5}

Finally why did the authors not used ROC curve to assess diagnostic accuracy (validity) of their model (combination of linear and ratio LCR measurements)?

Siamak Sabour, MD, PhD

Clinical Epidemiologist

Shahid Beheshti University of Medical Sciences

Tehran, Iran

s.sabour@sbmu.ac.ir

REFERENCES

1. Souki MQ, Souki BQ, Franco LP, Becker HM, Araújo EA. Reliability of subjective, linear, ratio, and area cephalometric measurements in assessing adenoid hypertrophy among different age groups. *Angle Orthod.* 2012 E-Pub.
2. *Epidemiology, biostatistics and preventive medicine*, Jeckel, 1st edition, 2008.
3. *Modern Epidemiology*, K. Rothman, 3rd edition, 2010.
4. *Epidemiology beyond the basics*, Moyses Szklo and F. Javier Nieto, 2nd edition, 2007.
5. Bland JM, Altman DG. Agreed statistics: measurement method comparison. *Anesthesiology.* 2012 Jan; 1165: 182–185.