

# Reliability of shade selection using an intraoral spectrophotometer: common mistakes in reliability analysis

S. Sabour · M. Moezizadeh · E. Vahid Dastjerdi

Received: 30 May 2012 / Accepted: 24 January 2013 / Published online: 13 February 2013  
© Springer-Verlag Berlin Heidelberg 2013

We were interested to read the paper by Witkowski S and colleagues published in the June 2012 issue of *Clinical Oral Investigations*. The authors' aim was to evaluate the accuracy and reproducibility of human tooth shade selection using a digital spectrophotometer. They report the mean colour difference from the mean metric for measurement precision (reliability)! Also, least square test was used to assess inter- and intra-observer reliability [1]. Why did the authors not use well-known statistical tests for reliability analysis such as intra-class correlation coefficient for quantitative variables, weighted kappa for qualitative ones [2–5] or other methods such as coefficient of variance for repeatability [2–5], considering that the mean of the variables and using paired *t* test, least square test or correlation coefficient (*r*) are among the common mistakes in reliability analysis [2–5]? Regarding reliability or agreement, it is good to know that not only avoiding common mistakes in reliability analysis but also taking into account clinical importance instead of statistical significance is a crucial issue in clinical reliability researches [2–5].

The authors computed the confidence interval (CI) value 5.23 (4.66–5.86) to represent the accuracy (validity) of the measurements! As the authors point out in their conclusion, the accuracy and reproducibility of dental shade selection using the tested spectrophotometer with respect to examiner and illumination conditions reflected the reliability of this device [1]. A narrow CI has nothing to do with accuracy. We

have seven well-known statistical tests for validity (accuracy) analysis, as follows: sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio positive (LR+), likelihood ratio negative (LR–) and finally odds ratio [2–5]. Briefly, accuracy and precision (validity and reliability) are two completely different methodological issues in clinical researches, having their own statistical tests, and should not be confused with each other; otherwise, misdiagnosis, mistreatment and mismanagement of the patients will be the result of such researches.

**Conflicts of interest** None

## References

1. Witkowski S, Yajima ND, Wolkewitz M, Strub JR (2012) Reliability of shade selection using an intraoral spectrophotometer. *Clin Oral Investig* 16(3):945–949
2. Jekel JF, Katz DL, Elmore JG, Wild D (2007) *Epidemiology, biostatistics and preventive medicine*, 3rd edn. Elsevier, Philadelphia
3. Rothman KJ, Greenland S, Lash TL (2018) *Modern epidemiology*, 3rd edn. Lippincott Williams and Wilkins, Philadelphia
4. Szklo M, Javier Nieto F (2007) *Epidemiology beyond the basics*, 2nd edn. Jones and Bartlett, Sudbury
5. Bland JM, Altman DG (2012) Agreed statistics: measurement method comparison. *Anesthesiology* 116(1):182–185

S. Sabour (✉)

Department of Clinical Epidemiology, Faculty of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran  
e-mail: s.sabour@sbmu.ac.ir

M. Moezizadeh · E. V. Dastjerdi  
Faculty of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran