Journal of Oral Research

LETTER TO THE EDITOR

Comment on: "Transcultural adaptation and reliability of the Spanish version of a questionnaire of oral hygiene advice given by dentists in Chile"

DOI: *10.17126/joralres.2015.032*

In the last issue of Journal of Oral Research have been published the entitled paper: "Transcultural adaptation and reliability of the Spanish version of a questionnaire of oral hygiene advice given by dentists in Chile"[1] which leads to this editorial letter.

One of the most trending and challenging topics in research is the measurement process and its employed tools. Scales are developed and continuously assessed to account the minimum measurement error [and even measurement biases] thus ensuring adequate the results' external validity.

For scales, the following standards in selection, construction and application should be considered: 1. Substantial changes in the structure; application format, language or content leads to scale re-validation due to the new conditions. If not, arguments should be presented justifying the decision for no re-validation. 2.In the case of translated scales, its reliability and validity should be assessed in the new population to be applied. 3.When comparability of scales in two different languages is intended, researchers must perform statistical tests to assess the degree of reproducibility between them[2]. Having in mind the aforementioned paper, item number 2 is addressed; nevertheless potential pitfalls for result generalization were detected.

Reliability [often called precision] is the extent in which a measurement is error free. In conditions where the measurements are repeated, results should be similar. This concept is related with the instrument stability [regardless of the assessor] and the time in which is applied to the population[3]. As the term implies, a reliable instrument is one that performs in consistent, predictable ways; the score produced by the instrument that should not change unless there has been an actual change in the variable the instrument is measuring and, thus, that any observer change in scores can be attributed to actual change in that variable[4].

Validity [often called accuracy] is the degree in which an instrument [i.e. scale or index] measures what its intended to assess; this is a latent variable (3). The adequacy of a scale as a measure of a specific variable is an issue of validity. Whereas reliability concerns how much a variable influences a set of items, validity concerns whether the variable is underlying cause of item covariation. There are essentially three types of validity that correspond to these operations: content, criterion-related and construct[4].

Statistical approaches for scales psychometric properties [reliability and validity] are widespread discussed in statistical journals and books. There is consensus on what tests should be performed and its presentation outline for scientific papers in the context of public health [dental] research. Having in mind the aim of the authors, in the following lines, I'll describe the employed statistical methods and the suggested approaches for this.

Instrument-related reliability was assessed with the Cronbach's alpha [α] coefficient accepting a cut-off point of 0.70 so it's assumed that the dependent variable [global score] was approached as a numerical outcome. Despite its the most widely used statistical test in internal consistency assessment, due to its several limitations McDonald's Omega is preferred instead of the Cronbach's α[5]. Although, accepted values should be between 0.70–0.90, confidence intervals must be reported as well.

For time-related reliability, test-retest reliability [TRT-R], which is the most accurate method, was reported. Nevertheless, they don't report any specific statistical approach. Several statistical procedures can be performed to assess it. In addition, authors intend to report limits of agreement for TRT-R, however results are not depicted in the paper. So authors could statistically address this

ISSN Online 0719-2479 - ©2014 - Official publication of the Facultad de Odontología, Universidad de Concepción - **www.joralres.com**

matter by computing the Lin's Concordance Correlation Coefficient, the most recommended statistical approach since includes both precision and accuracy in its formula, rather than traditional methods as Intraclass Correlation Coefficient, or Analysis of Variances. Also, the statistics summary for the Lin's CCC can be accompanied with the Bland & Altman's plots which are a graphical depict of the limits of agreement[2].

In the case of validity two types of it were evaluated: construct and content. In that paper, and considering the underlying statistical theory, transcultural validity should be understood as a previous step for psychometric properties assessment rather than a type of validity[2]. Face validity [the paper approach] and content validity are sometimes confused because both may concern the extent to which item content appears relevant to the construct of interest. An important difference, however, is that content validity is defined in terms of specific procedures, and those procedures are generally more structured and rigorous than informal assessments of face validity[4].

For the content validity, authors may conduct factor analysis [exploratory factor analysis–EFA, in the first step], which reflects the underlying factors [latent variables] that are being measured[1]. Also it allows to compute items variability and in the most of the cases, this is considered cumulative evidence to redefine the factor structure of any scale. Construct validity, the extent to which a measure behaves the way that the construct it purports to measure should behave with regard to established measurements of other constructs[3]. The recommended statistical methods to approach construct validity includes the multitrait-multimethod matrix[1,3] and approaches like confirmatory factor analysis [CFA] a member of the structural equation modeling techniques[6].

Authors also report Cohen's κ and W-Cohen's $\kappa$to examine test-retest reliability. Cohen's Kappa should be used to examine inter-examiner agreement rather than TRT-R. Inter-examiner agreement should be assessed using Lin's CCC since the measurement [global scale score] was approached as a numerical variable and taking in to account its capabilities[2].

Regarding the sample size, for validation studies sample size can be assessed with the scale number of items [5-10 subjects/item][2,6]. So, with 22 items, an approximately sample size should be between 110 and 220. Although required sample sizes are low for the initial steps [translation, face validity and pilot testing], adequate number of participants are important in this studies. Finally, methodological issues as sample size requirements and proper statistical methods should be used to achieve accurate and generalizable results leading to scientific evidence that could effectively support the governments' public health decisions.

MIGUEL SIMANCAS MSc
Clinical Epidemiology. Applied Statistics
Student. Professor. Associate Researcher.
Public Health Research Group. Faculty of Dentistry.
Research Department. University of Cartagena.

REFERENCES.

1. Sánchez R, Echeverry J. Validating scales used for measuring factors in medicine. Rev Salud Pública. 2004: 6; 302–18.
2. Aravena P, Moraga J, Cartes-Velásquez R, Manterola C. Validez y confiabilidad en investigación Odontológica. Int J Odontostomat. 2014: 8 (1); 69-75.
3. DeVellis RF. Scale development. Theory and applications. 3rd Ed. Thousand Oaks, CA: SAGE Publications; 2012.
4. Williams B, Onsman A, Brown T. Exploratory factor analysis: A five-step guide for novices. Australasian J Paramed. 2010; 8 (3):1-13.
5. Harrington D. Confirmatory Factor Analysis. Pocket Guides to Social Work Research Methods. 1st Ed. Oxford University Press. New York, NY: Oxford University Press, Inc; 2009.