doi:10.5937/sejodr3-1274
Short Communication
SEJODR

StaTips Part II: Assessment of the repeatability of measurements for continuous data

Perinetti, Giuseppe *

** Department of Medical, Surgical and Health Sciences, University of Trieste, Italy*

ABSTRACT

Primary research requests the recording of data through measurements, such as linear and angular parameters in cephalometric, strength of adhesion, skeletal maturation stages and so on. In all of these cases, an analysis of the method error, i.e. repeatability of the measurements, is a fundamental part of the study. In this paper, the case of the continuous data set will be taken into consideration.

Perinetti G. StaTips part II. South Eur J Orthod Dentofac Res. 2016;3(2):33-34.

FRAMING THE PROBLEM

Primary research requests the recording of data through measurements, such as linear and angular parameters in cephalometric, strength of adhesion, skeletal maturation stages and so on. In all of these cases, an analysis of the method error, i.e. repeatability of the measurements, is a fundamental part of the study. Indeed, any recording is subjected to some degree of error that has to be quantified in order to ensure that results are reliable. Repeatability analysis does not only include the case of replicate recordings (by the same operator through the same parameter), but also the case of agreement between raters (same parameter) or different parameters (same operator). Moreover, it follows proper statistical procedures according to the type of the data (continuous, ordinal and nominal), as for the tests used to evaluate differences among the groups. [1] In this paper, the case of the continuous data sets will be taken into consideration, while the corresponding procedures for ordinal/nominal parameters will be included in the forthcoming StaTips.

SYSTEMATIC AND RANDOM ERROR

When dealing with repeatability of the measurements, an important distinction has to be made between the components of the total method error, which are the systematic and random error. Systematic error is also referred as 'bias', and it is a reproducible inaccuracy leading to a measured value that is consistently greater or lower than the true value. Random error is a non-reproducible inaccuracy leading to a measured value that may be greater or lower than the true value. Therefore, while the effects from systematic error may be predictable, those from the random error are not. A proper repeatability analysis has to evaluate both the systematic and random errors.

Corresponding Author:
Giuseppe Perinetti
Struttura Complessa di Clinica Odontoiatrica e Stomatologica, Ospedale Maggiore,
Piazza Ospitale 1, 34129 Trieste, Italy
e-mail: G.Perinetti@fmc.units.it

EVALUATION OF THE SYSTEMATIC ERROR

Systematic error is tested by evaluating the significance of the difference between (or among) data sets. Therefore, for continuous data sets with the required assumption, the paired t-test may be used. [1] Alternatively, in the case when data sets fail to show required assumptions for the use of parametrical tests, non-parametrical tests, i.e. Wilcoxon test, may be used instead. For further details, see StaTips Part I. [1] A common mistake seen in several published studies resides in the use of the t-test alone to show repeatability of the measurements. The t-test should be used only to exclude the existence of bias, not to assess the repeatability of the measurement, being it unable to quantify the random error.

EVALUATION OF THE RANDOM ERROR

Several methods are available to evaluate the random error including the Dahlberg's formula, [2] the method of moments variance estimator (MME), [3] and the Bland and Altman plot. [4] A substantial difference between these approaches is that the Dahlberg's formula and the MME can quantify the absolute random error, while the Bland and Altman plot is a graphical representation of the spreading between replicate measurements, which is of easy interpretation but does not quantify the absolute random error.

Calculation of the Dahlberg's formula and the MME is of easy execution with details reported in Table. An Excel datasheet can be easily used to carry out the desired outcomes. A comparative analysis [3] between the Dahlberg's formula and the MME, examining the effects of different sample sizes and different levels of systematic error between replicated measurements on the accuracy of estimates of random error, showed that: 1) irrespective of the procedure, a sample of at least 25 cases should be used for replication of the measurements; 2) in case of smaller sample sizes, confidence intervals should be reported along the mean estimates of errors; and 3) the error estimation obtained by the MME is less subjected to systematic error as compared to the Dahlberg's formula. Therefore, the MME is preferable over the Dahlberg's formula unless (a rare situation) existence of systematic error between replicate measurements can be excluded. For calculation of confidence intervals for the MME (in the case of fewer than 25 replications), refer to a previous report. [3]

The Bland and Altman plot is a graphical method where the differences between pairs of measurements (or between two parameters) are plotted against the corresponding averages of the two measurements (Figure 1). The Authors recommend plotting the difference (M1-M2) between the first (M1) and second (M2) measurements against the average (M1+M2)/2 of the results obtained from the two recordings. In the ideal case, there should not be any correlation between the difference and the average, irrespective of the existence of bias (systematic error). As an alternative, the difference as a % of the sum (M1-M1)/(M1+M2) or the ratio (M1/M2) may be used instead of the difference.

Greater the repeatability/agreement, more the values will be clustered around the mean of the differences (referred to as 'bias'). When values are within 1.96 standard deviations of the bias (referred to as 'limits of agreement'), the recording is considered repeatable (or agreement between methods is acceptable). Assuming that differences are normally distributed, the limits of agreement will correspond to the 95% prediction interval.

Finally, when dealing with two different parameters, i.e. when testing the reliability of a new method against another already known, the differences can be plotted against the latter method provided if can be considered as a "gold standard." [5] Of note, the Bland and Altman plot remains a visual analysis and whether consider the replicate measurements repeatable (or in agreement) or not depends on the level of precision that is needed in that particular circumstance. Authors have to carefully evaluate case-by-case, and eventually report also a quantification of the random error, preferably through the MME.

FIGURE LEGEND

Figure 1. An example of the Bland and Altman plot for repeatable and not repeatable measurements.

1a. a case of repeatable measurements where all the differences between replicate recordings are within the limits of agreements;

1b. a case of not repeatable measurements with four differences between replicate recordings outside the limits of agreement. M1 and M2, represent the replicate measurements; SD, standard deviation of the mean of the difference. Mean of the difference (bias), solid blue line; upper and lower limits of agreement, dotted red lines. In this example, arbitrary units are used.

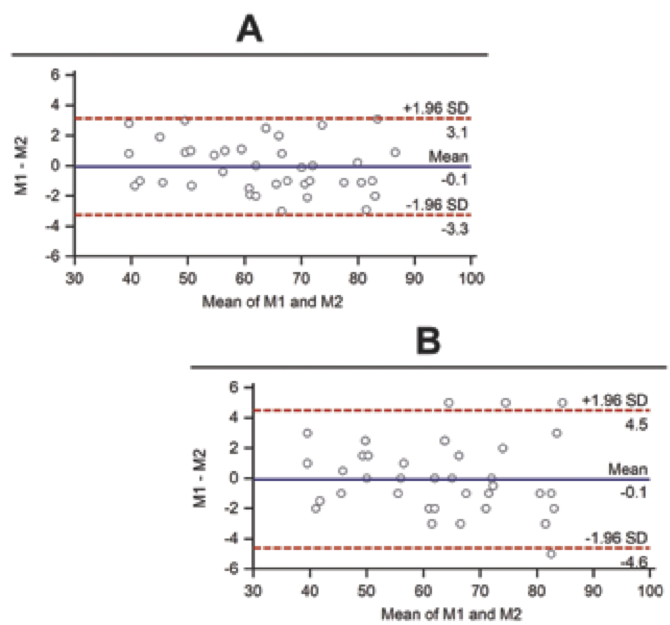

Table 1. A brief comparison of the calculations of the Dahlberg's formula and method of moments variance estimator (MME)

Error estimator procedure	Calculation	Legend
Dahlberg's formula	$S_D = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{2n}}$	• ***d*** is the difference between the pairs of replicate measurements • ***n*** is the number of cases • S_D is the estimate of the random error
Method of moments variance estimator (MME)	$S_M = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{2(n-1)}}$	• ***d*** is the difference between the pairs of replicate measurements • ***n*** is the number of cases • $\bar{d}$ is the difference between the • S_M is the estimate of the random error

REFERENCES

1. Perinetti G. StaTips Part I. Choosing statistical test when dealing with differences. South Eur J Orthod Dentofac Res. 2016;3(1):4-5.
2. Dahlberg G. Statistical methods for medical and biological students. New York: Interscience Publications; 1940.
3. Springate SD. The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. Eur J Orthod. 2012;34(2):158-63.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
5. Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. Stat Med. 2008;27(5):778-80.