Elsevier Editorial System(tm) for American Journal of Orthodontics & Dentofacial Orthopedics
Manuscript Draft

Manuscript Number:

Title: HOW TO REPORT RELIABILITY IN ORTHODONTIC MEASUREMENTS. Part 2

Article Type: Statistics and Research Design

Keywords: statistics; research design

Corresponding Author: Dr. Shin-Jae Lee, DDS, MS, PhD, PhD.

Corresponding Author's Institution: Seoul National University School of Dentistry

First Author: Richard E Donatelli, DMD, MS.

Order of Authors: Richard E Donatelli, DMD, MS.; Shin-Jae Lee, DDS, MS, PhD, PhD.

Title Page

HOW TO REPORT RELIABILITY IN ORTHODONTIC RESEARCH

Part 1

Authors:

Richard E. Donatelli, DMD, MS[1]

Shin-Jae Lee, DDS, MS, PhD, PhD[2]*

[1]Clinical Assistant Professor, Department of Orthodontics, University of Florida College of Dentistry, Gainesville, Florida.

[2]Professor and Chair, Department of Orthodontics and Dental Research Institute, Seoul National University School of Dentistry, Seoul, Korea.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology [2012-0007545].

***Corresponding author:**

Shin-Jae Lee, DDS, MSD, PhD, PhD.

Professor and Chair, Department of Orthodontics and Dental Research Institute

Seoul National University School of Dentistry

101 Daehakro, Jongro-Gu, Seoul 110-749, Korea

Tel, 82 2 2072 3019; Fax, 82 2 2072 3817;

E-mail: nonext@snu.ac.kr

HOW TO REPORT RELIABILITY IN ORTHODONTIC MEASUREMENTS. Part 2

ABSTRACT

Proper statistical analysis is an absolutely essential tool for both clinicians and researchers attempting to implement evidence-based decisions. When analyzing reliability, statistical graphical representation is the best method. Other previously published error studies of 2D measurements, such as cephalometric landmarks, have inappropriately applied one-dimensional approaches, such as linear- or angular measurements. The aim of this present article is to illustrate a graphic presentation method that can be applied to two-dimensional (2D) data sets. We propose that this 2D visualization technique, that can visualize errors in both *x*- and *y* axes simultaneously, should be utilized when reporting the reliability of a 2D data set. Our prediction error analysis of soft tissue changes after orthognathic surgery will be presented as an example. By applying different colors to each ellipse, this method can also identify any between-group differences that may be present.

INTRODUCTION

(*Background*)

Reliability studies are conducted in experimental or survey situations to assess the level of observer variability in the measurement procedures to be used in data acquisition and/or investigation.[1] Several related measurements may be taken on a single individual. For example, measurements of the teeth, dentition, jaw bone, and soft tissue covering are *correlated* (or *clustered*). A problem may arise if clustered observations are erroneously treated and analyzed as separate (*independent* or *non-clustered*) observations.[2-6] Orthodontics, a unique specialty in dentistry, also has unique problems with which conventional methods cannot properly deal. This is because orthodontic treatment involves numerous variables within the dentition, skeletal configuration, and soft tissue responses which are all clustered within an individual. Therefore, to properly consider the correlation among variables, a more sophisticated method than the ordinary analysis and/or interpretation method should be applied.[7-9]

(*Raising an issue*)

The commonly utilized orthodontic cephalometric points, which have *x*- and *y* Cartesian coordinates (axes), are examples of measurements that should not be considered separately but as clustered data sets. Let us suppose that we need to know the identification error of a specific cephalometric landmark, or, by using cephalometric analyses, we are going to measure the accuracy of predicting an orthognathic surgical result. In these cases, we are analyzing deviations between predicted and actual results. An issue may arise since the location of a cephalometric landmark has two measurements, both in *x*- and *y* axes. In other words, in a cephalometric error study, a single variable has two values that are correlated. Even in the polar system, which utilizes angles and distances instead of the *x*- and *y* axes to identify points, there are still two variables to each cephalometric point. Therefore, how can we properly express the reliability measure of a single cephalometric landmark? How we can report a reliability measure of a 2D variable? In our investigation, with the exception of only a couple of articles,[7,10] we could not identify in the orthodontic literature an error report or a reliability measure that simultaneously expressed 2D errors in the 2D plane.

(*Purpose of this manuscript*)

We have previously suggested a simple modification of the Bland-Altman plot for use in both simultaneous intra- and inter-observer reliability situations.[(#Part 1 citation)] However, since the Bland-Altman plot demonstrates the pattern and magnitude of an error in only one-dimension at a time, we therefore need a 2D approach that would be more appropriate in reporting 2D reliability than the conventional aforementioned 1-dimensional approach. By proposing a 2D graphic presentation style for a cephalometric 2D data set that can visualize both *x*- and *y*-axis errors simultaneously, the aim of this present article is to illustrate a graphical presentation method that can analyze the error in 2D data sets.

CRITICISMS OF PREVIOUS REPORTS EVALUATING EITHER ANGULAR OR LINEAR MEASUREMENTS AND THE LIMITATION OF THE BLAND-ALTMAN PLOT FOR 2D CEPHALOMETRIC DATA

Traditionally, in studies with cephalometrics, errors of measurement are considered to arise from both the actual identification of the landmarks and the linear or angular measurements derived from those landmarks. Since cephalometrics has not been an exact science, we, orthodontists, recognize that certain latitude must be granted to a person tracing cephalometric radiographs and that errors are likely to occur.[11] As mentioned, traditionally in cephalometrics, landmark identification errors or prediction errors have been most commonly reported by means of one-dimensional variables, such as linear- and/or angular measurements, i.e., linear measurements between two landmarks *separately* in either horizontal (*x*-axis) or in vertical (*y*-axis)[12-17] or degrees of angles among three points.[16-18]

The problem associated with the reliability of linear or angular measurements is because a linear measurement is formed from two points and an angle is formed from three points. Measuring errors using linear and/or angular measurements cannot pinpoint the culprit of the error from out of the related points. In general, a distance measurement is more precise and/or reliable than an angular measurement. For example, when reporting the reliability of a measured angle, such as ANB (SNA – SNB), the source of error could be partly due to the unreliability of locating any of the three related points (Nasion, Point A, and Point B), and the variation among the location of the points can also be compounded when determining the measured angle.[19] In some situations, the points themselves may not be significantly different, but when added together, produce a very different angle.[20]

With regard to the 1-dimensional graphical visualization of the error report, Bland and Altman developed a simple, intuitive, and easy method to show a reliability measure between two variables.[21] The simple descriptive analysis described by Bland and Altman[21] permits the assessment of the agreement between two imperfect clinical measurements or the repeatability of duplicate observations. As long as there is only one variable, the Bland-Altman plot can be utilized to evaluate both intra-observer and inter-observer reliability. This method has become increasingly popular in orthodontic publications.[(*#Part 1 citation*)] Nonetheless, when a variable has a 2D entity, the Bland-Altman plot cannot visualize both *x*- and *y* axis errors simultaneously in the 2D plane.

RELIABILITY REPORTS FOR 2D DATA IN THE 2D PLANE

In this section, we use a real clinical example for illustrative purpose. The example's data is from a clinical study by Suh *et al* (2012), which suggested the appropriateness of a new method (Method 2) over the conventional least squares method (Method 1) to predict soft tissue change after orthognathic surgery. The subjects consisted of patients (N = 69) who had undergone surgical correction of Class III mandibular prognathism by only mandibular setback surgery.[7] In this example, we selected two simple but important cephalometric landmarks, soft tissue pogonion and soft tissue menton, from among the variables in the original paper. After applying the conventional least squares prediction method to the *test* data set (also called the *validation* data set), the result of the prediction errors (or the systematic error, also called *bias*) did not show a significant difference either in *x*- or in *y*-axis. However, since overestimations and underestimations of the predictions essentially cancel themselves out when mean values are derived, a comparison test using the mean values in a reliability report is not appropriate.[7] A graphical representation is again the best method.

In a 2D situation, a scattergram can be plotted in 2D space. In a scattergram, a negative value indicates that the prediction is more posterior in the *x*-axis or more superior in the *y*-axis as compared to the actual result (**Figure**). If, and only if, errors in both the *x*- and *y*-axes were normally distributed (Gaussian) and the *x*- and *y*-axes did not show linear association between them (no correlation), the plots would form a circle. In most cases, the plots form an ellipse with a certain degree of deformation. After plotting the errors, ellipses can be depicted (**Figure**). The ellipsoid satisfies $(\mathbf{z} - \boldsymbol{\mu})^{T} \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \leq \chi^2(\alpha)_2$, where $\mathbf{z}$ is the two dimensional (*x*- and *y* coordinates) vector for the error; $\boldsymbol{\mu}$ is the mean vector for $\mathbf{z}$; $\boldsymbol{\Sigma}^{-1}$

is the inverse matrix of the covariance matrix; and $\chi^2(\alpha)_2$ is the upper 95th percentile of a chi-square distribution with 2 degrees of freedom.[22]

A scattergram with an ellipse representing plot points is essentially a 2D extension of the Bland-Altman plot. A Bland-Altman plot has two horizontal lines, the lower and upper agreement limits, which indicate which plotted data points are "within the mean difference ± 2 standard deviations (SD)." The range between the lower- and upper limits also indicates the so-called 95% limits of agreement. Likewise, the contour of an ellipse within a scattergram also indicates the 95% confidence boundary. If any points are outside the 95% confidence boundary of ellipse on the graph, they can be called *outliers* just like plot points that are outside the horizontal lines of the Bland-Altman plot graph.

There are several advantages when utilizing a 2D scattergram and a 95% confidence ellipse:

1. It can visualize not only the reliability but also the result of method comparisons. This can be done easily by assigning a different type and/or different color of dots for the second observer or for the repeated measurements of an alternative method being compared. The mode of comparison can be expanded to more than two groups by changing colors and point characters. In the **Figure**, for example, we assigned red circles and blue diamond dots, respectively, for the first and second methods. After comparing the size of the 95% confidence ellipses, Method 1 illustrated significantly larger ellipses than Method 2. Therefore, upon observing the graphs in this example, it is immediately obvious that Method 2 showed a significantly more accurate, and higher predictive performance than Method 1.
2. The form and shape of the ellipse imply the correlation between the *x*- and *y*-axis errors. A perfect circle is indicative of perfect independence and normality between the *x*- and *y*-axis errors of a variable. A more deformed ellipsoid indicates greater correlation between the *x*- and *y*- axis values.
3. The scattergram and 95% confidence ellipse can leave something to interpret. For example, soft tissue pogonion graph (**Figure**, *left*) shows more errors in the vertical than in the horizontal axis. On the other hand, errors in the soft tissue menton graph (**Figure**, *right*) are distributed more horizontally than in the vertical direction. This may partly be related to the definition of the cephalometric landmarks themselves, pogonion and menton. For example, when locating the most anterior point of the chin, pogonion, it is likely easier to determine the position on the *x*-axis coordinate than on the *y*-axis, thus producing the large vertical variation in the distributions for pogonion. By the same token, since menton is the most inferior point of the chin, there is greater horizontal variation in the *x*-axis than in the *y*-axis. Without graphic visualization, the previous interpretation would have been difficult.

In this article, the language **R** program (Vienna, Austria) was used. **R** is a free software for statistical computing.[23] Detailed codes of the example data sets and plots for use with language **R** are available by request to the authors. Recently, there are greater numbers of published studies involving three-dimensional (3D) imaging technology. We envision that an easy 3D statistical graphic presentation for 3D data reliability will also be introduced in the near future.

CONCLUSIONS

We illustrated a graphical presentation method that can be applied to 2D data sets. We proposed a 2D graphical presentation style for a cephalometric 2D data set that can visualize both *x*- and *y*-axis errors simultaneously. The proposed method can also distinguish between-group differences, if any, by applying different colors to each ellipse. Several advantages and the usefulness of applying the 2D scattergram and the 95% confidence ellipse were also discussed. We hope that the proposed method may encourage accurate and meaningful reporting of reliability analyses in orthodontic research.

REFERENCES

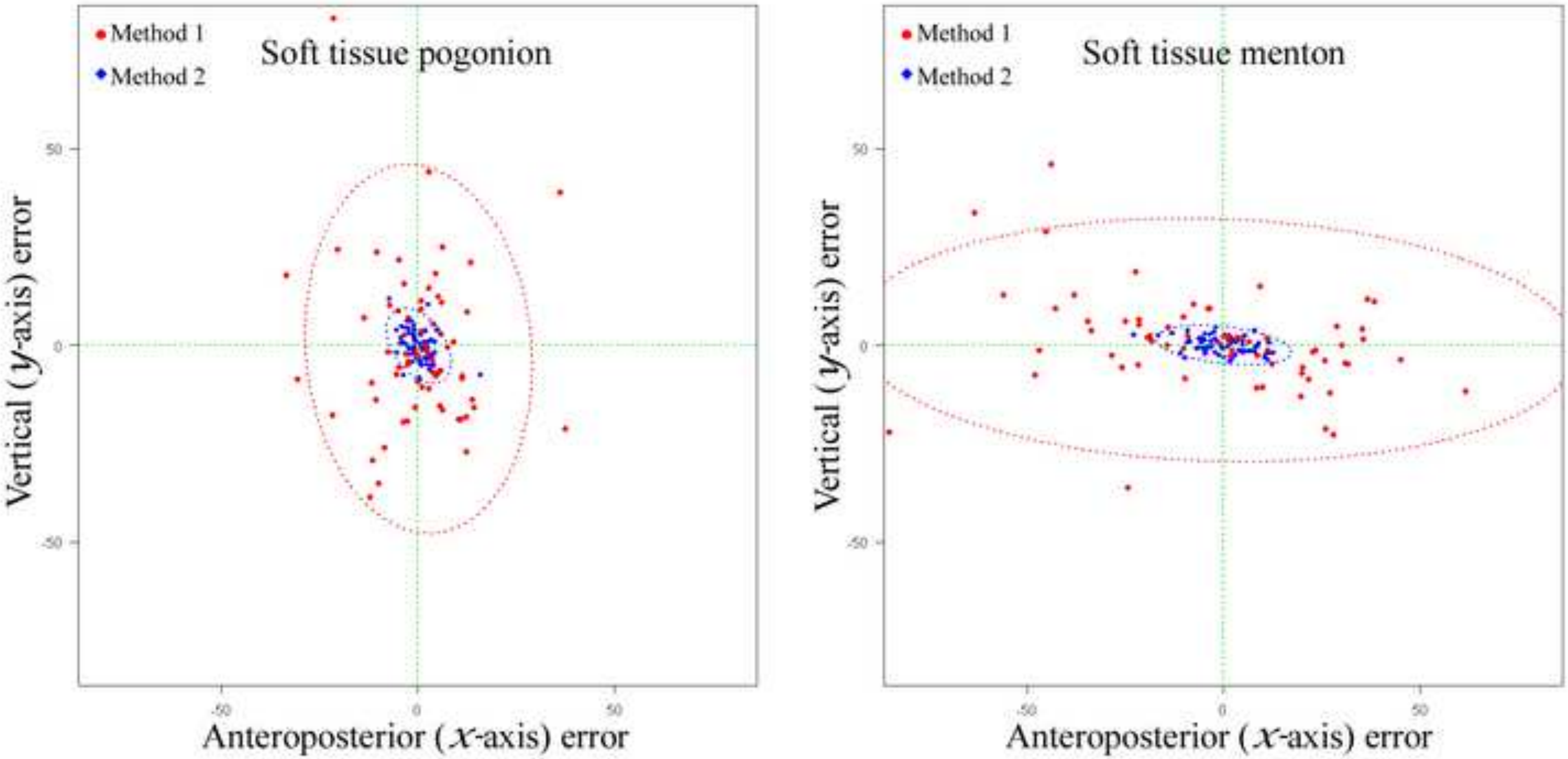
1. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.
2. Pandis N, Walsh T, Polychronopoulou A, Eliades T. Cluster randomized clinical trials in orthodontics: design, analysis and reporting issues. Eur J Orthod 2012.
3. Lee SJ, Lee S, Lim J, Park HJ, Wheeler TT. Method to classify dental arch forms. Am J Orthod Dentofacial Orthop 2011;140:87-96.
4. Lee SJ, Ahn SJ, Lim WH, Lee S, Lim J, Park HJ. Variation of the intermaxillary tooth-size relationship in normal occlusion. Eur J Orthod 2011;33:9-14.
5. Lee SJ, An H, Ahn SJ, Kim YH, Pak S, Lee JW. Early stature prediction method using stature growth parameters. Ann Hum Biol 2008;35:509-517.
6. Lee SJ, Lee S, Lim J, Ahn SJ, Kim TW. Cluster analysis of tooth size in subjects with normal occlusion. Am J Orthod Dentofacial Orthop 2007;132:796-800.
7. Suh HY, Lee SJ, Lee YS, Donatelli RE, Wheeler TT, Kim SH et al. A more accurate method of predicting soft tissue changes after mandibular setback surgery. J Oral Maxillofac Surg 2012;70:e553-562.
8. Li EN, Lim J, Kim K, Lee SJ. Distribution-free tests of mean vectors and covariance matrices for multivariate paired data. Metrika 2012;75:833-854.
9. Lim J, Li EN, Lee SJ. Likelihood ratio tests of correlated multivariate samples. Journal of Multivariate Analysis 2010;101:541-554.

10. Tanikawa C, Yagi M, Takada K. Automated cephalometry: system performance reliability using landmark-dependent criteria. Angle Orthod 2009;79:1037-1046.
11. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: poor reproducibility. Am J Orthod Dentofacial Orthop 2009;136:478 e471-477.
12. McCollum AGH, Gardener GJM, Evans WG, Becker PJ. Soft-tissue changes related tomandibular advancement surgery. Semin Orthod 2009;15:161-171.
13. McCollum AGH, Dancaster JT, Evans WG, Becker PJ. Sagittal soft-tissue changes related to the surgical correction of maxillary-deficient Class III malocclusions. Semin Orthod 2009;15:172-184.
14. Ksiezycki-Ostoya BK, McCollum AGH, Becker PJ. Sagittal soft-tissue changes of the lower lip and chin associated with surgical maxillary impaction and consequent mandibular autorotation. Semin Orthod 2009;15:185-195.
15. Kaipatur NR, Flores-Mir C. Accuracy of computer programs in predicting orthognathic surgery soft tissue response. J Oral Maxillofac Surg 2009;67:751-759.
16. Naoumova J, Soderfeldt B, Lindman R. Soft tissue profile changes after vertical ramus osteotomy. Eur J Orthod 2008;30:359-365.
17. Joss CU, Vassalli IM, Thuer UW. Stability of soft tissue profile after mandibular setback in sagittal split osteotomies: a longitudinal and long-term follow-up study. J Oral Maxillofac Surg 2008;66:1610-1616.
18. Rustemeyer J, Groddeck A, Zwerger S, Bremerich A. The accuracy of two-dimensional planning for routine orthognathic surgery. Br J Oral Maxillofac Surg 2010;48:271-275.
19. van der Linden FP. Sheldon Friel memorial lecture 2007: myths and legends in orthodontics. Eur J Orthod 2008;30:449-468.
20. Eckhardt CE, Cunningham SJ. How predictable is orthognathic surgery? Eur J Orthod 2004;26:303-309.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.
22. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Pearson Prentice Hall; 2007.
23. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.

FIGURE LEGENDS

Figure Scattergrams and 95% confidence ellipses for the error that were obtained from a conventional method 1 (*red*) and an alternative method 2 (*blue*). The plots clearly indicate that the error in both *x*- and *y*-axes was greater in Method 1 than in Method 2. This implies that the application of the method to an individual subject may give rise to errors to this extent with 95% probability. Refer to **Part 2** text for explanation.

Figure
Click here to download high resolution image



ICMJE Statement on Conflict of Interest for each author
Click here to download ICMJE Statement on Conflict of Interest for each author: ICMJE_Form_2_LeeShinJae.pdf