European Journal of Orthodontics 29 (2007) 105–108
doi:10.1093/ejo/cjl065
© The Author 2007. Published by Oxford University Press on behalf of the European Orthodontic Society.
All rights reserved. For permissions, please email: journals.permissions@oxfordjournals.org.

An evaluation of the errors in cephalometric measurements on scanned cephalometric images and conventional tracings

Korkmaz Sayinsu, Fulya Isik, Göksu Trakyali and Tülin Arun
Department of Orthodontics, Faculty of Dentistry, Yeditepe University, Istanbul, Turkey

SUMMARY The aim of this investigation was to compare the classic method of tracing by hand with a computerized method, where the lateral cephalograms were scanned at 300 dpi and digitized onscreen. The inter- and intra-observer errors were investigated for tracing and digitizing errors.

Thirty lateral cephalograms were scanned into digital format at 300 dpi, displayed on a high-resolution monitor, and processed twice by two operators using Dolphin Imaging Software 9.0. The same radiographs were then traced and measured manually by the same two operators. Intraclass correlation coefficients (ICC) were used for detecting intra- and inter-rater agreement for each cephalometric variable.

The results showed that each operator was consistent in the repeated measurements; all ICC were greater than or equal to 0.90 and none of the 95 per cent confidence limits on these ICC had a lower boundary less than 0.84. Inter-rater agreement also showed correlation coefficients greater than 0.75. The angles, maxillary height, maxillary depth, *y*-axis, FMA, and nasolabial, and the distance N perpendicular point A had a wider reliability interval and lower correlation than the other parameters tested. The findings demonstrated that the use of computer software for cephalometric analysis carried out on scanned images does not increase the measurement error when compared with hand tracing.

Introduction

In contemporary orthodontics, lateral cephalograms are used for the assessment of treatment changes induced by the appliances used. Therefore, it is important to keep the method error to a minimum in order to see the valid small changes achieved by treatment (Kamoen *et al.*, 2001). Hence, errors arising from the acquisition of radiographs, tracing, landmark identification, and measurements have been investigated in an effort to minimize related errors (Baumrind and Frantz, 1971a, b; Gravely and Benzies, 1974; Cohen, 1984; Houston *et al.*, 1986; Battagel, 1993; Chen *et al.*, 2000, 2004; Turner and Weerakone, 2001).

Tracing on paper using hand instruments is reported to compare favourably with the results of digitized radiographs and the findings of studies using manual methods could be considered perfectly valid (Richardson, 1981; Sandler, 1988). Manual tracing was found to yield more reproducible results especially for the points articulare and gonion which are constructed on a tracing, but only estimated using the digitizer (Sandler, 1988). Other points were easier to visualize and locate when the outline of the structure could be traced first, such as the apex of the maxillary incisor root (Houston, 1983).

On the other hand, hand measurements are time-consuming and there is a risk of misreading the measuring instruments and registering data to the computer (Sandler, 1988). If digitization is carried out, then the angles and distances are automatically calculated which can eliminate the errors in drawing lines between landmarks and in measurements with a protractor. Moreover, the digital image can be manipulated to process the image and alter its visual appearance which can facilitate landmark identification (Jackson *et al.*, 1985).

Although radiographic film is quite stable and can retain its information for many years, due to its physical nature, it is not always a dependable archive medium (Geelen *et al.*, 1998). Film deterioration has been a major source of information loss in craniofacial biology (Melsen and Baumrind, 1995); therefore, digital archiving of lateral cephalograms is a valuable method for orthodontic clinics.

A previous study revealed that computer-aided cephalometric analysis does not introduce more measurement error when the localization of the landmarks is determined by hand (Gravely and Benzies, 1974). However, other research has shown that there are statistically significant differences in landmark identifications between original and digitized cephalometric radiographs (Chen *et al.*, 2000). More recent research carried out by the same authors concluded that the differences between the measurements derived from the landmarks on original cephalometric radiographs and those identified on their digitized counterparts were statistically significant but clinically acceptable. The inter-observer errors of cephalometric measurements on digitized images were generally comparable with those from original radiographs (Chen *et al.*, 2004).

The aim of this investigation was to compare the classic method of tracing by hand with a computerized method, where the lateral cephalograms were scanned at 300 dpi and digitized onscreen. The inter- and intra-observer errors were investigated for tracing and digitizing errors.

Materials and methods

Thirty lateral cephalograms were randomly selected from the patient files at the Department of Orthodontics, Yeditepe University. These radiographs were scanned into digital format at 300 dpi using an Epson 1680 Pro scanner, with 1600 dpi imaging 40 800 pixels per line and 48-bit colour depth for both film and reflective scanning, and displayed on a 15-inch 1024 × 768 high-pixel resolution Benq FP581 monitor with pixel pitch of 0.297 mm, a contrast ratio of 450:1, and a brightness of 250 cd/m^2. All the scanned bitmap images of radiographs were then processed twice by two operators (KS and FI) using Dolphin Imaging Software 9.0 (Los Angeles, California, USA). The same radiographs were then traced twice with a 0.1 mm drawing pen by the same two operators, and measured using a ruler and protractor. There was at least a 3-week interval between the two landmark identification sittings of any radiograph (Figure 1).

Nine dental, 21 skeletal, and three soft tissue parameters were measured, which consisted of 18 angular measurements, 13 linear measurements, and four ratios (Table 1).

Intraclass correlation coefficients (ICC) were used to determine intra- and inter-rater agreement for each cephalometric variable. ICC, derived from analysis of variance, assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. It is a measure of the homogeneity of elements within clusters and has a maximum value of 1 when there is complete homogeneity (Kish, 1965).

Results

The results showed that each operator was consistent in the repeated measurements; all ICC were greater than or equal to 0.90 and that none of the 95 per cent confidence limits on these ICC had a lower boundary that was less than 0.84. ICC of 0.75 or above are usually considered to be good and above 0.9 to be excellent (Table 2). Inter-rater agreement also showed correlation coefficients greater than 0.75 (Table 2). The angles maxillary height, maxillary depth, *y*-axis, FMA, and nasolabial, and the distance N perpendicular had a wider reliability interval and lower correlation than other parameters tested.

Discussion

In modern orthodontics, quantitative, systematic, and objective measurements based on hard and soft tissue landmarks determined on cephalometric films are used on a daily basis. Precision and reproducibility in data obtained from cephalometrics is important for the orthodontist. Errors in conventional methods arise from radiographic acquisition, landmark identification, and measurement (Houston *et al.*, 1986; Forsyth *et al.*, 1996a,b). The progress in computer technology in orthodontics has not only resulted in ease of image archiving, image manipulation, transmission, and the possibility of enhancement, but also raised questions on the validity as well as the reproducibility of digital cephalometrics (Forsyth *et al.*, 1996a).

ICC calculation was used in the present study in order to determine if the results for two intra- or inter-group measurements presented congruity. *t*-tests were not used, because these compare the means of two groups, which could have caused mathematical errors in the results. Since a deviation in a few values could affect group means, correlation and agreement was carried out in the assessment of the data. The data in previous similar studies should be evaluated accordingly.

Since there was high agreement between the repeated measurements of each investigator, only one set of measurements were used in the inter-observer agreement evaluation. For both manual and digital measurements, inter-rater agreement indicated a high correlation (Table 2). The angles

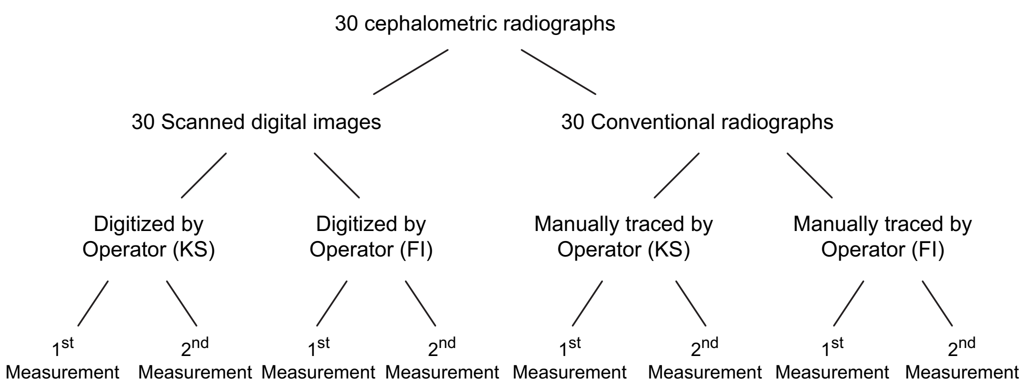

Figure 1 Flow chart of the study design showing the number of tracings and measurements undertaken by each operator.

Table 1 The cephalometric variables used in the study.

SNA (°)	Angle determined by points S, N, and A
SNB (°)	Angle determined by points S, N, and B
ANB (°)	Angle determined by points A, N, and B
Maxillary depth (°)	Angle formed between FH and NA planes
GoMeSN (°)	Angle formed between Go–Me and SN planes
Saddle (°)	Angle determined by points N, S, and Ar
Ar (°)	Angle determined by points S, Ar, and Go
Go (°)	Angle determined by points Ar, Go, and Me
Maxillary height (°)	Angle determined by points N, CF, and A
FMA (°)	Angle formed between FH plane and the mandibular plane
y-axis (°)	Angle formed between FH plane and S–Gn
SNOcc (°)	Angle formed between SN and occlusal planes
SN (mm)	Distance between points S and N
SAr (mm)	Distance between points S and Ar
NperA (mm)	Perpendicular distance from point A to a line perpendicular to the FH plane from point N
PogNB (mm)	Perpendicular distance from pogonion to the plane between points N and B
Ar–Go (mm)	Distance between points Ar and Go
N–Me (mm)	Distance between points N and Me
Ans–Me (mm)	Distance between points Ans and Me
Jarabak (ratio)	The ratio between posterior and anterior face heights (S–Go/N–Me)
AnsMe/NMe (ratio)	Ratio of lower (Ans–Me) to total (N–Me) face height
SAr/ArGo (ratio)	The ratio between posterior cranial base (S–Ar) and ramus (Ar–Go)
Go (ratio)	The ratio between the upper and lower parts of the gonial angle bisected by a line from point N
U1–SN (°)	Angle formed between the axis of the maxillary incisor to SN plane
IMPA (°)	Angle formed by the intersection of the mandibular incisor axis to the mandibular plane
U1–NA (°)	Angle formed by the intersection of the maxillary incisor axis to the plane between points N and A
L1–NB (°)	Angle formed by the intersection of the mandibular incisor axis to the plane between points N and B
Interincisal (°)	Angle formed by the intersection of the mandibular incisor axis to the maxillary incisor axis
Overjet (mm)	Horizontal distance between the tips of the maxillary and mandibular central incisors
Overbite (mm)	Vertical distance between the tips of the maxillary and mandibular central incisors
U1–NA (mm)	Perpendicular distance from the tip of the maxillary incisor to the plane between points N and A
L1–NB (mm)	Perpendicular distance from the tip of the mandibular incisor to the plane between points N and B
Nasolabial (°)	Angle determined by points collumella, SN, and UL
ULE (mm)	Perpendicular distance from the upper lip point to E line
LLE (mm)	Perpendicular distance from the lower lip point to E line

maxillary height, maxillary depth, *y*-axis, FMA, and nasolabial, and the distance N perpendicular, showed lower correlation and a relatively wider reliability interval. All these parameters, showing lower correlations, except nasolabial angle, are measurements related to the Frankfort horizontal plane, which passes through porion and orbitale. Porion has also been previously reported to cause problems regarding precision and accuracy (Chen *et al.*, 2000). Nasolabial angle on the other hand, depends on landmarks that are placed on a curve with wide radii which show proportionally greater errors of measurement (Baumrind and Frantz, 1971a). This type of error can be made regardless of the method (digital–manual) used for measurement. Even points articulare and gonion, which were estimated when digitized compared with construction of these points in manual drawing, displayed a very high correlation in intra- and inter-group analyses. Higher errors regarding these parameters, which could exhibit large deviations in a few manual tracings, may arise due to the comparison of means used in *t*-tests.

An important source of error in landmarks is image quality. Dolphin software allows for enhancement of the cephalogram, which is advantageous especially while precisely marking soft tissue profile landmarks. On the other hand, according to Geelen *et al.* (1998), image quality is already determined during exposure of analogue films and processing of the image, and little can be done to subsequently improve image quality; authors who share this idea have suggested that analogue has more detail than digital, and even though digital can be enhanced, this would only increase reproducibility and not precision (Macrì and Wenzel, 1993). However, the loss of detail that occurs when an image is compressed into JPEG format does not significantly affect the diagnostic quality of the image when standard compression settings are used (MacMahon *et al.*, 1991; Goldberg *et al.*, 1994). If the film is scanned and transferred to digital format, such as in the present study, the quality of the original film is one of the most important criteria in the validity of the result.

Conclusion

The validity and reproducibility of the measurements with the Dolphin Imaging Software and with the conventional method are highly correlated. When the advantages of digital imaging such as archiving, transmission, and enhancement are taken into consideration, the digitized method could be preferred in daily use and for research purposes without loss of quality.

Address for correspondence

Dr Korkmaz Sayinsu
Department of Orthodontics
Faculty of Dentistry
Yeditepe University
Bagdat cad. No. 238
Goztepe
TR-81006 Istanbul
Turkey
E-mail:drkorkmaz@yeditepe.edu.tr

Table 2 Intraclass correlation coefficients and 95 per cent confidence interval for intra- and inter-rater agreement.

	Operator I, Dolphin/manual	Operator II, Dolphin/manual	Inter-operator, Dolphin	Inter-operator, manual
SNA	0.97 (0.94–0.99)	0.97 (0.95–0.98)	0.97 (0.95–0.98)	0.96 (0.93–0.98)
SNB	0.99 (0.98–0.99)	0.97 (0.95–0.97)	0.98 (0.96–0.99)	0.97 (0.94–0.98)
ANB	0.98 (0.95–0.99)	0.97 (0.94–0.98)	0.95 (0.91–0.98)	0.96 (0.91–0.98)
SNB	0.99 (0.98–0.99)	0.97 (0.94–0.98)	0.99 (0.97–0.99)	0.97 (0.94–0.98)
SAr	0.98 (0.96–0.99)	0.97 (0.94–0.98)	0.98 (0.96–0.99)	0.96 (0.93–0.98)
Maxillary depth	0.90 (0.84–0.96)	0.95 (0.89–0.97)	0.92 (0.88–0.93)	0.90 (0.88–0.93)
GoMeSN	0.99 (0.98–0.99)	0.98 (0.96–0.99)	0.96 (0.93–0.98)	0.96 (0.92–0.98)
Saddle	0.95 (0.90–0.97)	0.97 (0.93–0.98)	0.98 (0.96–0.99)	0.93 (0.86–0.96)
Ar	0.92 (0.93–0.96)	0.98 (0.96–0.99)	0.98 (0.96–0.99)	0.97 (0.94–0.98)
Go	0.95 (0.90–0.97)	0.96 (0.93–0.98)	0.96 (0.92–0.98)	0.94 (0.89–0.97)
Saddle + Ar + Go	0.96 (0.92–0.98)	0.93 (0.90–0.96)	0.98 (0.96–0.99)	0.90 (0.88–0.95)
Jarabak	0.98 (0.95–0.99)	0.95 (0.90–0.98)	0.96 (0.92–0.98)	0.96 (0.92–0.98)
ANS–Me	0.97 (0.91–0.99)	0.92 (0.88–0.95)	0.94 (0.90–0.97)	0.90 (0.89–0.94)
Maxillary height	0.92 (0.89–0.96)	0.90 (0.93–0.97)	0.90 (0.88–0.95)	0.90 (0.87–0.97)
SArRamus	0.94 (0.89–0.97)	0.96 (0.93–0.98)	0.91 (0.89–0.95)	0.91 (0.89–0.96)
Go (ratio)	0.96 (0.92–0.98)	0.93 (0.88–0.96)	0.98 (0.96–0.99)	0.92 (0.89–0.96)
FMA	0.93 (0.88–0.96)	0.95 (0.92–0.97)	0.95 (0.90–0.97)	0.91 (0.87–0.95)
y-axis	0.94 (0.84–0.97)	0.91 (0.89–0.96)	0.91 (0.88–0.96)	0.93 (0.90–0.96)
SNOcc	0.96 (0.90–0.98)	0.97 (0.81–0.98)	0.98 (0.96–0.99)	0.94 (0.92–0.97)
U1–SN	0.91 (0.91–0.96)	0.97 (0.94–0.98)	0.94 (0.89–0.97)	0.95 (0.90–0.97)
IMPA	0.95 (0.88–0.97)	0.99 (0.98–0.99)	0.98 (0.96–0.99)	0.96 (0.92–0.98)
L1–NB	0.90 (0.85–0.95)	0.99 (0.98–0.99)	0.98 (0.97–0.99)	0.90 (0.87–0.95)
PogNB	0.95 (0.91–0.97)	0.96 (0.93–0.98)	0.96 (0.93–0.98)	0.96 (0.92–0.98)
Nasolabial	0.90 (0.84–0.94)	0.91 (0.91–0.96)	0.95 (0.90–0.97)	0.94 (0.88–0.97)
ULE	0.97 (0.95–0.98)	0.98 (0.96–0.99)	0.98 (0.96–0.99)	0.97 (0.95–0.98)
LLE	0.99 (0.99–0.99)	0.99 (0.98–0.99)	0.96 (0.93–0.98)	0.97 (0.94–0.98)
NperA	0.91 (0.88–0.95)	0.93 (0.90–0.96)	0.93 (0.89–0.97)	0.91 (0.87–0.98)

Acknowledgement

We would like to express our sincere thanks to Rana Konyalioglu for her kind help with the biostatistics.

References

Battagel J M 1993 A comparative assessment of cephalometric errors. European Journal of Orthodontics 15: 305–314

Baumrind S, Frantz R C 1971a The reliability of head film measurements. 1. Landmark identification. American Journal of Orthodontics 60: 111–127

Baumrind S, Frantz R C 1971b The reliability of head film measurements. 2. Conventional angular and linear measures. American Journal of Orthodontics 60: 505–517

Chen Y J, Chen S K, Chang H F, Chen K C 2000 Comparison of landmark identification in traditional versus computer-aided digital cephalometry. Angle Orthodontist 70: 387–392

Chen Y J, Chen S K, Yao J C, Chang H F 2004 The effects of differences in landmark identification on the cephalometric measurements in traditional versus digitized cephalometry. Angle Orthodontist 74: 155–161

Cohen A M 1984 Uncertainty in cephalometrics. British Journal of Orthodontics 11: 44–48

Forsyth D B, Shaw W C, Richmond S 1996a Digital imaging of cephalometric radiography, part 1: advantages and limitations of digital imaging. Angle Orthodontist 66: 37–42

Forsyth D B, Shaw W C, Richmond S, Roberts C T 1996b Digital imaging of cephalometric radiographs, part 2: image quality. Angle Orthodontist 66: 43–50

Geelen W, Wenzel A, Gotfretsen E, Kruger M, Hansson L G 1998 Reproducibility of cephalometric landmarks on conventional film, hardcopy, and monitor-displayed images obtained by the storage phosphor technique. European Journal of Orthodontics 20: 331–340

Goldberg M A, Pivovarov M, Mayo-Smith W W 1994 Application of wavelet compression to digital radiographs. American Journal of Roentgenology 163: 463–468

Gravely J F, Benzies P M 1974 The clinical significance of tracing error in cephalometry. British Journal of Orthodontics 1: 95–101

Houston W J B 1983 The analysis of errors in orthodontic measurements. American Journal of Orthodontics 83: 382–390

Houston W J B, Maher R E, McElroy D, Sherriff M 1986 Sources of error in measurements from cephalometric radiographs. European Journal of Orthodontics 8: 149–151

Jackson P H, Dickson G C, Birnie D J 1985 Digital image processing of cephalometric radiographs: a preliminary report. British Journal of Orthodontics 12: 122–132

Kamoen A, Dermaut L, Verbeeck R 2001 The clinical significance of error measurement in the interpretation of treatment results. European Journal of Orthodontics 23: 569–578

Kish L 1965 Survey sampling. John Wiley & Sons, Inc., New York

MacMahon H, Doi K, Sanada S 1991 Data compression: effect on diagnostic accuracy in digital chest radiography. Radiology 178: 175–179

Macrì V, Wenzel A 1993 Reliability of landmark recording on film and digital lateral cephalograms. European Journal of Orthodontics 15: 137–148

Melsen B, Baumrind S 1995 Clinical research application of cephalometry. In: Athanasiou A (ed). Orthodontic cephalometry. Mosby-Wolfe, St Louis, pp. 181–202

Richardson A 1981 A comparison of traditional and computerized methods of cephalometric analysis. European Journal of Orthodontics 3: 15–20

Sandler P J 1988 Reproducibility of cephalometric measurements. British Journal of Orthodontics 15: 105–110

Turner P J, Weerakone S 2001 An evaluation of the reproducibility of landmark identification using scanned cephalometric images. Journal of Orthodontics 28: 221–229

Copyright of European Journal of Orthodontics is the property of Oxford University Press / UK and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.