European Journal of Orthodontics 32 (2010) 721–728
doi:10.1093/ejo/cjq009
Advance Access Publication 16 June 2010

A comparison of hand-tracing and cephalometric analysis computer programs with and without advanced features—accuracy and time demands

Georgios Tsorovas and Agneta Linder-Aronson Karsten
Department of Orthodontics, Institute of Odontology, Karolinska Institutet, Huddinge, Sweden

Correspondence to: Georgios Tsorovas, Department of Orthodontics, Institute of Odontology, Karolinska Institutet, Box 4064, 14104 Huddinge, Sweden. E-mail: geotsoros@gmail.com

SUMMARY The aim of this study was to evaluate the basic and advanced features of five different cephalometric analysis computer programs. The level of measurement agreement with hand-tracing and time demands was examined. The material consisted of 30 digital lateral radiographic images. Twenty-three measurements were calculated by one operator both manually and using five different cephalometric analysis software programs. Intraclass correlation coefficient (ICC) was used to detect differences in measurement agreement between hand-tracing and basic features as well as between hand-tracing and advanced features. Coefficient of variation (CV) was used to assess intra-user error and a Student's *t*-test to determine time differences.

Of the 23 measurements tested for each procedure, one [(Ii to NB (mm)] showed better agreement with hand-tracing when the advanced features were used, 20 showed good agreement with hand-tracing for both basic and advanced features, while two (AB on FOP and Ii to A/Pog) showed poor intra-user reproducibility. Hand-tracing took a significantly longer time ($P < 0.001$) than both the basic and advanced features. The advanced features took a significantly longer time ($P < 0.001$) than the basic features.

Both basic and advanced features showed good measurement agreement with the hand-tracing technique. The use of the basic features minimizes the time requirements for analysis. A computerized tracing technique, which consists of either basic or advanced feature, can be regarded as less time consuming and equally reliable to hand-tracing as far as cephalometric measurements are concerned.

Introduction

Since the introduction of cephalometry by Broadbent (1931), the technique has been an important tool to orthodontists studying dental malocclusions and underlying skeletal discrepancies. The applications for cephalometric analysis include case diagnosis, treatment planning, evaluation of treatment results, and prediction of growth (Steiner, 1960). Manual analysis is performed by tracing radiographic landmarks on acetate overlays and using these landmarks to measure the desired linear and angular values. This process can be time-consuming and the linear and angular cephalometric measurements obtained manually with a ruler and a protractor may be prone to error.

Rapid advances in computer science have led to widespread application of computers in cephalometry (Rudolph *et al.*, 1998). When using computer software programs for cephalometric analysis, the landmarks are usually digitized first. The software can thereafter generate the values of cephalometric measurements instantaneously when the locations of all the required landmarks have been entered. Digital cephalometric films can be integrated with the patient records to establish a computer-based filing system and to take advantage of image processing, storage, and transmission (Forsyth *et al.*, 1996a).

Many studies have investigated the reproducibility of both landmarks and measurements of hand-tracing versus digitized cephalometry. Jackson *et al.* (1985) compared some common cephalometric landmarks sampled by manual point identification on film with those acquired by digital sampling on images displayed on a video monitor. They found that the results from the digital image system were comparable with those obtained with the hand-tracing method. Döler *et al.* (1991) showed an improvement in image quality of digital cephalograms when using various digital enhancement and filtering techniques. Macrì and Wenzel (1993) concluded, however, that the reliability of landmark location on digital images was inferior to conventional film when a low-cost black-and-white video camera and a spatial resolution of 512 × 512 pixels was used for digitization, and that digital image processing only increased reliability when good quality original films were used.

Various investigators have evaluated the use of computerized cephalometrics and the digitizing process of cephalometric radiographs (Nimkarn and Miles, 1995; Lim and Foong, 1997; Geelen *et al.*, 1998; Rudolph *et al.*, 1998; Chen *et al.*, 2000; Liu *et al.*, 2000). Those investigators compared several digitizing methods, such as phosphor plates, video imaging, and flatbed scanners with digital lateral cephalogram analogue methods. However, the results concerning the preferred method were contradictory. Geelen *et al.* (1998) compared the reproducibility of landmarks on conventional film with hard copy and monitor-displayed storage phosphor images. In that study, there was a large variation in landmark identification between the imaging modalities. A pilot study by Gotfredsen *et al.* (1997) regarding the effect of irreversible digital image compression demonstrated that the loss of resolution was not associated with a significant difference in reproducibility. From these studies, it appears that landmark recording on digital images is as reproducible as with conventional films.

Many commercially available or customized programs have been developed to perform cephalometric analysis directly on screen-displayed digital film. Such applications may substantially eliminate the need for hard copies of cephalometric films. In addition, benefits, such as the transmission and ease of processing (Forsyth *et al.*, 1996a), the elimination of the scanning procedure, and the convenient integration with the patients' files etc., cannot be overlooked. Furthermore, the substantial reduction in radiation exposure that can be achieved when direct digital cephalogram methods are used (Seki and Okano, 1993) is one additional reason for the gradual transition from analogue to digital cephalometrics.

The majority of the software available offers a number of features unfamiliar in the hand-tracing method. Most of these programs consist of basic and advanced features; the former is composed of insertion of the digital film in the software and analysis of the film without any alteration of the image. The latter includes several applications, starting from altering the properties of the image (brightness, contrast, zoom etc.) to an automatic landmark identification function. With regard to automatic computerized identification of cephalometric landmarks, it has been demonstrated that these automated systems are at present unable to compete with manual identification, in terms of accuracy of landmark position (Forsyth *et al.*, 1996b). Earlier studies revealed that computer-aided cephalometric analysis does not introduce more measurement error than hand-tracing, as long as landmarks are identified manually (Gravely and Benzies, 1984; Enlow and Hans, 1996). Therefore, manually identifying landmarks on screen-displayed digital images for cephalometric analysis may still be the best strategy.

Although the automatic landmark identification procedure has been investigated, none of the additional features offered by the software programs have been adequately assessed in terms of clinical usefulness. The addition of all these extra features is claimed to offer improved accuracy in the process of analysis of a lateral cephalogram. However, research that confirms the above is limited. Furthermore, no information on the extra time required when using these features has been provided.

The aim of this study was to assess whether measurements calculated when using the basic and advanced features of five different cephalometric analysis computer programs available on the market are in good agreement with those derived when using the hand-tracing technique. The null hypotheses tested were:

1. there is no difference in agreement between the measurements derived using the basic features of computerized tracing programs and those obtained by hand-tracing, and
2. there is no difference in agreement between the measurements derived using the advanced features of computerized tracing programs and those obtained by hand-tracing.

Furthermore, an evaluation of the time needed to perform the analysis using all three methods (hand-tracing and basic and advanced features of the programs) was performed.

Materials and methods

Thirty digital lateral cephalometric films collected randomly from patients who had received orthodontic treatment in the Orthodontic Department, Institute of Odontology, Karolinska Institutet, were used. Ethical approval was obtained from the Ethics committee of Stockholm (Protocol 2006/4:10). Exclusion criteria were (1) unerupted or missing incisors and (2) unerupted teeth overlying the apices of the incisors. The resolution of the digital images was 300 d.p.i. with 256 grey levels, and the image size was 1360 × 1018 pixels. The 30 digital films were transferred to conventional films (Fujifilm® dry imaging film DI-AL, Tokyo, Japan) using a Fujifilm® FM-DPL printer. Calibration of both the digital image and the hard copy was based on the measurement of a known distance between two points on the phosphor plate. Neither the digital nor the conventional film contained patient data. The 30 printed films were hand-traced by one observer (GT). Twenty-seven commonly used skeletal and dental landmarks were selected, which produced 23 measurements (Figure 1). Trace foil (3M Unitek® Corporation, Monrovia, California, USA), a 4H pencil, and a cephalometric protractor (3M Unitek® Corporation) were used to measure the variables. These served as the control to compare the measurements deriving from the digital analysis. Furthermore, the procedure of landmark identification, tracing, and measurement was followed step-by-step, and the time spent for each procedure was recorded in seconds. The same observer undertook these time registrations using a digital watch.

The decision regarding which programs were to be included in the study was made using the following procedure. Fourteen companies that produce cephalometric analysis software were contacted via e-mail with a request

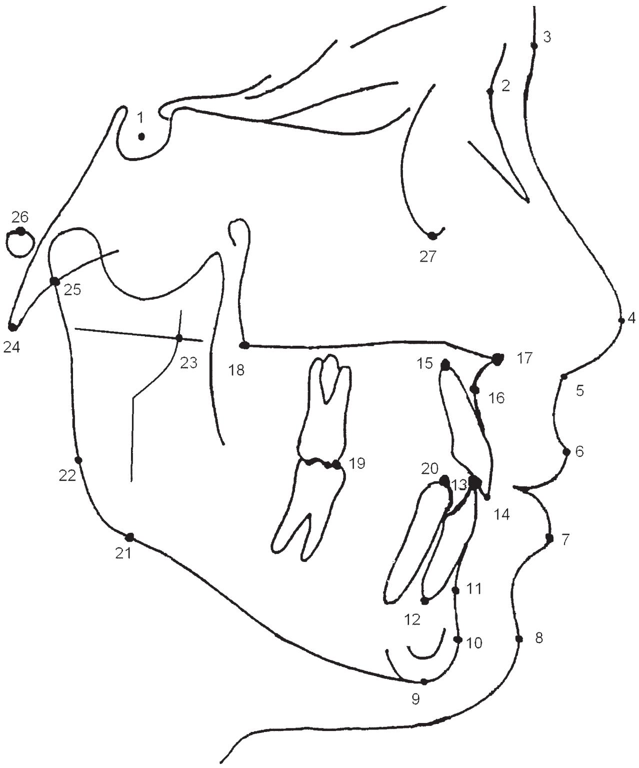

Figure 1 Location and definitions of the 27 landmarks used in the study: 1: Sella (S), the midpoint of sella turcica; 2: Nasion (N), junction of the frontal and nasal bones at the naso-frontal suture; 3: Glabella (G′), the most anterior point on the forehead, in the region of the supra-orbital ridges; 4: Pronasale (Pr′), the most anterior point on the nasal tip; 5: Subnasale (Sn′), the junction of the columella of the nose with the philtrum of the upper lip; 6: Labrare Superios. (Ls), the muco-cutaneous junction of the upper lip and philtrum; 7: Labrare Inferios. (Li), the muco-cutaneous junction of the lower lip and philtrum; 8: Soft Pogonion (Pg′), the most anterior point on the soft tissue chin; 9: Menton (Me), the most inferior point on the bony chin; 10: Pogonion (Pg), the most anterior point on the bony chin; 11: Point B, the deepest point in the concavity of the anterior mandible between the alveolar crest and pogonion; 12: Lower incisor apex, the root apex of the lower central incisor; 13: Lower incisor tip, the tip of the crown of the lower central incisor; 14: Upper incisor tip, the tip of the crown of the upper central incisor; 15: Upper incisor apex, the root apex of the upper central incisor; 16: Point A, the deepest point in the concavity of the anterior maxilla between anterior nasal spine and the alveolar crest; 17: Anterior nasal spine (ANS), the anterior limit of the floor of the nose, at the tip of anterior nasal spine; 18: Posterior nasal spine (PNS), the posterior limit of the floor of the nose, at the tip of posterior nasal spine; 19: Lower molar crown, the tip of the mesial cusp of the lower first molar; 20: Lower first premolar tip, the tip of the crown of the lower first premolar; 21: Inferior gonion, a mid-planed point at a tangent to the inferior border of the mandible near Gonion; 22: Posterior gonion, a mid-planed point at a tangent to the posterior border of the mandible near gonion; 23: Ad1, a landmark located at the intersection of the line between PNS and basion with the posterior nasopharyngeal wall; 24: Basion (Ba), the most inferior point on the anterior margin of the foramen magnum; 25: Articulare (Ar), a mid-planed point located at the intersection of the posterior border of the ramus with the inferior surface of the cranial base; 26: Porion (Po), the most superior point of the bony external auditory meatus; 27: Orbitale (Or), the most inferior point on the infra-orbital margin. Location and definitions of the 23 measurements used in the study: SN–FH (°), angle determined by the SN and the Frankfort Horizontal (FH) plane; SNA (°), angle determined by points S, N, A; SNB (°), angle determined by points S, N, B; ANB (°), angle determined by points A, N, B; AB on FOP (mm), linear distance between points A and B functional occlusal plane parallel to the (line formed by the tip of the mesio-buccal cusp of the lower first molar and the tip of the buccal cusp of the lower first premolar); Is–ANS/PNS (°), angle formed by the intersection of the maxillary incisor axis to the ANS/PNS plane; Is–NA (°), angle formed by the intersection of the maxillary incisor axis to the NA plane; Is–NA (mm), perpendicular distance from the tip of the maxillary incisor to the plane between points N and A; Ii–MP (°), angle formed by the intersection of the mandibular incisor axis to the Mandibular plane (MP) formed by the points inferior to gonion and menton; Ii–NB (°), angle formed by the intersection of the mandibular incisor axis to the plane between points N and B; Ii–NB (mm), perpendicular distance from the tip of the mandibular incisor to the plane between points N and B; Ii–A/Pog (mm), perpendicular distance from the tip of the mandibular incisor to the plane between point A and Pog; Gonial angle (°), angle formed between the Mandibular plane and the plane formed by points posterior Gonion and Articulare; SN–MP (°), angle formed between the SN plane and the MP plane; ANS/PNS–SN (°), angle formed between ANS/PNS plane and the SN plane; ANS/PNS–MP (°), angle formed between ANS/PNS plane and the MP plane; N–S–Ba (°), angle formed between the SN plane and the SBa plane; N–ANS/N–Me (ratio), ratio between upper anterior face height and total face height; Pog–NB (mm), perpendicular distance from Pg to the plane between points N and B; Ad1–pm (mm), distance between the Ad1 point and the PNS point; ULE (mm), perpendicular distance from the upper lip point to the E-line; LLE (mm), perpendicular distance from the lower lip point to the E-line; G′–Sn′–Pg′ (°) angle formed between Glabella, Subnasale and soft tissue Pogonion.

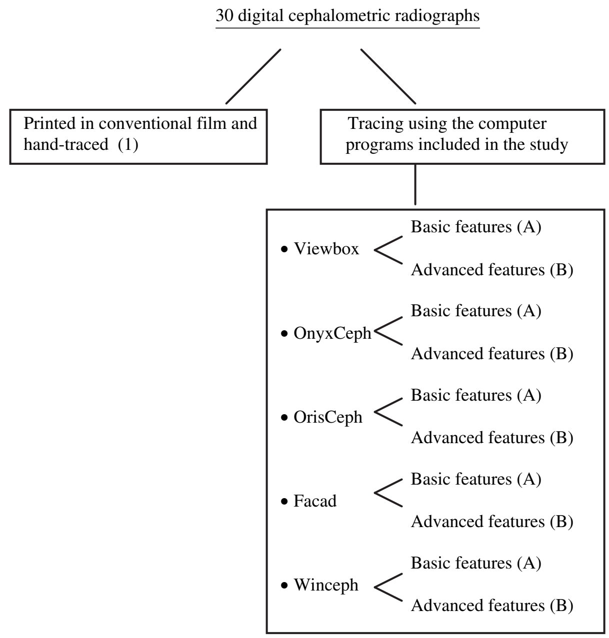

Figure 2 Flow chart of the study design. The level of agreement was analysed using Intraclass Correlation Coefficient between measurements deriving from method (1) and (A) as well as between method (1) and (B) for each program.

to allow free of charge use of their product. The five companies that gave permission were included in the study [Viewbox® (dHAL Software, Kifisia, Greece), OnyxCeph® (Image Instruments GmbH, Frankfurt, Germany), OrisCeph® (Elite Computer Italia, Vimodrone, Italy), Facad® (Ilexis AB, Linköping, Sweden), and Winceph® (Rise Corporation, Sendai, Japan)]. The 30 digital films were analysed twice using the five programs by the same observer, once using the basic features and once the advanced features. The time interval between each procedure was at least 1 month.

The digital images were displayed on a 1280 × 1024–pixel monitor (Philips®, Amsterdam, The Netherlands) and the landmarks were identified. The basic features of each program included insertion of the digital image in actual dimensions and the following analysis without the aid of any enhancement tool. In addition to the above, the advanced features included the use of any enhancement tool that the observer felt was necessary in order to produce the most precise landmark identification, according to his perception. Examples of such enhancement tools included zoom in and zoom out, brightness and contrast adjustments, noise reduction, and sharpening. The time corresponding to film insertion and landmark identification was recorded in this case since the software produced the measurements. The same observer, who has experience in cephalometric analysis, completed all measurements. A flow chart of the study design is shown in Figure 2.

Intraclass correlation coefficients (ICCs) were calculated in order to determine the agreement for each cephalometric value between those deriving from hand-tracing and those when the basic features of the programs were used. ICC was also used to determine the agreement for each cephalometric value between those of hand-tracing and those when the advanced features of the programs were used. The intra-user error for each method was assessed using Dahlberg's standard error formula (total error = $\sqrt{\Sigma d^2/2n}$). In order to determine measurement error, the coefficient of variance (CV% = standard error/mean value × 100) measurement was used (Atkinson and Nevill, 1998). Intra-user reproducibility was considered poor for a CV greater than 10 per cent. In order to calculate intra-user agreement, 15 films were randomly selected, in both their digital and printed form, and reanalysed using all three methods (hand-tracing and basic and advanced features for all programs), with a time interval of at least 1 month between the first and second analysis. This procedure was completed by the same observer. A Student's *t*-test was used to study differences regarding timing, and a Kolmogorov–Smirnov test to assess whether time registrations were normally distributed. A statistically significant limit was set at $P < 0.001$.

Comparison between the total time needed for hand-tracing and that required to perform the analysis using the basic and advanced features for each computer program was undertaken. Furthermore, the time needed for the basic features was compared with that required for the advanced features for each software program. Finally, the time necessary for landmark identification with the hand-tracing procedure, which can be considered as the equivalent of the whole digital procedure, was compared with the time needed for both the basic and the advanced feature procedures. The level of cephalometric measurement agreement between the two methods (basic and advanced) and the hand-tracing procedure was set as follows: low agreement for ICC ≤0.75, good agreement for ICC >0.75 (Fayers and Hays, 2005).

Results

The ICC and 95 per cent confidence interval (CI) for the level of agreement between hand-tracing and basic features as well as between hand-tracing and advanced features for each program are shown in Tables 1 and 2. For 21 of the 23 measurements, there were no significant differences in the level of agreement between the use of basic features and hand-tracing or between the use of advanced features and hand-tracing. The first exception was the AB on FOP measurement, which showed an ICC ≤0.75 (low agreement) for all programs when the basic features of the software were used. However, the same measurement showed an ICC >0.75 (good agreement) for all programs when the advanced features of the software were used. The second measurement that showed significant differences in agreement according to whether the basic or advanced features were used was Ii–NB (mm). For this measurement, agreement with the hand-tracing procedure was also better when the advanced features of the software were used. This was true for four of the five software programs used in the study. The only exception was the OnyxCeph® program where the ICC for Ii–NB (mm) was above 0.75 for both the basic and the advanced features (ICC of 0.76 and 0.84, respectively).

Regarding intra-user error, 21 of the 23 measurements constantly showed a CV below 10 per cent regardless of the method (Table 3, available as online supplementary data). Only AB on FOP and lower incisor to A–Pog measurements showed poor reproducibility for all methods (CV ranging from 28 to 38 and 15 to 24 per cent, respectively).

The mean times registered for the hand-tracing and the digital cephalometric analysis are presented in Table 4. Both the basic and the advanced feature procedures took significantly less time ($P < 0.001$) than the total time needed for hand-tracing. Furthermore, the basic features procedure took significantly less time ($P < 0.001$) than the advanced features procedure, independent of the software. Lastly, the time recorded to perform the analysis using the basic features of all programs was always significantly less ($P < 0.001$) than that needed to perform the landmark identification part of the hand-tracing procedure. In contrast, the time needed to perform the analysis using the advanced features of all programs was always significantly longer ($P < 0.001$) than that necessary to perform the landmark identification part of the hand-tracing procedure.

Discussion

It is important to know whether spending additional time performing a cephalometric analysis will provide more accurate findings. In this research, the basic and advanced

Table 1 Intraclass correlation coefficient (ICC) and 95% confidence interval (CI) for hand-tracing (HD) and Viewbox® (VB), Facad® (FD) and Winceph® (WC) basic features and for the advanced features of the same programs. A, advanced features show better agreement; P, poor intra-user reproducibility.

Measurement	HD versus VB basic		HD versus VB advanced		HD versus FD basic		HD versus FD advanced		HD versus WC basic		HD versus WC advanced	
	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI
SN–FH (°)	0.87	0.73–0.94	0.96	0.90–0.98	0.84	0.72–0.93	0.97	0.93–0.94	0.85	0.79–0.92	0.96	0.92–0.98
SNA (°)	0.89	0.85–0.98	0.96	0.92–0.98	0.90	0.84–0.98	0.97	0.95–0.98	0.89	0.86–0.94	0.96	0.95–0.98
SNB (°)	0.83	0.82–0.92	0.98	0.95–0.99	0.89	0.89–0.95	0.97	0.94–0.99	0.97	0.93–0.99	0.98	0.95–0.99
ANB (°)	0.89	0.77–0.95	0.94	0.88–0.97	0.88	0.77–0.94	0.95	0.89–0.97	0.87	0.73–0.94	0.92	0.89–0.96
AB on FOP (mm)	0.75*	0.53–0.87	0.81	0.84–0.91 A,P	0.75*	0.65–0.88	0.81	0.72–0.91 A,P	0.74*	0.68–0.92	0.86	0.74–0.90 A,P
Is–ANS/PNS (°)	0.91	0.83–0.96	0.95	0.90–0.98	0.89	0.84–0.97	0.96	0.93–0.98	0.93	0.86–0.97	0.94	0.87–0.97
Is–NA (°)	0.96	0.85–0.98	0.94	0.91–0.97	0.92	0.88–0.97	0.95	0.90–0.97	0.96	0.91–0.98	0.95	0.90–0.98
Is–NA (mm)	0.85	0.74–0.93	0.91	0.86–0.95	0.86	0.71–0.94	0.87	0.80–0.94	0.85	0.74–0.93	0.89	0.77–0.94
Ii–MP (°)	0.92	0.80–0.97	0.94	0.91–0.97	0.93	0.82–0.97	0.94	0.87–0.97	0.92	0.82–0.97	0.94	0.88–0.97
Ii–NB (°)	0.93	0.83–0.97	0.96	0.92–0.98	0.93	0.83–0.97	0.96	0.91–0.98	0.93	0.83–0.97	0.96	0.92–0.98
Ii–NB (mm)	0.72*	0.47–0.88	0.85	0.75–0.93 A	0.67*	0.48–0.84	0.88	0.75–0.95 A	0.74*	0.67–0.88	0.85	0.77–0.93 A
Ii–A/Pog (mm)	0.76	0.68–0.94	0.89	0.78–0.95 P	0.78	0.66–0.94	0.90	0.81–0.95 P	0.86	0.68–0.94	0.88	0.76–0.94 P
Gonial angle (°)	0.81	0.75–0.91	0.93	0.90–0.97	0.83	0.67–0.91	0.96	0.94–0.99	0.83	0.67–0.91	0.95	0.90–0.97
SN–MP (°)	0.97	0.93–0.99	0.97	0.93–0.99	0.97	0.94–0.99	0.96	0.92–0.98	0.96	0.92–0.98	0.96	0.92–0.98
ANS/PNS–SN (°)	0.91	0.83–0.96	0.92	0.88–0.96	0.91	0.82–0.96	0.96	0.91–0.98	0.92	0.83–0.96	0.92	0.84–0.96
ANS/PNS–MP (°)	0.97	0.90–0.99	0.97	0.94–0.99	0.97	0.94–0.99	0.98	0.95–0.99	0.98	0.96–0.99	0.97	0.94–0.99
N–S–Ba (°)	0.95	0.90–0.98	0.94	0.90–0.97	0.94	0.98–0.97	0.95	0.90–0.98	0.95	0.90–0.98	0.97	0.94–0.99
N–ANS/N–Me	0.82	0.80–0.88	0.94	0.91–0.96	0.86	0.75–0.89	0.94	0.91–0.96	0.82	0.69–0.86	0.94	0.90–0.97
Pog–NB (mm)	0.83	0.78–0.92	0.91	0.81–0.96	0.78	0.58–0.89	0.91	0.81–0.96	0.83	0.68–0.92	0.91	0.81–0.95
Ad_1–pm (mm)	0.85	0.83–0.89	0.89	0.86–0.95	0.79	0.73–0.90	0.89	0.86–0.93	0.85	0.73–0.94	0.93	0.88–0.96
ULE (mm)	0.97	0.94–0.99	0.98	0.95–0.99	0.95	0.91–0.98	0.98	0.95–0.99	0.97	0.94–0.99	0.98	0.95–0.99
LLE (mm)	0.95	0.90–0.98	0.95	0.90–0.98	0.90	0.80–0.95	0.96	0.91–0.98	0.94	0.88–0.97	0.96	0.92–0.98
G'–Sn′–Pg′ (°)	0.97	0.93–0.98	0.95	0.90–0.99	0.96	0.92–0.98	0.99	0.98–0.99	0.97	0.93–0.98	0.95	0.90–0.98

*Low level of agreement between the specific method and hand-tracing.

Table 2 Intraclass correlation coefficient (ICC) and 95% confidence interval (CI) for hand-tracing (HD) and OrisCeph® (OR) basic features as well as for hand-tracing and OnyxCeph® (ON) basic features and for the advanced features of the same programs, A, advanced features show better agreement; P, poor intra-user reproducibility.

Measurement	HD versus OR basic		HD versus OR advanced		HD versus ON basic		HD versus ON advanced	
	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI
SN–FH (°)	0.92	0.82–0.97	0.93	0.92–0.96	0.89	0.75–0.94	0.92	0.86–0.94
SNA (°)	0.90	0.85–0.98	0.96	0.92–0.98	0.89	0.84–0.98	0.93	0.90–0.97
SNB (°)	0.83	0.67–0.91	0.97	0.95–0.99	0.96	0.93–0.98	0.97	0.94–0.99
ANB (°)	0.82	0.74–0.88	0.93	0.85—0.96	0.87	0.75–0.94	0.90	0.85–0.96
AB on FOP (mm)	0.70*	0.56–0.88	0.81	0.72–0.91 A, P	0.69*	0.56–0.88	0.86	0.72–0.93 A, P
Is–ANS/PNS (°)	0.92	0.84–0.96	0.92	0.83–0.96	0.92	0.83–0.96	0.95	0.89–0.98
Is–NA (°)	0.95	0.89–0.98	0.93	0.85–0.96	0.90	0.83–0.98	0.93	0.85–0.96
Is–NA (mm)	0.86	0.68–0.94	0.87	0.73–0.93	0.87	0.70–0.94	0.90	0.80–0.95
Ii–MP (°)	0.92	0.80–0.96	0.94	0.88–0.97	0.92	0.82–0.96	0.94	0.88–0.97
Ii–NB (°)	0.92	0.81–0.97	0.96	0.92–0.98	0.93	0.85–0.97	0.97	0.94–0.99
Ii–NB (mm)	0.72*	0.42–0.87	0.85	0.70–0.93 A	0.76	0.45–0.85	0.84	0.69–0.92
Ii–A/Pog (mm)	0.79	0.76–0.94	0.89	0.77–0.95 P	0.80	0.64–0.93	0.89	0.79–0.95 P
Gonial angle (°)	0.84	0.70–0.92	0.95	0.91–0.97	0.83	0.67–0.91	0.95	0.90–0.98
SN–MP (°)	0.98	0.96–0.99	0.95	0.89–0.97	0.97	0.93–0.98	0.97	0.94–0.99
ANS/PNS–SN (°)	0.91	0.82–0.96	0.95	0.89–0.98	0.91	0.82–0.96	0.95	0.90–0.98
ANS/PNS–MP (°)	0.97	0.95–0.99	0.97	0.94–0.99	0.97	0.93–0.98	0.97	0.95–0.99
N–S–Ba (°)	0.98	0.96–0.98	0.97	0.95–0.98	0.95	0.89–0.97	0.96	0.91–0.98
N–ANS/N–Me	0.80	0.62–0.90	0.97	0.94–0.99	0.81	0.73–0.92	0.94	0.90–0.99
Pog–NB (mm)	0.86	0.73–0.93	0.92	0.88–0.95	0.82	0.75–0.91	0.92	0.87–0.97
Ad_1–pm (mm)	0.78	0.72–0.92	0.91	0.87–0.95	0.76	0.72–0.90	0.89	0.81–0.95
ULE (mm)	0.92	0.76–0.97	0.98	0.95–0.99	0.95	0.89–0.97	0.97	0.93–0.99
LLE (mm)	0.96	0.91–0.98	0.96	0.91–0.98	0.89	0.79–0.95	0.96	0.91–0.98
G′–Sn′–Pg′ (°)	0.98	0.97–0.99	0.98	0.96–0.99	0.97	0.93–0.98	0.98	0.95–0.99

*Low level of agreement between the specific method and hand-tracing.

Table 4 Mean time in minutes (±SD) required for each procedure of hand-tracing and computer cephalometric analysis.

Hand-tracing cephalometric analysis									
Landmark location	Cephalometric tracing	Measurement	Total						
3.8 ± 0.7*	3.1 ± 0.7	9.1 ± 1.4	15.9 ± 2.2**						
Computer cephalometric analysis									
Viewbox®		Facad®		WinCeph®		OrisCeph®		OnyxCeph®	
Basic	Advanced	Basic	Advanced	Basic	Advanced	Basic	Advanced	Basic	Advanced
2.4 ± 0.3***	4.8 ± 0.2	2.6 ± 0.2***	4.9 ± 0.2	2.6 ± 0.3***	5.0 ± 0.2	2.7 ± 0.2***	4.8 ± 0.2	2.7 ± 0.2***	4.9 ± 0.2

*Significantly less ($P < 0.001$) than the advanced features procedure and significantly more ($P < 0.001$) than the basic feature procedure for all five programs.
**Significantly more ($P < 0.001$) than both the basic and advanced features procedure for all programs.
***Significantly less ($P < 0.001$) than the advanced features procedure of the same software.

features of five different cephalometric analysis computer programs were evaluated with respect to the level of their measurement agreement with the hand-tracing procedure and with their time demands. ICC and CI calculations were used to determine if the results between the basic features and hand-tracing, as well as between the advanced features and hand-tracing presented congruity. The ICC 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, 1995). CI is also an indicator of precision. A wider CI indicates lower precision, while a narrower CI indicates greater precision (Kirkwood and Sterne, 2001). *t*-tests were not used to assess agreement level as these compare the means of two groups, which could have resulted in mathematical errors. Since the group means could be easily affected by a deviation in a few values, correlation and agreement were preferred for the assessment of the data. This means that the data in previous similar studies should be evaluated accordingly.

For assessing intra-user error, the CV was used. CV values below 10 per cent were found, which are considered to be low (Machin *et al.*, 2007). A Student's *t*-test was used for the assessment of mean time differences. This was possible after using a Kolmogorov–Smirnov test to verify the normal distribution of the data registrations for each of the procedures under investigation. Overall, intra-user error assessment showed that the operator was consistent in the repeated measurements.

Twenty-one of the 23 measurements showed a CV below 10 per cent for all programs and methods used. This is not surprising since most of these measurements have been methodically studied (Geelen *et al.*, 1998, Chen *et al.*, 2000, Santoro *et al.*, 2006), which confirm their good reproducibility.

The AB on FOP and Ii to A/Pog measurements were the only ones that consistently showed poor intra-user reproducibility regardless of the method used. This finding is in agreement with those of similar studies (Geelen *et al.*, 1998; Chen *et al.*, 2000). Santoro *et al.* (2006) also found that these two measurements showed poor intra-user reproducibility for digital as well as hand-traced measurements. The high intra-user error for AB on FOP is usually explained by the difficulties in locating its landmarks, mainly point A and the FOP, which are noted for their poor reproducibility. Chen *et al.* (2000) described the identification of point A as very demanding because of overlapping of ANS and the upper incisors in the two-dimensional projection of the skull. Interestingly, it was only AB on FOP and Ii to A/Pog measurement that showed consistently poor reproducibility and not any other of the measurements based on point A, such as SNA or Is to NA (both mm and angle). This fact highlights the importance of FOP and confirms that it is difficult to define (Chen *et al.*, 2000) as its structures overlap. Furthermore, the AB on FOP is a small linear measurement that makes it sensitive even to small errors.

The last principle may also apply to Ii to A/Pog high intra-user error. Santoro *et al.* (2006) also failed to justify this finding with a more convincing argument. This did not allow any useful conclusions to be drawn regarding the superiority of the advanced features as far as these two measurements are concerned since the higher ICC and narrower CI for these two measurements, when the advanced features of the programs were used, may appear as a result of user inconsistency and be independent of the method.

Twenty-one measurements showed good reproducibility. Twenty revealed no differences in cephalometric measurement agreement with the hand-tracing procedure with regard to whether the basic or advanced features of each computer program were used. Only Ii–NB (mm) showed an ICC <0.75 for four out of the five programs (OnyxCeph® was the exception) when the basic features were used. For 20 of the 21 measurements that showed good reproducibility, both null hypotheses are accepted. The null hypothesis should be rejected for Ii–NB(mm), but only for the basic features.

As expected, time findings revealed that the hand-tracing procedure was, by far, the most time-consuming. Hand-tracing analysis took almost six times longer than when the basic features were used and almost four times longer than when the advanced features were used. Chen *et al.* (2004) also found the hand-tracing procedure to be the most time consuming. They concluded that hand-tracing time expectations are high, independent of user experience. Experience in hand-tracing does not increase the speed of the actual measurements which, in the present study, took more than 50 per cent of the time needed for the entire analysis procedure. Even more interesting were the findings when the time demands for the basic and advanced features of each program were analysed. For all the programs used in the study, the use of the advanced features required significantly more time (almost double) than when the basic features were used. The least expected finding was that the time needed for the landmark identification part of the hand-tracing analysis was significantly different from both the basic and advanced time requirements for all programs. It could reasonably have been expected that there should be no statistically significant difference between them since they are basically the same procedure; identification of points on either conventional film or on the monitor. It is difficult to determine whether or not this is a constant finding. Most studies have not differentiated timing between the different parts of the tracing procedure (neither the conventional nor the digital). As far as the basic features are concerned, it can be assumed that the observer may have been influenced by the simplicity of the whole procedure, thus deciding about landmark positioning faster. Nevertheless, this assumption cannot explain why the advanced features procedure took significantly more time than the landmark identification part of the hand-tracing procedure. Chen *et al.* (2004) concluded that the more experienced the observer in the analysis performed, the less time he/she requires for landmark identification. In the present study, the observer had more experience in hand-tracing analysis than in the digital procedure. Therefore, this could be a possible explanation. Overall, among the cephalometric analysis computer programs, there were no significant differences. Nevertheless, until further research is available on the subject, the responsibility lies with the clinician to judge to what extent he/she uses lateral cephalometric analysis software.

Conclusion

Cephalometric measurements calculated using the basic and the advanced features of five different cephalometric analysis computer programs were in good agreement with those derived when using the hand-tracing technique. The hand-tracing procedure took a significantly longer time than both the basic and the advanced features procedure, while the advanced features required almost twice the time of the basic features.

Supplementary data

Supplementary material is available at *European Journal of Orthodontics* online.

Acknowledgements

We would like to express our sincere gratitude to Professor Jan Huggare and the Orthodontic Department of the Karolinska Institutet for providing all the necessary facilities for this study. Dr Konstantinos Parikakis is accredited for sharing his experience in the field of digital cephalometric analysis, and Ms. Elisabeth Berg and Ms. Anna Wiklund for their statistical advice, modelling, and data analysis.

The following companies are acknowledged for providing their software: Image Instruments GmbH for OnyxCeph®, Elite Computer Italia for OrisCeph®, dHAL for Viewbox®, Ilexis AB for Facad®, and Rise Corporation for WinCeph®.

References

Atkinson L G, Nevill A M 1998 Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine 26: 217–238

Broadbent B H 1931 A new X-ray technique and its application to orthodontia. Angle Orthodontist 1: 93–114

Chen S K, Chen Y J, Yao C C, Chang H F 2004 Enhanced speed and precision of measurement in a computer-assisted digital cephalometric analysis system. Angle Orthodontist 74: 501–507

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

Döler W, Steinhöfel N, Jäger A 1991 Digital image processing techniques for cephalometric analysis. Computers in Biology and Medicine 21: 23–33

Enlow D H, Hans M G 1996 Cephalometry. In: Enlow D H (ed). Essentials of facial growth. W B Saunders Company, Philadelphia, pp. 259–260

Fayers P, Hays R (eds.) 2005 Assessing agreement levels. In: Quality of life in clinical trials: methods and practice Oxford University Press, Oxford, pp. 114–127.

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

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

Geelen W, Wenzel A, Gotfredsen 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

Gotfredsen E, Wenzel A, Bosch C 1997 Direct digital cephalometry. In: Lemke A V, Vannier M W, Inamura K (eds). Computer assisted radiology and surgery. Excerpta Medica, Amsterdam, pp. 54–58

Gravely J F, Benzies P M 1984 The clinical significance of tracing error in cephalometry. British Journal of Orthodontics 11: 44–48

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

Kirkwood B, Sterne J (eds.) 2001 Analysis of numerical outcome. In: Essentials of medical statistics 2nd edn. Blackwell Publishing Limited, Oxford, pp. 50–80.

Kish L (ed.) 1995 Cluster sampling and subsampling. In: Survey sampling Wiley-Interscience, New York, pp. 170–173.

Lim K F, Foong K W 1997 Phosphor-stimulated computer cephalometry: reliability of landmark identification. British Journal of Orthodontics 24: 301–308

Liu J K, Chen Y C, Cheng K S 2000 Accuracy of computerized automatic identification of cephalometric landmarks. American Journal of Orthodontics and Dentofacial Orthopedics 118: 535–540

Machin D, Campbell M J, Walters S J (eds.) 2007 Reliability and method comparison studies. In: Medical statistics: textbook for the health sciences 4th edn., John Wiley & Sons Ltd, Oxford, pp. 202–213.

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

Nimkarn Y, Miles P G 1995 Reliability of computer-generated cephalometrics. The International Journal of Adult Orthodontics and Orthognathic Surgery 10: 43–52

Rudolph D J, Sinclair P M, Coggins J M 1998 Automatic computerized radiographic identification of cephalometric landmarks. American Journal of Orthodontics and Dentofacial Orthopedics 113: 173–179

Santoro M, Jarjoura K, Cangialosi T J 2006 Accuracy of digital and analogue cephalometric measurements assessed with the sandwich technique. American Journal of Orthodontics and Dentofacial Orthopedics 129: 345–351

Seki K, Okano T 1993 Exposure reduction in cephalometry with a digital photostimuable phosphor imaging system. Dentomaxillofacial Radiology 22: 127–130

Steiner C C 1960 The use of cephalometrics as an aid to planning and assessing orthodontic treatment. American Journal of Orthodontics 46: 721–735