Journal of Orthodontics/Vol. 28/2001/221–229

An Evaluation of the Reproducibility of Landmark Identification Using Scanned Cephalometric Images

P. J. TURNER

Department of Orthodontics, Warwick Hospital, Lakin Road, Warwick, CV34 5BW, UK

S. WEERAKONE

Department of Orthodontics, Good Hope Hospital, Rectory Road, Sutton Coldfield, West Midlands, B75 7RR, UK

Abstract

Objective: *A method of cephalometric analysis is described in which cephalometric x-rays were scanned using a flat-bed scanner and transparency hood. Then the image was displayed on a computer monitor for point identification and subsequent cephalometric analysis using dedicated software. The reproducibility of point identification using this method was compared with two other, commonly used, methods.*

Material and methods: *The study material comprised 25 lateral skull x-rays taken as part of routine orthodontic assessment. Repeat cephalometric point identification was carried out on each x-ray using 3 methods:*

1. *On-screen digitization of the scanned bitmap image (Screenceph method)*
2. *Tracing followed by digitization of the identified points and*
3. *Direct digitization.*

Results: *For the 8 angular and 4 linear cephalometric measurements examined the Screenceph method compared favourably with the two conventional methods. The median difference between methods was 0·5 degrees and 0·2 mm. Using constructed Cartesian axes to examine the x, y discrepancy between repeat measurements and comparing Screenceph to tracing followed by digitization, there were significant differences in 3 instances at the 5% level and 2 instances at the 1% level. These differences represented median scores of 0·14 to 0·32 mm greater for Screenceph. Comparing Screenceph to direct digitization 15 significant differences out of the 28 measurements were noted: six at the 5% level and 9 at the 1% level. The actual difference in median scores ranged from 0·2 mm to 0·53 mm.*

Conclusion: *The results demonstrated that Screenceph is sufficiently accurate to use in a clinical setting but is not yet sufficiently exact for use in research projects owing to hardware limitations.*

Index words: Computerized Cephalometric Analysis.

Introduction

The analysis of cephalometric lateral skull X-rays is critically dependent on the accurate location of carefully defined anatomical and constructed landmarks. Errors in landmark identification, both systematic and random, are a significant source of error (Baumrind and Frantz, 1971a,b; Midtgård *et al*., 1974; Cohen, 1984; Houston *et al*., 1986), so that the methodology used to identify and record landmarks must be meticulous.

Three techniques are commonly used to identify and record landmarks in cephalometric studies. These are:

1. Overlay tracing of the lateral skull radiograph on an X-ray viewer, followed by direct measurement of cephalometric lines and angles on the tracing paper using a ruler and protractor.
2. Overlay tracing of the radiograph to identify anatomical and constructed points followed by transfer of the tracing to a digitizer linked to a computer.
3. Direct digitization of the lateral skull X-ray using a digitizer linked to a computer.

Several studies have examined the accuracy and reproducibility of landmark identification using these different methods. Direct digitization of radiographs is reported to be the most reproducible and therefore the most accurate method (Richardson, 1981; Sandler, 1988), although the difference between methods is small and statistically significant in only a few instances.

Compared to other methods, direct digitization of X-rays involves fewer stages to record landmarks, and because the angles and distances are automatically calculated using computer software there is less margin for error (Houston, 1982; Cohen, 1984). However, as Richardson pointed out, this highly accurate measurement technique is not necessarily going to reduce overall landmark error when the points being digitized are poorly defined. Furthermore, the design of a digitizer's cursor can obscure structures peripheral to the landmark of interest and the cross-hairs of the cursor can be difficult to distinguish against a dark background (Houston, 1982). This problem does not occur when digitizing a tracing.

Landmark identification using tracing paper and hand

0301-228X/01/020000+00$02.00

instruments compares favourably with the results of digitized X-rays and the results of studies using this method can be considered perfectly valid (Richardson, 1981; Sandler, 1988). Tracing alone was found to produce more reproducible results in certain circumstances: for example, the points articulare and gonion can be 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 upper incisor root (Houston, 1982).

Conversely, taking hand measurements from tracings is by far the most time consuming and tedious method, and carries the possibility of errors caused by misreading the measuring instruments and transcribing the data to computer (Sandler, 1988)

Computerized Analysis of Cephalometric Lateral Skull Radiographs

With the development of computer technology it has become possible to 'capture' a radiographic image and display this on a computer monitor. Various methods have been reported, such as mounting the image on a viewing box, and 'capturing' it using a television or video camera. More recently, digital acquisition has been achieved using a photo-stimulatable storage phosphor plate sandwiched into a standard cassette (Buckwalter and Braunstein, 1992; Cowen *et al.*, 1993; Geelan *et al.*, 1998).

The captured radiographic image is displayed on a computer monitor as an array of small points (pixels), each with a particular shade of grey: the contrast and density of this image can be altered in the same way as a television picture. The digital image is concurrently held in the computer's memory as a corresponding array of numbers (each representing a value on the grey scale) and as such can be manipulated mathematically, offering the possibility of image processing to alter its visual appearance on the monitor (Jackson *et al.*, 1985). For example, it is possible to alter the X-ray image from negative to positive, manipulate contrast and brightness and filter the image. The perceived advantage of these techniques is that they can greatly facilitate landmark identification and, therefore, overall accuracy.

As a further development a screen cursor can be guided over the image on the monitor, using either a keyboard or mouse, and using appropriate software can record anatomical landmarks. Additional software then calculates the cephalometric values. Cohen *et al*. (1984) developed a cellular logic image processing system (CLIP4) to automatically identify the cephalometric landmarks menton and sella.

Using image enhancement following video capture of the image Jackson *et al*. (1985) found comparable results with manual tracing. Cohen and Linney (1986) captured the X-ray image using a TV camera linked to a personal computer. They measured sella turcica and menton, and compared their results to those obtained using a reflex metrograph, finding comparable accuracy. Geelan *et al*. (1998) compared images captured by a storage phosphor technique, video capture, and normal tracing. They found no large differences between groups. Lowey (1993) compared methods of digitization to measuring images captured using a video camera and found small but statistically significant results between the different measurement techniques, which he felt could be ignored clinically but may be of relevance for research purposes. He concluded that an increase in resolution of his system of capture would improve the results.

Comparing Methods of Landmark Identification

The results of the investigations mentioned above are not directly comparable owing to the way in which repeat tracings of lateral skull x-rays have been examined and the different approaches used in the statistical analysis of the results.

Where the method of point identification is being compared between successive recordings of the same radiograph, it is appropriate to construct Cartesian axes around the radiograph in order to measure the horizontal and vertical distance of each point from the ordinate and abscissa, respectively. These distances can then be compared between recordings and between methods. This approach is more revealing than comparing the values of cephalometric lines and angles of successive tracings where errors in the vertical or horizontal plane (the envelope of error) can be hidden by the cephalometric analysis (Richardson, 1981).

For example, B point is more difficult to identify in the vertical plane than in the horizontal (Baumrind and Frantz, 1971a,b) and the angle SNB would fail to identify differences in the vertical position of B point between tracings. Conversely, using Cartesian axes, comparison of the vertical and horizontal distances of a point and between successive readings *would* show any differences. Having said this, in any study comparing methods of point identification, it is useful to include the differences between common cephalometric lines and angles in order to quantify the practical significance of the technique on cephalometric error.

When the position of landmarks are compared between successive recordings any difference noted (in the horizontal or vertical plane) can range from zero upwards. Negative differences are meaningless, as the researcher will never know the true position of a landmark (Houston, 1982). Plotting the differences on a graph for a series of X-rays would reveal a skewed curve, rather than a normal curve. In his study comparing methods, Sandler (1988) found one-third of the data to be skewed and all kurtosed at the 5 per cent level: nearly two-thirds were significantly kurtosed at the 1 per cent level. The application of parametric statistics to skewed data may not be appropriate and non-parametric techniques should be used (Houston, 1982).

Aim of the Project

This project describes a new method of cephalometric analysis in which conventional lateral skull radiographs were scanned into a computer using a flat bed scanner and saved as bitmap image files. The recorded X-ray image was subsequently digitized directly on-screen using a mouse pointer linked to appropriate software. This system has several potential advantages: the need for a digitizing tablet is eliminated and instead a scanner with transparency hood is used, a stage in the analysis of X-rays is eliminated and the stored image can be manipulated to assist in point identification.

Prior to the study the apparatus was checked for its accuracy by repeat recording of an image of known exact dimensions. Consistent and accurate measurements of known distances were obtained.

The new method of direct on-screen digitization of cephalometric radiographs was compared with two currently used methods of cephalometric landmark registration. In the first of these, points were identified and marked directly onto tracing paper, which was in turn digitized, and in the second method direct digitization of the X-rays was undertaken. The null hypothesis assumed there was no difference in the accuracy and reproducibility in point identification between methods.

Materials and Methods

Twenty-five lateral skull radiographs were selected consecutively from the records of patients who had attended for orthodontic assessment at Stratford upon Avon Hospital, UK. All the X-rays were taken on a Planmeca PM 2002 Proline X-ray machine using Kodak T-MAT G/RA film and a Kodak Lanex Regular rare earth intensifying screen. The X-rays were considered of good quality rather than exceptional quality and as such represented typical lateral cephalometric skull X-rays taken on a modern machine. Inclusion criteria were as follows:

1. The film was of sufficient quality to permit identification of the landmarks
2. There were no unerupted or partially erupted teeth that would have hindered landmark identification.

For each method the radiographs were recorded twice with a 1-month interval between each recording and a 4-month interval between each method. To avoid operator fatigue no more than 10 radiographs were digitized at any one time.

In the bottom left hand corner of each radiograph three pin-prick fiducial points were punched into the film outside the area of interest. The two vertical fiducial points were employed to construct Cartesian axes by which to measure the horizontal and vertical distances of the recorded cephalometric landmarks, and facilitate a comparison of methods. The abscissa was sufficiently low and the ordinate sufficiently far to the left to ensure that all measurements recorded were positive.

In addition, all three points were utilized by the ScreenCeph computer program to re-calibrate the change in image size (see later). Fourteen cephalometric points were used in the study (Figure 1), together with eight angular (Figure 2) and four linear (Figure 3) cephalometric measurements. In addition, the vertical and horizontal distance of each point, in relation to the constructed Cartesian axes were recorded.

Landmark Identification Using Computerized Recording of Scanned Images.

The lateral skull radiographs were scanned in using an Astra Umax—600P flatbed scanner fitted with a transparency hood. The optical resolution of the CCD (Charge Coupled Device) on this scanner was 300 × 600 dpi (dots per inch). Images were scanned and digitized using 'ScreenCeph'—Cephanalysis and Surgical Planning software (Version 1.4) for Windows developed by one of the authors (SW).

Images were captured at a scanning resolution of 800 × 800 dpi using a 256 gray scale palette and a magnification factor of 12 per cent. As the final image size is determined

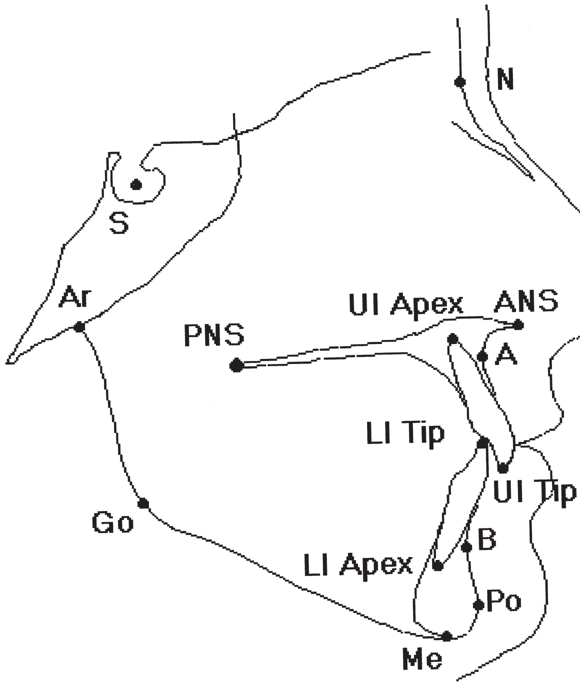

FIG. 1 Cephalometric points.

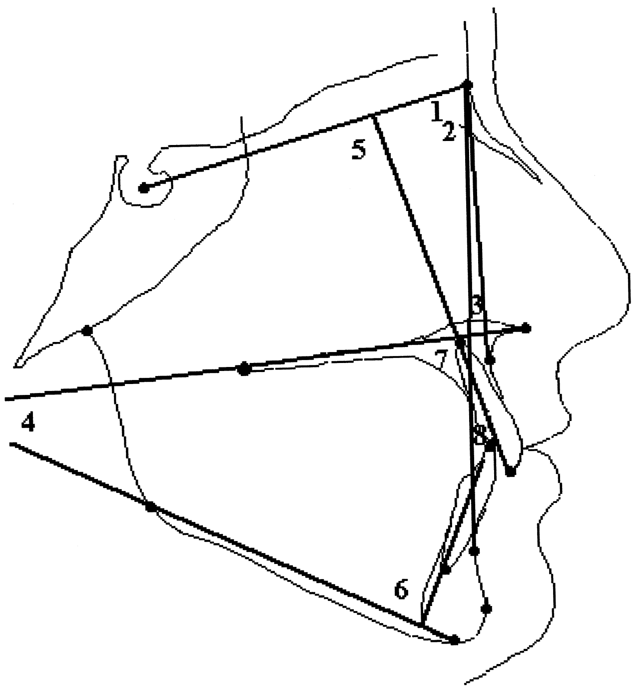

FIG. 2 Angular cephalometric measurements.

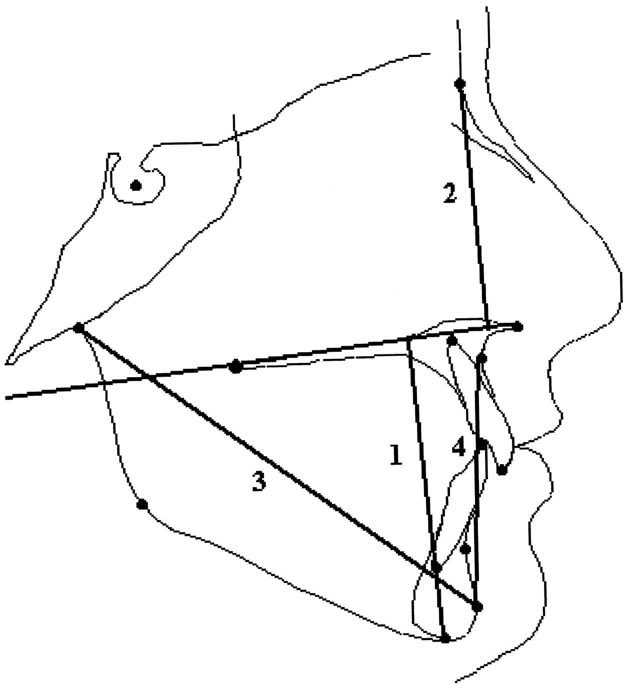

FIG. 3 Linear cephalometric measurements.

by the scanning resolution and magnification factor these two settings were kept constant for this study. An aspect ratio of 1:1 was used to eliminate image distortion along *x* and *y*-axes. It took approximately 90 seconds to scan each radiograph at this resolution. The images were stored as Bitmap image files (BMP) each requiring about 450 Kb of disk space.

Radiographic images were subsequently opened using ScreenCeph Cephanalysis program and digitized on a 17-inch colour monitor at a screen resolution of 1074 × 728 pixels. The digitizing window is approximately 9 inches wide and 8 inches high on a 17-inch monitor.

The landmarks were located using a cross-wire mouse cursor and recorded by clicking a mouse button. The *x* and *y* co-ordinates of these points were subsequently used to calculate various angular and linear measurements used in the rest of the study. Each X-ray was calibrated for the change in image size by using the three fiducial points marked on the X-ray. These points were punched at a pre-determined distance apart on the X-ray film using a metal punch constructed for the study.

Landmark Identification Using Tracing Paper Followed by Digitization

Tracing was carried out in a darkened room using an illuminated viewing screen with a black surround to reduce extraneous light. Each X-ray was firmly secured to the surface of a viewing box and a sheet of fine grade, semi-matt acetate tracing paper taped over the X-ray. Using a hard 4H pencil landmarks were identified by a single point, in a predetermined order. For bilateral structures and double images the mid-point was chosen by construction.

Following point identification the tracings were secured to a GTCO digitizing tablet linked to a PC running the GELA 1.7 digitization program and GLP1.27 interface. Each cephalometric point marked on the tracing paper was subsequently digitized, again in the same order. From these digitized points the computer software calculated an *x* and *y* value (in relation to the constructed Cartesian axes) and several commonly used cephalometric angles and lines. The resolution of the digitizer was 0·1 mm

Direct Landmark Identification Using a Digitizer

In this method the X-rays were secured to a GTCO digipad linked to a PC running Gela 1.7 digitizer program and GLP1.27 interface. Points were identified and digitized in the same predetermined order from which the computer software calculated the *x* and *y* value for each point and the cephalometric angles and lines. Again, the level of accuracy of the digitizer was 0·1 mm.

Statistical Analysis

For the statistical analysis the cephalometric data files of both digitizations for all three methods was automatically converted into ASCII format by the GLP1.27 interface software and then exported to the SurveyPlus database and statistical package.

Non-parametric statistical analysis was applied to the data: the difference between repeat measurements for each method was demonstrated using the median value and 80th percentile value as a measure of spread. Differences between methods were examined using the Wilcoxon test.

In order to relate the significance of the results to the clinical situation commonly used angular and linear cephalometric measurements were also calculated for each of the three methods examined.

Method Error

For all three methods, the 25 radiographs were measured twice. The difference between recordings were compared using a one sample *t*-test to detect systematic error and the coefficient of correlation to examine random error.

Results

Method Error

To detect for systematic error *t*-tests applied to the repeat measurement for each method (Table 1). Out of the total 120 measurements made—40 for each method—four values showed significant difference at the 5 per cent level and two at the 1 per cent level, suggesting that the material was largely unaffected by systematic error. Of note is the linear distance between the lower incisor to the APo line, which shows a significant difference ($P < 0{\cdot}01$). This is explained by the fact that the actual linear distance measured is extremely small and that any variation in the repeat measurement is more likely to show a significant difference: 0·3 mm difference for a 3-mm linear measurement is a 10 per cent difference, but for a 30-mm measurement is a 1 per cent difference.

TABLE 1 t-*Tests to detect systematic error of repeat tracings for all three methods*

Variable	Tracing		Digitizing		ScreenCeph	
	T	P	T	P	T	P
SNA	1·66	0·11	1·42	0·17	1·27	0·22
SNB	0·92	0·37	0·60	0·55	0·17	0·866
ANB	1·44	0·16	2·16	0·04*	1·27	0·214
Max–Mand	0·63	0·54	0·74	0·46	0·56	0·577
UI–SN	1·21	0·24	1·33	0·19	1·32	0·20
LI–Mand	0·47	0·64	0·36	0·72	0·93	0·36
UI–max	1·53	0·14	1·48	0·15	1·10	0·28
I–I–Angle	1·52	0·14	1·13	0·27	1·39	0·178
Max–Me–len	0·67	0·51	0·47	0·64	0·29	0·776
N–Max–len	0·27	0·79	1·10	0·28	0·25	0·802
Ar–Pog–len	0·87	0·39	1·17	0·25	0·10	0·92
LI–Apo–len	1·97	0·06	2·79	0·01**	1·81	0·083
Sella–hrz	0·49	0·63	0·81	0·42	0·41	0·683
Nasion–hrz	0·55	0·58	2·16	0·04*	0·72	0·48
Art–hrz	1·33	0·20	1·35	0·188	0·88	0·388
Gonion–hrz	0·07	0·95	0·52	0·60	0·22	0·83
Menton–hrz	0·91	0·37	1·10	0·28	0·45	0·66
Pog–hrz	0·27	0·79	0	1	1·58	0·128
A–Point–hrz	0·72	0·48	1·78	0·088	1·75	0·093
B–Point–hrz	0·35	0·73	0·36	0·723	1·58	0·128
ANS–hrz	0·47	0·64	0·71	0·485	1·26	0·218
PNS–hrz	0·57	0·57	0·22	0·825	0·85	0·403
UI–Tip–hrz	0·26	0·80	0·08	0·93	1·45	0·161
UI–Apex–hrz	1·99	0·058	1·68	0·106	1·25	0·225
LI–Tip–hrz	0·24	0·82	0·61	0·55	0·73	0·472
LI–Apex–hrz	0·34	0·73	1·13	0·261	0·91	0·373
Sella–vrt	0·77	0·45	1·35	0·19	1·66	0·109
Nasion–vrt	0·57	0·57	0·44	0·67	0·94	0·358
Art–vrt	0·27	0·79	1·43	0·16	1·77	0·09
Gonion–vrt	0·51	0·61	0·82	0·42	1·25	0·224
Menton–vrt	1·42	0·17	0·69	0·49	0·09	0·925
Pog–vrt	2·02	0·055	1·07	0·29	2·16	0·041*
A–Point–vrt	3·25	0·003*	3·28	0·008**	0·95	0·35
B–Point–vrt	0·43	0·67	0·63	0·537	0·22	0·829
ANS–vrt	0·23	0·82	0·54	0·59	0·43	0·674
PNS–vrt	0·34	0·73	0·88	0·387	0·54	0·596
UI–Tip–vrt	1·15	0·26	0·33	0·74	0·42	0·675
UI–Apex–vrt	1·45	0·16	0·35	0·73	0·11	0·912
LI–Tip–vrt	1·88	0·073	1·05	0·30	0·99	0·3344
LI–Apex–vrt	0·29	0·77	1·02	0·317	1·27	0·215

$*P < 0·05$; $**P < 0·01$.

The vertical measurement of A point compared to the Cartesian axes is possibly unreliable for the methods of tracing followed by digitization and for direct digitization ($P < 0·01$).

Correlation coefficients between repeat measurements was used to detect for random error (Table 2). All values of r were at 0·95 or above apart from ANS-horizontal (r = 0·93, tracing method), Angle ANB (r = 0·94 digitizing method), and maxillary mandibular plane angle (r = 0·93, ScreenCeph method). It can be surmised that the subject material was largely unaffected by random error for any of the three methods.

Median and Percentile Scores for Each of the Three Methods of Measurement

The median and 80th percentile scores for repeat measurement of all 25 X-rays were calculated for each of the three methods (Table 3). For all three methods the median difference between repeat measurements was 1·5 degrees or less for the majority of angular measurements. Exceptions were the ScreenCeph measurement of the upper incisor to the line S–N (1·6 degrees) and the inter-incisal angle (1·9 degrees). Using the 80th percentile as a measure of spread most angular measurements were within 2·0 degrees. A few exceeded 2·5 degrees and only one, the inter-incisal angle, exceeded 3 degrees (3·1 degrees) For repeat linear measurements the median difference for the 3 methods were 1·5 mm or less. The spread was mostly within 2 mm. These results compare favourably with those previously published (Houston, 1982; Cohen, 1984; Jackson *et al.*, 1985; Sandler, 1988).

Comparison of Methods

Cephalometric values. The Wilcoxon test was used to compare the results of the ScreenCeph method of measurement with those of tracing followed by digitization and direct

TABLE 2 *Correlation coefficients for repeat measurement of the three methods to detect random error*

Variable	Tracing		Digitizing		ScreenCeph	
	T	P	T	P	T	P
SNA	0·97	0	0·96	0	0·98	0
SNB	0·99	0	0·99	0	0·99	0
ANB	0·95	0	0·94	0	0·96	0
Max–Mand	0·98	0	0·97	0	0·93	0
UI–SN	0·98	0	0·99	0	0·98	0
LI–Mand	0·98	0	0·99	0	0·99	0
UI–max	0·97	0	0·98	0	0·98	0
I–I–Angle	0·99	0	0·99	0	0·99	0
Max–Me–len	0·99	0	0·99	0	0·97	0
N–Max–len	0·97	0	0·97	0	0·9	0
Ar–Pog–len	0·99	0	0·99	0	0·98	0
LI–Apo–len	0·99	0	0·99	0	0·99	0
Sella–hrz	0·99	0	1	0	0·99	0
Nasion–hrz	0·99	0	1	0	0·99	0
Art–hrz	0·98	0	0·99	0	0·98	0
Gonion–hrz	0·99	0	0·99	0	0·99	0
Menton–hrz	1	0	1	0	0·99	0
Pog–hrz	1	0	1	0	1	0
A–Point–hrz	0·98	0	0·98	0	0·98	0
B–Point–hrz	1	0	1	0	1	0
ANS–hrz	0·93	0	0·95	0	0·91	0
PNS–hrz	0·97	0	0·98	0	0·9	0
UI–Tip–hrz	1	0	1	0	1	0
UI–Apex–hrz	0·98	0	1	0	0·98	0
LI–Tip–hrz	1	0	1	0	1	0
LI–Apex–hrz	0·98	0	0·98	0	0·97	0
Sella–vrt	1	0	1	0	0·99	0
Nasion–vrt	0·99	0	1	0	0·98	0
Art–vrt	1	0	0·99	0	0·98	0
Gonion–vrt	1	0	0·99	0	0·99	0
Menton–vrt	1	0	1	0	1	0
Pog–vrt	0·99	0	0·99	0	0·99	0
A–Point–vrt	0·96	0	0·96	0	0·99	0
B–Point–vrt	0·99	0	0·98	0	0·99	0
ANS–vrt	0·99	0	0·99	0	0·98	0
PNS–vrt	0·99	0	0·99	0	0·98	0
UI–Tip–vrt	1	0	1	0	1	0
UI–Apex–vrt	0·98	0	0·98	0	0·98	0
LI–Tip–vrt	1	0	1	0	1	0
LI–Apex–vrt	0·99	0	0·99	0	0·99	0

digitization, respectively (Table 4). For angular and linear cephalometric values, the ScreenCeph method compares favourably with both traditional methods. The only significant differences noted were for the upper incisor to S–N plane (ScreenCeph—direct digitization) and for articulare pogonion length (ScreenCeph—tracing). If the median values between methods are compared the differences for the three methods is mostly under 0·5 degrees and 0·2 mm, and in most cases much less.

Comparison of landmark identification using Cartesian axes. Comparing values obtained for the various cephalometric points in relation to Cartesian axes some significant differences are noted (Table 4). Comparing the median scores of ScreenCeph to tracing followed by digitization 20 out of 28 ScreenCeph medians gave a higher value, although in 15 of these the difference was within 0·3 mm and the remaining four within 0·4 mm. Comparing ScreenCeph to direct digitization the former had higher median scores in 25 out of 28 measurements, which were under 0·3 mm in 18 instances and under 0·4 mm in the remainder. Looking at the measure of spread (80th percentile) 21 out of 28 values were greater for ScreenCeph over tracing, though within 0·8 mm, and 22 out of 28 values greater for ScreenCeph over digitization though within 0·9 mm.

Using the Wilcoxon test to compare ScreenCeph with direct digitization 15 values (out of 28) showed significant differences (six at the 5 per cent level and nine at the 1 per cent level). Comparing ScreenCeph to tracing followed by digitization there were five significant differences (three at the 5 per cent level and two at the 1 per cent level)

Discussion

To summarize the results of this study, for the eight angular and four linear cephalometric measurements recorded and compared between methods, the only significant differences found (at the 5 per cent level) were the upper incisor to S–N (ScreenCeph versus digitization) and the articulare–

TABLE 3 *Median and 80th percentile values in mm for repeat tracings of each method*

Variable	Tracing		Direct digitization		ScreenCeph	
	Median	80th percentile	Median	80th percentile	Median	80th percentile
SNA	0·40	0·90	0·40	0·90	0·49	0·94
SNB	0·20	0·40	0·20	0·40	0·26	0·69
ANB	0·20	0·70	0·20	0·90	0·19	0·54
Max–Mand	0·70	1·40	0·60	1·10	0·56	1·65
UI–SN	1·00	2·30	0·70	1·90	1·55	2·43
LI–Mand	0·70	2·10	0·90	1·40	1·17	1·74
UI–Max	1·30	2·90	0·80	1·70	1·45	2·39
I–I–Angle	1·40	2·40	1·50	2·20	1·81	3·08
Max–Me–len	0·40	0·70	0·30	0·80	0·34	1·16
N–Max–len	0·50	1·10	0·30	0·70	0·51	0·72
Ar–Pog–len	0·40	0·70	0·40	0·70	0·61	1·06
LI–Apo–len	0·30	0·80	0·20	0·60	0·37	0·56
Sella–hrz	0·80	0·80	0·20	0·50	0·73	1·25
Nasion–hrz	0·40	1·00	0·20	0·50	0·72	1·03
Art–hrz	0·50	0·80	0·20	0·40	0·52	0·92
Gonion–hrz	0·40	0·70	0·50	0·90	0·27	0·66
Menton–hrz	0·30	0·60	0·30	0·50	0·56	0·93
Pog–hrz	0·10	0·20	0·10	0·30	0·34	0·77
A–Point–hrz	0·50	0·90	0·30	0·60	0·41	0·96
B–Point–hrz	0·30	0·40	0·10	0·40	0·41	0·68
ANS–hrz	0·70	2·90	0·50	2·50	0·95	2·18
PNS–hrz	0·80	1·20	0·40	0·90	0·89	1·33
UI–Tip–hrz	0·30	0·50	0·10	0·30	0·44	0·70
UI–Apex–hrz	0·70	1·30	0·40	0·60	0·80	1·53
LI–Tip–hrz	0·40	0·60	0·20	0·40	0·28	0·58
LI–Apex–hrz	0·60	1·50	0·70	1·00	0·63	1·26
Sella–vrt	0·20	0·40	0·20	0·30	0·53	1·15
Nasion–vrt	0·40	0·90	0·30	0·50	0·77	1·30
Art–vrt	0·20	0·40	0·10	0·20	0·52	0·96
Gonion–vrt	0·50	1·00	0·50	1·30	0·74	1·31
Menton–vrt	0·40	0·70	0·20	0·40	0·40	0·81
Pog–vrt	0·80	1·30	0·70	0·90	0·50	1·24
A–Point–vrt	1·50	2·30	0·70	2·10	1·04	1·53
B–Point–vrt	0·80	1·40	0·50	2·20	0·92	1·44
ANS–vrt	0·90	1·10	0·30	1·00	0·52	1·67
PNS–vrt	0·50	1·00	0·30	0·70	0·66	1·30
UI–Tip–vrt	0·50	1·00	0·30	0·50	0·53	1·07
UI–Apex–vrt	0·90	1·30	0·60	2·10	0·97	2·06
LI–Tip–vrt	0·50	1·10	0·20	0·50	0·51	0·87
LI–Apex–vrt	0·70	1·20	0·60	1·40	0·98	1·45

pogonion length. The median values of the differences between methods were within 0·5 degrees and 0·2 mm. These differences compare favourably with published results and it can be surmised that for cephalometric analysis in routine clinical practice the ScreenCeph method is perfectly acceptable

Turning to the more exacting method of comparing landmark identification using Cartesian axes significant differences were noted when comparing ScreenCeph to the two other methods. Out of the 28 measurements (14 horizontal and 14 vertical) comparing ScreenCeph to tracing followed by digitization, there were significant difference in three instances at the 5 per cent level and two instances at the 1 per cent level. These differences represented median scores of 0·14–0·32 mm greater for ScreenCeph.

Comparing ScreenCeph to direct digitization 15 significant differences out of the 28 measurements were noted: six at the 5 per cent level and nine at the 1 per cent level. The actual difference in median scores ranged from 0·2 to 0·53 mm.

The median and 80th percentile values for direct digitization are smaller than those of tracing followed by digitization and the ScreenCeph method. This is also true for the measure of spread between methods. In summary, the results suggest that direct digitization of X-rays remains the most accurate method of measurement for the present time and should continue to be used for research purposes.

The difference of significance between methods for cephalometric values compared to Cartesian values was explained in the introduction, and is most likely because small errors in landmark identification either cancel each other out or are not picked up by conventional cephalometric angular and linear measurements.

Within each method and particularly for the ScreenCeph method, the pattern that emerges for accuracy of point location is typical. The largest median values and 80th percentile scores are for the horizontal and vertical position of the incisor apices, and for A-and B-point vertical position. Not surprisingly, these errors are reflected in the cephalometric values for incisor angulation.

TABLE 4 *Wilcoxon signed rank test comparing tracing and direct digitization methods to the ScreenCeph method of measurement*

Variable	Trace – ScreenCeph		Digitize–ScreenCeph	
	Z	P	Z	P
SNA	0·27	0·788	0·01	0·99
SNB	1·45	0·135	1·14	0·253
ANB	0·22	0·83	0·09	0·932
Max–Mand	0·12	0·904	1·17	0·243
UI–SN	1·09	0·275	2·22	0·026*
LI–Mand	0·40	0·687	1·30	0·192
UI–max	0·36	0·716	1·65	0·098
I–I–Angle	1·20	0·231	1·52	0·128
Max–Me–len	0·38	0·706	1·09	0·276
N–Max–len	0·28	0·778	1·13	0·258
Ar–Pog–len	2·21	0·027*	1·53	0·125
LI–Apo–len	0·36	0·721	0·57	0·567
Sella–hrz	0·74	0·46	2·09	0·037*
Nasion–hrz	0·66	0·51	2·96	0·003**
Art–hrz	0·61	0·545	2·11	0·035*
Gonion–hrz	0·06	0·954	0·87	0·381
Menton–hrz	1·88	0·059	2·95	0·003**
Pog–hrz	3·27	0·001**	3·32	0·009**
A–Point–hrz	0·12	0·904	1·05	0·294
B–Point–hrz	2·14	0·032	2·06	0·039*
ANS–hrz	0·26	0·798	0·93	0·353
PNS–hrz	0·12	0·903	2·77	0·006**
UI–Tip–hrz	2·31	0·021*	3·4	0·001**
UI–Apex–hrz	0·15	0·882	2·76	0·006**
LI–Tip–hrz	0·09	0·924	1·28	0·201
LI–Apex–hrz	0·77	0·443	0·67	0·501
Sella–vrt	3·66	0·001**	3·3	0·001**
Nasion–vrt	1·63	0·106	2·69	0·007**
Art–vrt	2·23	0·026*	2·89	0·004**
Gonion–vrt	0·82	0·412	0·03	0·978
Menton–vrt	0·54	0·587	2·39	0·017*
Pog–vrt	0·31	0·757	0·09	0·924
A–Point–vrt	1·95	0·051	0·61	0·545
B–Point–vrt	0·09	0·925	0·09	0·927
ANS–vrt	0·22	0·830	1·09	0·276
PNS–vrt	0·31	0·757	1·91	0·057
UI–Tip–vrt	0·17	0·861	2·73	0·006**
UI–Apex–vrt	0·69	0·493	1·59	0·112
LI–Tip–vrt	0·30	0·767	2·31	0·021
LI–Apex–vrt	0·83	0·404	0·52	0·599

$*P < 0{\cdot}05$; $**P < 0{\cdot}01$.

It was noted that the identification of the point sella, using Cartesian values, was greater for ScreenCeph than the other methods. The flashing cursor used in ScreenCeph changes its greyscale value to the opposite of the background image as it is moved over the screen. However, over the sella turcica the cursor fails to contrast significantly with the background and can become indistinct. This may explain the increased median and spread for this point. This source of error can be eliminated by altering the software algorithm.

The ScreenCeph method

This method offers several potential advantages over conventional cephalometric analysis, and with future improvement in image resolution is likely to become comparable to direct digitization for accuracy of point location. Flat bed scanners are already an easily obtainable and common peripheral device that can be purchased, together with a transparency hood, for a few hundred pounds. For this study the scanned image files were stored in bitmap format (450 K), but could just as easily be stored in GIF or JPEG format resulting in a smaller file size. The routine storage of X-ray images with instant access becomes a real possibility.

Resolution and Magnification

The X-rays were scanned using a UMAX flatbed 600P scanner with a CCD optical resolution of 300 × 600 dpi. When scanning images it is possible to select a specific scan density (between 1 and 4800 dpi). The images for this study were scanned at 800 dpi and displayed at a monitor resolution of 1024 × 768 dpi. When scanning the images it was difficult to improve on the automatic settings of the scanner for brightness and contrast. The ScreenCeph program was recalibrated for each image as part of the analysis, and the magnification factor between the scanned and displayed image automatically applied.

A potential limitation of using an on-screen digitizer is that the screen resolution is a determining factor and point identification could be off by as much as 50 per cent of the distance between two screen pixels. This can be calculated using the following formula:

X error = screen width/screen resolution in inches

Y error = screen height/screen resolution in inches

For a 17-inch monitor, using a screen resolution of 1024 × 768, the corresponding values are:

X error = 12/1024 = 0·0117 inch (0·2976 mm)

Y error = 9/768 = 0·0117 inch (0·2975 mm)

These are potentially large errors in comparison to the limitations of the digitizer, which has a resolution accuracy of 0·1 mm and could alone explain the difference between methods. Computer monitor resolution is continually improving with resolutions of 1280 × 1040 already available on 17-inch monitors and 1600 × 1200 on 19-inch monitors. These increased resolutions will further reduce the errors caused by pixel size. For example, on a 19-inch monitor at 1600 × 1200 resolution X error = 0·009 inch (0·2328 mm), Y error = 0·008 inch or (0·2222 mm).

This study compared methods of point identification without using any image enhancement. However, the full range of image enhancement options are built into the ScreenCeph program, such as contrast and brightness manipulation, negative images, gradient operators, and filters. These image enhancement techniques offer potential advantages in point recognition and improving accuracy (Jackson *et al.*, 1985).

References

Baumrind, S. and Frantz, R. C. (1971a)
The reliability of head film measurements. 1. Landmark Identification,
American Journal of Orthodontics, **60,** 111–127.

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

Buckwalter, K. A. and Braunstein, E. M. (1992)
Digital skeletal radiography,
American Journal of Roentgenology, **158,** 1071–1080.

Cohen, A. M. (1984)
Uncertainty in cephalometrics,
British Journal of Orthodontics, **11,** 44–48.

Cohen, A. M. and Linney, A. D. (1986)
A low cost system for computer-based cephalometric analysis,
British Journal of Orthodontics, **13,** 105–108.

Cohen, A. M., Ip, H. H-S. and Linney, A. D. (1984)
A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis,
British Journal of Orthodontics, **11,** 143–154.

Cowen, A. R., Workman, A. and Price, J. S. (1993)
Physical aspects of photostimulable phosphor computed radiography,
British Journal of Radiology, **66,** 332–345.

Geelan, W., Wenzel, A., Gotfredsen, E., Kruger, M. and 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.

Houston, W. J. B. (1982)
A comparison of the reliability of measurement of cephalometric radiographs by tracings and direct digitization,
Swedish Dental Journal Supplement, **15,** 99–103.

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

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

Lowey, M. N. (1993)
The development of a new method of cephalometric and study cast mensuration with a computer controlled, video image capture system: Part I,
British Journal of Orthodontics, 20, 203–214.

Midtgård, J., Björk, G. and Linder-Aronson, S. (1974)
Reproducibility of cephalometric landmarks and errors of measurements of cephalometric cranial distances,
Angle Orthodontist, **44,** 56–61.

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.

Appendix 1: computer programs used in the study

The ScreenCeph cephanalysis program can be obtained from Mr S. Weerakone, Department of Orthodontics, Good Hope Hospital, Rectory Road, Sutton Coldfield, West Midlands B75 7RR, UK.

The GELA cephanalysis program and GLP1.27 Gela Interface program can be obtained via the Consultants Orthodontists Group of the British Orthodontists Society and are supplied by Mr N. W. T. Harradine, Department of Orthodontics, Bristol Dental Hospital, Lower Maudlin Street, Bristol BS1 2LY, UK.

SurveyPlus Windows version 1.1 can be obtained from Dr Brian Bonner, Dental Health Services Research Unit, University of Dundee, Dundee DD1 4HN, UK.