European Journal of Orthodontics 19 (1997) 397–409

A critical appraisal of measurement of the soft tissue outline using photographs and video

Philip E. Benson and Stephen Richmond*

University Dental Hospital of Manchester and *Department of Child Dental Health, Dental School, University of Wales College of Medicine, Cardiff, UK

SUMMARY Radiographs are essential in the planning and assessment of treatment, as well as the study of growth patterns, but before subjecting a child to X-rays, the clinician must ensure that it is justifiable on clinical grounds. The use of cephalometric radiography, particularly in the young patient with a cleft of the lip and/or palate, has been questioned.

The aim of this project was to investigate the validity and reproducibility of using a photograph or video image, compared with a radiograph, for measurement of the soft tissue profile of the face. A radiographic phantom head was used, which consisted of a dry bone skull encased in a soft tissue substitute. The method involved comparing direct measurement of the head in a Reflex Metrograph with the results of computerized digitizing of a cephalometric radiograph, digitizing a video image from a computer screen and measurement of a photograph.

It was found that digitizing of the soft tissue outline of a radiographic phantom head from a video image was neither a valid, nor a reproducible method of replacing cephalometric radiographs in the measurement of the soft tissue profile of the face. A photograph may be a clinically acceptable alternative, but errors from this method are likely to be larger than those due to digitization of a radiograph.

Introduction

Radiographs are essential in the planning and assessment of treatment, as well as the study of growth patterns, but before subjecting a child to X-rays, the clinician must ensure that it is justifiable on clinical grounds. The use of cephalometric radiography, particularly in the young patient with a cleft of the lip and/or palate, has been questioned (Shaw *et al.*, 1992; Mackay *et al.*, 1994). Shaw *et al.* (1992) have suggested that examination of the soft tissue profile may indicate the extent of maxillary hypoplasia, and supply more significant information about future growth patterns, than examination of the hard tissues. The use of a photographic or video image for the evaluation of the soft tissue profile would reduce the amount of radiation to which the patient is exposed.

Video photography has many advantages over conventional film photography. Video images are produced instantly and therefore do not require costly and time-consuming development. They may be repeated immediately if they prove to be unsatisfactory. The information is stored digitally, ensuring considerable saving of space. Digital information may also be used to interact with other data, such as digitally stored radiographs, to produce a picture that combines the hard tissue image of a radiograph with the soft tissue clarity of a video (Sarver and Johnston, 1990). The information is easily recalled, is simple to edit and, unlike a photograph, the video image will not fade with time. There are many computer systems on the market that will capture and manipulate patient images.

Video images have been used in dentistry for a number of purposes. McCutcheon *et al.* (1977) described a video scanning system designed to measure lip and jaw motion. Videotape has been used for assessment of speech (Neely and Bradley, 1964) and facial appearance (Morrant, 1992). Video systems have been developed as an aid to diagnosis and treatment planning, particularly with patients undergoing orthognathic surgery (Sarver *et al.*, 1988; Takahashi *et al.*, 1989; Sarver and Johnston, 1990, 1993). The accuracy of video systems has recently been examined and questioned (Hing, 1989; Lowey, 1993a,b; Konstiantos *et al.*, 1994).

Aim of the study

The aim of this study was to find the most appropriate method of assessing soft tissue profile for use in research. To carry this out, the validity and reproducibility of linear and angular measurements of the soft tissue outline recorded by three different techniques were compared. The three procedures were a conventional cephalometric radiograph, a photograph and a video of the facial profile.

Materials and methods

Measurements were carried out on a radiographic phantom head (3M Health Care Ltd, Morley Street, Loughborough, Leicestershire, UK) which consists of a dry bone sample encased in polymethylmethacrylate. The radiographic phantom head was not anatomically ideal, therefore metal markers were placed to avoid random error due to landmark identification (Houston, 1983). The position of the landmarks and definitions used in the study are shown in Figure 1.

Four techniques were investigated:

1. Direct measurement in a Reflex Metrograph.
2. Computerized digitization of a lateral cephalometric radiograph.
3. Direct measurement of a photographic image.
4. Measurement of a captured video image.

The Reflex Metrograph technique

This is an instrument that allows three-dimensional measurements of objects without direct contact, as previously described by Butcher and Stephens (1981). The accuracy of the instrument has been reported (Speculand *et al.*, 1988). The accuracy of the Reflex Metrograph used in this study was tested using a pair of Vernier callipers (Neill Tools Ltd, Napier Street, Sheffield, S11 8HB) calibrated to an accuracy of 0.05 mm. The callipers were set to 100 mm and were placed on the Metrograph table. The calliper tips were measured horizontally in the x and y planes, and vertically in

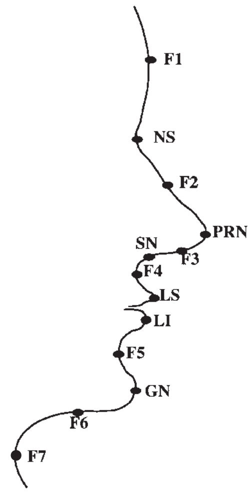

Figure 1 Landmarks and definitions used in the study:

Abbr.	Definition	Description
F1	Fiducial 1	Fiducial point at most anterior point on forehead
NS	Soft tissue nasion	The deepest point in the fronto-nasal curvature
F2	Fiducial 2	Fiducial point midway between NS and PRN
PRN	Pronasale	The most prominent point on apex of the nose
F3	Fiducial 3	Fiducial point midway between PRN and SN
SN	Subnasale	The deepest point in nasolabial curvature
F4	Fiducial point 4	Fiducial point midway between SN and LS
LS	Labrale superius	The most prominent point on the prolabium of the upper lip
LI	Labrale inferius	The most prominent point on the prolabium of the lower lip
F5	Fiducial 5	Fiducial point midway between LI and GN
GN	Soft tissue gnathion	The most antero-inferior point on the soft tissue of the chin
F6	Fiducial 6	Fiducial point midway between GN and F9
F7	Fiducial 7	Fiducial point on anterior neck

Table 1 Measurements carried out in the study.

Distances (mm)	Nose length	NS–PRN
	Fiducial 2–Fiducial 3	F2–F3
	Nose depth	SN–PRN
	Fiducial 3–Fiducial 4	F3–F4
	Upper lip length	SN–LS
	Fiducial 4–Fiducial 5	F4–F5
	Lower lip length	LI–GN
	Fiducial 5–Fiducial 6	F5–F6
	Fiducial 6–Fiducial 7	F6–F7
Angles	Nose angle	NS–PRN–SN
	Fiducial 1–Fiducial 2–Fiducial 3	F1–F2–F3
	Sagittal lip relation	F4–F1–F5
	Fiducial 3–Fiducial 4–Fiducial 5	F3–F4–F5
	Fiducial 4–Fiducial 5–Fiducial 6	F4–F5–F6
	Fiducial 5–Fiducial 6–Fiducial 7	F5–F6–F7

the z plane, 30 times on two occasions a month apart.

The radiographic phantom head was placed on the Reflex Metrograph table and the points outlined in Figure 1 were digitized. The head was removed from the table, the Metrograph reset, then the head replaced. This was repeated until 10 recordings of the points had been carried out. The co-ordinates were fed into a microcomputer. Nine linear and six angular measurements were calculated (Table 1). This procedure was repeated 1 week later.

The radiographic technique

Cephalometric radiographs of the head were taken. After each exposure, the head was removed from the cephalostat and the equipment settings adjusted. The head was repositioned, the settings readjusted and exposure carried out. This was repeated until five satisfactory radiographs of the head, displaying all the fiducial points, were obtained. The procedure was repeated 1 week later, until a total of 10 satisfactory radiographs had been produced.

The radiographs were measured using a computer-controlled digitizer connected to an IBM-compatible computer. In-house software was utilized written in Microsoft Quick Basic. The digitizer was checked for accuracy using a photographically etched graticule. Thirty measurements were made of 100 mm distances on the graticule in the x and y dimensions, on two occasions a month apart.

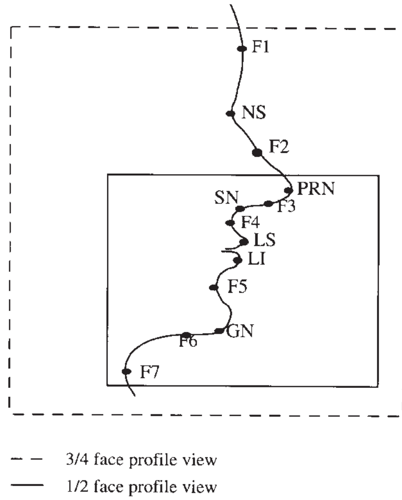

Figure 2 Landmarks used in three-quarter and half face profile views (see Figure 1 caption for definition of landmarks).

The radiographs were placed on the illuminated surface of the digitizing tablet, orientated with the Frankfort plane horizontal. The landmarks identified are defined in Table 1 and shown in Figure 1. Each landmark was digitized directly from the radiograph and double digitization was carried out to a tolerance of 0.5 mm. The co-ordinates were entered into a computer and the distances and angles shown in Table 1 were calculated. A correction for the magnification of the radiograph was made. The films were redigitized after 1 week.

The photographic technique

Colour slides of the right profile of the phantom head were taken with a 90 mm macro lens (Elicar lenses, Luxfoto Ltd, Unit 3, Grovelands Avenue, Winnersh, Berks., UK) on a 35 mm SLR camera body (Nikon UK Ltd, 380 Richmond Road, Kingston-upon-Thames, Surrey, UK) using natural light. A camera-to-object distance of 2 metres was chosen, as this closely resembled the 7 feet chosen by Sarver *et al.* (1988). To assess the accuracy of measuring those areas which would be of particular interest in the study of the soft tissues of a child with a cleft of the lip and/or palate, an image of the complete head was compared with that of a three-quarter view of the face and a half view (Figure 2). Ten exposures of each view were taken.

A calibration marker was included in each image. The marker was designed to allow for accurate calibration of images in both the *x* and the *y* plane. It was constructed of a square of black perspex, 100 mm × 100 mm. Two squares, one of 50 mm × 50 mm, the other of 20 mm × 20 mm, were etched into one corner, to permit calibration of close-up views. The calibration marker was suspended from two plumb-lines to restrict rotation of the marker and to ensure that it maintained its position at right angles to the camera.

Measurement of each colour slide was carried out by projecting the image onto a solid screen. The image was calibrated by measuring the calibration square on the image using Vernier callipers and ensuring that the full head image view was projected 1:1. The three-quarter face image was projected to 2:1 and the half face image was projected to 4:1. Measurements of linear dimensions were carried out using the Vernier callipers and angles were measured with a specially adapted protractor. The distances and angles measured in the main study are defined in Table 1. An adjustment was made for magnification of the image when the measurements were analysed. Ten images were measured in any one session to allow for operator fatigue. Repeat measurements were carried out after an interval of 1 week.

The video technique

A video image of the right profile of the radiographic phantom head was captured with the video camera. A camera-to-object distance of 2 metres was chosen, as this closely resembled the optimum chosen by Sarver *et al.* (1988). The head was removed from its stand and the camera position adjusted. The head and camera were repositioned and a new image of the right profile was captured. This was repeated until 15 sequences of the right profile had been captured. The technique was repeated 1 week later to produce 15 more sequences. The calibration marker was included in each image for accurate calibration of images in both the *x* and *y* plane.

To assess the accuracy of measuring those areas which would be of particular interest in the study of the soft tissues of a child with a cleft of the lip and/or palate, an image of the complete head was compared with that of a three-quarter view of the face and a half view (Figure 2).

An image was acquired 'live' from the videotape using an image-capture facility called 'Screen Machine' (FAST Electronic GmbH, Landsberger Str. 76, D-8000 Munchen 2, Germany). Stored images were not used as their quality was considerably less than that of an image captured live. Measurement of the video images was carried out using an image-analysis application called 'Aequitas' (Dynamic Data Links Ltd, PO Box 31,Elsworth, Cambridge, CB3 8LG, UK). The zoom facility was used to view the image at ×4 magnification. Each image was individually calibrated using the calibration marker captured with the image. The distances and angles measured are outlined in Table 1. To prevent operator fatigue, only five images were measured in one session.

Statistics

Error assessment. Analysis of Variance (ANOVA) was carried out on replicated measurements for each technique, as outlined by Buschang *et al.* (1987). ANOVA was used because it is a significance test which allows a comparison of 'within-group' and 'between-group' variance. In this study, the test was constructed to compare the variance within a session of data collection and between two sessions a week apart. The total error of each technique was calculated and the error divided into that due to the rater, the image and the random error.

Reproducibility of the technique. Reproducibility represents the nearness of repeated measurements of the same object. Paired *t*-tests were carried out on each set of repeated measurements for the radiographs and photographs. Two-sample *t*-tests were carried out on the Reflex Metrograph readings as they were not strictly paired observations. Houston (1982) and Sandler (1988) have suggested that a non-parametric test such as a Wilcoxon matched pairs signed rank test is more appropriate when investigating the differences between replicated measurements. Normality tests carried out on the data in this investigation suggested that the measurement differences were normally distributed and a *t*-test was considered appropriate.

Validity of the technique. Validity depicts the extent to which the technique produces an accurate representation of the object being measured. In this study, the Reflex Metrograph reading was taken to represent the true value or 'Gold Standard'. Two-sample *t*-tests were therefore carried out between the first set of Reflex Metrograph readings and the first set of readings from the radiographs, photographs and the video to establish the validity of the three techniques.

Bland and Altman (1986) suggested an alternative method of assessing the agreement between two methods of clinical measurement, which they called the limits of agreement. This consists of a graphical method showing the difference between the measurements on the *y* axis against the mean for the two methods on the *x* axis. This technique is useful when the true value is not known and the mean of the two techniques is a useful estimate of the true value. In this study, the Reflex Metrograph reading was used to represent the true value and the limits of agreement method was not deemed appropriate.

Table 2 Average errors for the three techniques.

	Linear (mm)	Angular (degrees)
Reflex Metrograph	0.3	0.7
Radiograph	0.3	0.5
Photograph		
Full head	0.5	1.9
3/4	0.4	1.2
1/2	0.3	0.9
Video		
Full head	1.0	3.0
3/4	0.8	1.7
1/2	0.7	1.7

Results

The means and standard deviations of the measurements carried out to calibrate the Reflex Metrograph and the digitizer were considered sufficiently accurate to indicate satisfactory calibration of both instruments.

Reflex metrograph

Error assessment (Table 2). The average error was low (0.3 mm and 0.7 degrees). The greatest proportion of this was random error (average 82 per cent of the total error).

Reproducibility. The results of the two-sample *t*-tests between the repeated Reflex Metrograph measurements are given in Table 3. There were no statistically significant differences between repeated measurements of the linear distances between fiducials. However, three out of the four distances between soft tissue cephalometric landmarks showed statistical differences between repeated measurements, one at the 1 per cent level. This may reflect difficulties with landmark identification on the soft tissue outline of the radiographic phantom head; however, closer inspection of the results showed that the mean difference of these significant results was 0.3 mm with a standard deviation of 0.2 mm, suggesting that although the result was statistically significant, this reflected the narrow range of the readings.

Two of the six angular measurements were statistically different at the 5 per cent level. Closer examination of the results again revealed that the mean difference of one of these angular measurements was only 0.1 degrees, but the standard deviation was very low and the 95 per cent confidence limits reflected a very small difference between the two readings. The Metrograph results were considered sufficiently satisfactory in terms of reproducibility to use as the Gold Standard.

Radiographs

Error assessment (Table 2). The average error was low (0.3 mm and 0.5 degrees). The largest proportion of the error was due to random error (average 57 per cent of the total error), but a larger proportion than the Metrograph readings was due to image error (average 24 per cent).

Reproducibility. The results of the paired *t*-tests between the repeated cephalometric measurements are given in Table 3. One linear distance showed statistically significant dif-

Table 3 Reproducibility of the Reflex Metrograph and cephalometric radiograph.

Measurement	Reflex metrograph				Lateral cephalograph			
	T1		T2 – T1		T1		T2 – T1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Linear (mm)								
NS–PRN	63.5	0.3	–0.3	0.3	64.5	0.7	–0.4	0.7
PRN–SN	20.9	0.3	0.3	0.3	21.4	0.3	0.2	0.2
SN–LS	25.5	0.2	–0.3	0.3	24.6	0.4	0.1	0.6
LI–GN	37.8	0.2	–0.2	0.4	38.3	0.3	0.2	0.3
Fid 2–3	41.4	0.2	0.0	0.2	41.6	0.1	0.0	0.1
Fid 3–4	8.2	0.2	–0.1	0.3	8.3	0.2	0.0	0.2
Fid 4–5	40.6	0.2	0.1	0.2	40.9	0.2	0.0	0.3
Fid 5–6	36.1	0.1	0.0	0.3	36.1	0.1	–0.3	0.7
Fid 6–7	23.6	0.1	0.0	0.3	23.9	0.1	–0.1	0.1
Angular (degrees)								
NS–PRN–SN	65.1	0.7	0.7	1.0	68.6	0.8	0.5	0.4
Fid 1–2–3	166.2	0.2	–0.2	0.4	167.0	0.2	0.1	0.3
Fid 5–1–4	8.3	0.1	0.1	0.2	8.0	0.1	0.0	0.1
Fid 3–4–5	157.2	1.6	0.1	2.2	158.4	1.0	0.2	1.3
Fid 4–5–6	152.0	0.5	0.3	0.7	151.5	0.2	0.0	0.3
Fid 5–6–7	157.2	0.6	0.0	1.1	156.9	0.3	0.6	0.3

ferences between repeated readings at the 5 per cent level and two angular measurements demonstrated a difference at the 1 per cent level. As with the Reflex Metrograph readings, the differences were very small and reflected the small standard deviations and narrow confidence limits. The radiograph readings were considered highly reproducible.

Validity. The results of the two-sample *t*-tests between the Reflex Metrograph and the lateral cephalometric measurements are presented graphically in Figure 3. The results suggest a systematic bias between the two sets of readings. Seven of the nine linear and four of the six angular measurements showed statistical differences. Six out of the seven linear measurements from the radiographic readings that were statistically different were larger than their respective Metrograph readings. The direction of bias was less obvious for the angular readings. Two radiograph readings were larger than their respective Metrograph readings and two were smaller. Most of the differences were small and,

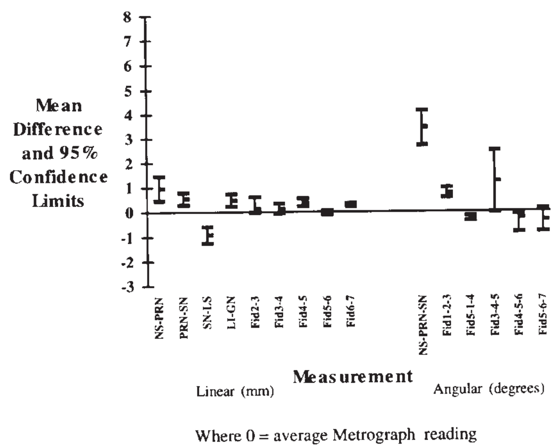

Figure 3 Validity of radiograph measurements—radiograph versus Metrograph readings (mean differences and 95 per cent confidence intervals).

although they were statistically different, would be considered clinically acceptable (1 mm or 1 degree).

Photographs

Error assessment (Table 2). The average total

Table 4 Reproducibility of the photographic and video images.

Measurement	Full head				Three-quarter head				Half head			
	T1		T2 – T1		T1		T2 – T1		T1		T2 – T1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Photographic images												
Linear (mm)												
NS–PRN	63.2	0.5	–0.9	0.6	64.4	0.4	–0.4	0.4				
PRN–SN	22.2	0.4	0.2	0.4	21.5	0.3	–0.3	0.3	20.4	0.3	0.2	0.4
SN–LS	24.2	0.8	–0.5	0.8	25.2	0.3	0.1	0.1	26.2	0.3	–0.4	0.3
LI–GN	39.0	0.6	0.4	1.0	38.1	0.5	–0.1	0.4	38.1	0.3	0.4	0.4
Fid 2–3	40.9	0.4	0.0	0.4	41.3	0.3	–0.0	0.4				
Fid 3–4	8.5	0.6	0.1	0.5	8.4	0.1	0.1	0.2	8.4	0.2	–0.1	0.1
Fid 4–5	40.1	0.5	–0.1	0.4	40.4	0.3	–0.2	0.2	40.6	0.2	0.0	0.2
Fid 5–6	34.9	0.3	–0.1	0.3	35.1	0.5	–0.1	0.2	35.4	0.2	0.1	0.2
Fid 6–7	22.8	0.2	–0.2	0.3	22.9	0.5	–0.1	0.2	22.4	0.3	–0.0	0.2
Angular (degrees)												
NS–PRN–SN	66.8	1.0	0.3	1.6	65.2	2.1	0.5	2.4				
Fid 1–2–3	166.5	3.0	–1.7	3.4	167.6	0.5	–0.2	1.2				
Fid 5–1–4	8.0	0.5	–0.4	0.7	8.3	0.3	0.1	0.6				
Fid 3–4–5	158.1	3.5	1.0	3.1	157.4	1.4	0.1	2.0	157.6	0.8	0.5	1.7
Fid 4–5–6	151.8	1.3	–0.6	1.5	152.4	1.0	–0.1	1.2	151.8	1.3	0.5	1.4
Fid 5–6–7	159.2	4.2	1.5	4.0	157.8	0.9	0.5	1.1	156.9	0.5	–0.2	0.5
Video images												
Linear (mm)												
NS–PRN	62.5	1.6	–0.3	1.6	63.8	1.2	–0.0	1.3				
PRN–SN	21.5	1.0	–0.7	1.3	21.7	1.0	–0.4	1.2	22.3	0.8	–0.3	0.9
SN–LS	24.5	1.1	–0.5	1.4	25.1	0.7	0.3	1.1	25.1	0.6	0.3	0.9
LI–GN	40.3	1.6	0.1	2.4	38.6	1.0	–0.8	1.3	38.2	0.8	–0.1	0.9
Fid 2–3	40.1	0.9	–0.5	1.2	39.9	0.6	–0.2	0.7				
Fid 3–4	10.5	0.8	–0.5	1.3	10.1	0.5	–0.1	0.6	9.7	0.6	0.2	0.6
Fid 4–5	41.0	0.9	0.5	1.3	40.5	0.7	0.1	0.7	40.0	0.4	–0.1	0.4
Fid 5–6	34.1	1.2	–0.3	1.4	34.7	0.6	–0.1	0.6	34.3	0.4	–0.0	0.5
Fid 6–7	24.5	1.1	0.2	1.2	24.6	0.7	–0.0	0.7	24.0	0.5	–0.0	0.7
Angular (degrees)												
NS–PRN–SN	69.2	4.4	2.0	3.7	65.8	2.9	0.2	2.9				
Fid 1–2–3	167.0	1.9	–0.1	3.2	168.9	1.0	0.2	1.4				
Fid 5–1–4	8.6	0.4	0.2	0.6	8.4	0.3	–0.1	0.6				
Fid 3–4–5	156.5	4.6	0.1	5.7	156.3	2.6	–1.4	3.3	157.6	3.0	–0.2	3.6
Fid 4–5–6	154.3	2.2	0.6	3.2	154.0	1.2	0.9	1.7	153.7	1.1	0.8	1.5
Fid 5–6–7	156.0	2.9	–1.6	3.4	156.9	1.9	0.1	2.8	157.1	1.0	–0.4	1.4

error was largest for the full head profile photographs (0.5 mm and 1.9 degrees) and progressively reduced for the three-quarter (0.4 mm and 1.2 degrees) and half face profile views (0.3 mm and 0.9 degrees). The full head and half face profile views showed similar proportions of random error (average 65 per cent of the total error) and rater and image error (average 35 per cent). The three-quarter view demonstrated smaller random error (average 52 per cent) and larger image error (average 42 per cent). The reason for this is not clear. The angular measurements showed larger errors than the linear measurements.

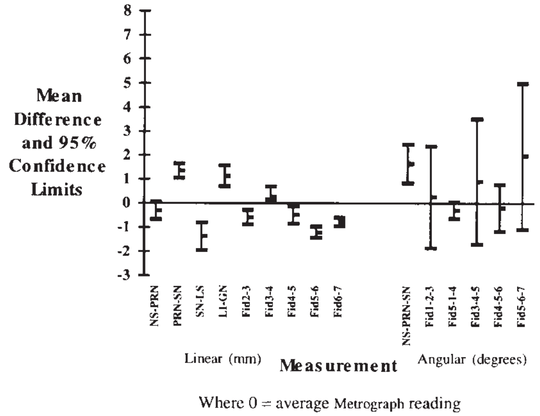

Figure 4 Validity of full head profile photographic images—full head photographic image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

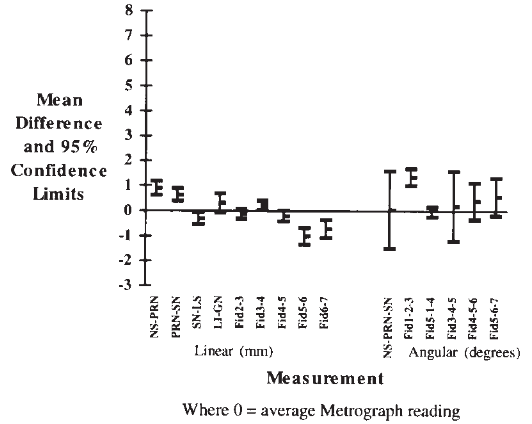

Figure 5 Validity of three-quarter face profile photographic images—full head photographic image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

Reproducibility. The results of the paired t-tests between the repeated photographic image measurements are given in Table 4. None of the angular measurements showed statistically significant differences between the first and second readings. In the case of the full head profile image, these results are probably a reflection of the large standard deviation and wide confidence limits for the differences between the first and second readings. The three-quarter and half face profile images showed much better agreement between the two sets of readings as shown by the more acceptable standard deviations and confidence limits.

The linear measurements for the photographic image repeat readings showed more statistical differences. The difference for these measurements did not exceed 0.4 mm for the three-quarter and half face profile images and, although the results were statistically significant, they were considered clinically acceptable. Overall, the reproducibility of the measurements carried out on the photographs was considered clinically acceptable for the three-quarter and half face profile images, but not for the full head profile image.

Validity. The results for the two-sample t-tests between the Reflex Metrograph and the photographic image measurements are presented graphically in Figures 4–6. The full head profile images showed many differences between the two means. Eight out of the nine linear measurements were statistically different, seven at the 1 per cent level of significance. Only one angular measurement was statistically different; however, the remaining angular measurements from the full head profile image showed large standard deviations and wide confidence limits.

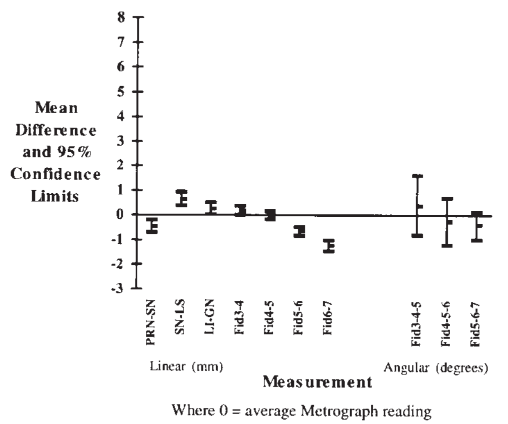

Figure 6 Validity of half face profile photographic images—full head photographic image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

The three-quarter and half face profile image measurements demonstrated much more consistent and valid results. Six of the nine linear measurements for the three-quarter view showed statistically significant differences from the

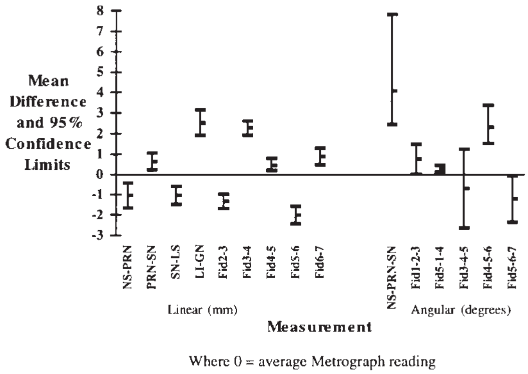

Figure 7 Validity of full head profile video images—full head video image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

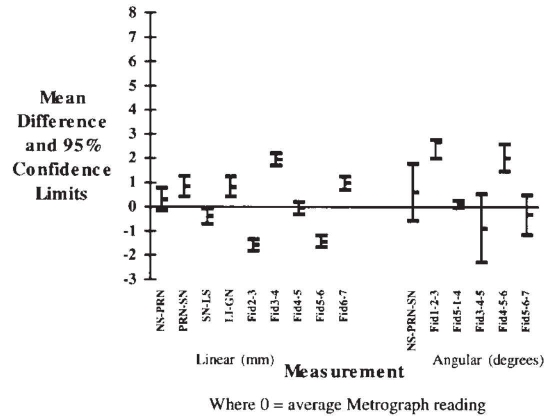

Figure 8 Validity of three-quarter face profile video images—three-quarter face video image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

Reflex Metrograph readings, but the average difference between the means for the two methods were small at 0.5 mm. Only one angular measurement showed a statistical difference and the difference between the two means was only 1.4 degrees.

The half face profile image measurements results were similar to the three-quarter face profile image results. Six out of the seven linear measurements showed statistically significant differences, but the average difference between the means for the two techniques was only 0.6 mm. The seventh measurement showed perfect agreement. None of the angular measurements showed any significant differences.

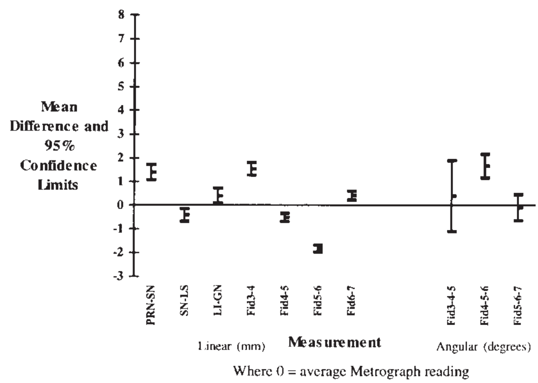

Figure 9 Validity of half face profile video images—half face video image versus Metrograph readings (mean differences and 95 per cent confidence intervals).

Video

Error assessment (Table 2). The full head video images had the largest average error at 1.0 mm and 3.0 degrees. The three-quarter and half face profile images had very similar average errors (0.8 mm compared with 0.7 mm and 1.7 degrees compared with 1.7 degrees). The proportion of errors showed that the full head image had a large average random error at 85 per cent of the total error, whereas the three-quarter and half face profile images demonstrated more error due to the image (average 19 per cent) and less to random error (average 75 per cent). As with the photographic images, the angular measurements showed larger errors than the linear measurements.

Reproducibility. The results of the paired *t*-tests between the repeated video image measurements are shown in Table 4. The full head profile image was the least reproducible statistically, with four measurements demonstrating statistically significant differences, two at the 1 per cent level. The three-quarter and half face profile views showed improved reproducibility statistically. One linear distance was statistically different for the three-quarter face profile view, this being between two soft tissue landmarks that may have been a reflection of the difficulty of landmark identification

No linear distances in the half view showed statistical differences between the first and second readings. The angular measurements for both three-quarter and half face profile images showed slightly less reproducibility, but looking at the results more closely, all the repeated angular measurements, except one, were within 2 degrees of each other. This could be considered clinically acceptable.

Validity. The results of the two-sample *t*-test between the Reflex Metrograph and the video image measurements are shown graphically in Figures 7–9. The full head profile video images showed poor validity. All the measurements, except one, showed significant differences with the two-sample *t*-test compared with the Reflex Metrograph measurements. Standard deviations were large and confidence limits were wide.

The three-quarter profile image showed reduced standard deviations and narrower confidence limits. However, seven of the nine linear measurements were statistically significantly different from the Metrograph measurements, six at the 1 per cent level of significance.

The average difference between the video and Metrograph measurement mean was 0.9 mm, which approaches the clinically acceptable limit of 1 mm. Only two angular measurements were significantly different to their Metrograph counterparts; however, the standard deviations and confidence limits remained unacceptably high and the average difference between the means of the two techniques using fiducial landmarks was 2.4 degrees, which is clinically unacceptable.

The standard deviations and confidence limits were lower for the linear measurements, even though the average difference between the means of the two techniques was the same as the three-quarter view at 0.9 mm. The angular measurements in the half face profile video image were no different to the three-quarter face profile video image. The soft tissue and fiducial landmark measurements showed no differences, therefore it appears that the soft tissue is accurately represented on the video image.

Discussion

Cephalometric radiographs provide very valuable information for the planning of treatment, as well as highlighting the effects of treatment and growth, but radiobiologists believe that there is no threshold below which X-rays are harmless (Smith, 1987). Every X-ray therefore causes finite damage to the patient and radiographic examinations must be kept to a minimum. Wall and Kendall (1983) have estimated the risks of fatal malignancy from dental radiology in Britain. By extrapolating the effects of high doses of radiation and taking into account the varying radiosensitivities of various organs, as well as age and sex differences, they predicted that the level of dental radiography present in 1981 resulted in three extra cases of fatal cancer per annum.

Sources of error

The radiographic image. Unlike a photographic image which records reflected light, the radiographic image is formed when silver halide crystals on the X-ray film are struck by X-ray photons that have passed through the subject. A radiograph has several characteristics that lead to a potential loss of accuracy of an image, particularly with respect to the different radiodensities of hard and soft tissue. Sharpness of the soft tissue outline is likely to be lost on a radiograph. In practice, these errors are usually small and clinically not significant; however, the potential for inaccuracy remains, due to the inherent way in which an X-ray image is produced. In this respect, a photographic or video image has an advantage over the radiograph, as they register reflected light, which is a more satisfactory way to record soft tissue outlines.

The results of this study suggest that measurement of a cephalometric radiograph is a valid and reproducible method of assessing the soft tissue outline. There was a small systematic bias between the radiograph and the Metrograph readings, but this was clinically insignificant and may have been due to inaccuracies in calculating the magnification of the radiograph.

The photographic image. The results of this study suggest that measurement of the soft tissue outline on a photographic image approached clinically acceptable validity and reliability, although this technique was not as accurate and reproducible as measurement of the radiograph.

Phillips *et al.* (1984) studied the errors in photocephalometry, which is a technique designed to correlate soft tissue measurements from photographs with the hard tissues of a cephalometric radiograph. They found that to produce a photographic image and a radiographic image with landmarks in the same position, certain compromises in the location of the camera were required. As a result of these compromises, there was a difference in the landmark identification errors between the radiograph and the photograph, which made a comparison between the soft tissue of the photograph and the hard tissue of the radiograph inaccurate.

There are a number of potential sources of error with a photograph. The resolution of the image will depend upon the size of the aperture and the shutter speed, when the photograph is taken, as well as the sensitivity of the film to light, which is related to the size of the grain. In addition, there will be distortion effects introduced by the lens.

A potential source of error with both photographs and radiographs, particularly when examining young children, will be distortion of the image due to movement of the patient when the photograph is taken. This is alleviated in the video technique because the camera may be kept live and a frame captured. No developing is required, therefore the image, which is produced instantly, may be immediately checked for suitability. If the frame is considered inappropriate, another may be captured and stored.

The video image. At the beginning of the experiment, it was decided to attempt to define clinically acceptable limits for validity. A general consensus was agreed that figures of 1 mm or 1 degree either side of the true figure were unlikely to be clinically significant. The mean differences for the video measurements approached these clinically acceptable boundaries, but further examination of the confidence intervals shows that although the lower limit occurred within these boundaries, the upper limits (furthest from the true figure) often fell outside. The results of this study, therefore, suggest that the video image produced with this equipment and software, even using a close-up image with small pixel size, is not a reliable and valid technique for measuring the soft tissue profile of a patient.

There are several potential sources of error.

1. *The nature of the image*. In this study, the Standard Composite Video (SCV) signal was employed, which is used for the transmission of colour television signals. The SCV signal is split into a chrominance signal, containing the values for colour hue and saturation, a luminance signal, which contains the values for the brightness, and the signals for line and frame refraction. The chrominance signal is split into two colour difference signals and this ultimately produces the signals for red, green and blue. A better system uses the RGB signal (red, green, blue colour signals) by which the base colours of the phosphors in a colour television tube are transmitted in separate signals. This allows the colour channels to be controlled individually. RGB monitors generally have a higher resolution than composite monitors.

2. *Image capture.* Errors in image capture occur in two areas. (a) Video camera: a Charge Coupled Device (CCD) camera was used in this study. CCD cameras contain elements of metal oxide silicon capacitors that build up a charge when light falls on them. The charge is transferred by the application of voltage to an amplifier that converts it to a video signal. The pixel array (a pixel is the smallest element of a computer picture making up the full image) on a typical CCD camera is 500 horizontal × 582 vertical, with a pixel size of 17 μm × 11 μm. Modern high-resolution cameras have a pixel array of 2029 horizontal × 2044 vertical with a pixel size of 9 μm square which would considerably improve the resolution. (b) Image-capture software: the image capture software used in this study, 'Screen Machine', employs a capture board with a geometric resolution of 640 × 512 pixels and colour resolution of approximately two million colours. The addition of a high-performance scientific capture board, with

resolutions up to 1024 × 1024 pixels, would improve the video image considerably, but with a substantial increase in the cost of the equipment.

3. *Image-analysis software.* There were a number of problems with the image-analysis software used in this study. These problems have been discussed with the software developers and they hope to remedy them in an updated version.

Cost–benefit analysis

The development of a new technique must not only be acceptable in terms of reproducibility and validity, but must also be justified in terms of cost. The equipment required to produce cephalometric radiographs is extremely expensive and the cost is likely to be prohibitive for research projects in developing countries.

The major running cost of the radiographic and photographic techniques is developing. This is not required with the video technique as the images can be captured directly from the camera and stored on the hard disk of a computer or on an optical disk. A 'hard copy' may be printed at a later date if required.

Another consideration in a cost–benefit analysis is the cost of storage. An optical disk will store 1 Gigabyte of information or approximately 1000 images. The same number of radiographs and photographs will use considerably more space.

It can be seen that after the initial capital outlay, the video technique is much cheaper in terms of the running costs and storage of images, compared with the radiographic and photographic technique. It would also be easier to catalogue and retrieve images from an optical disk, compared with the manual retrieval of a film or slide.

Conclusions

The following conclusions may be drawn from this study:

1. Cephalometric radiography was the most valid and reproducible technique for measuring the soft tissue outline.
2. The photographic technique showed clinically acceptable validity and reproducibility. This method would be appropriate for multicentre trials between developing and developed countries as it is non-invasive, cheap and mobile.
3. The video technique showed poor validity and reproducibility. Potential errors have been discussed and ways of improving the resolution of a video image have been suggested.

Address for correspondence

Philip Benson
Department of Clinical Dental Sciences
Liverpool University Dental Hospital and School of Dentistry
Pembroke Place
Liverpool L3 5PS
UK

Acknowledgements

The authors would like to thank Professors W. C. Shaw and K. D. O'Brien for their helpful advice, and Mr A. Jack for technical assistance.

References

Bland J M, Altman D G 1986 Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet i: 307–310

Buschang P H, Tanguay R, Demirjian A 1987 Cephalometric reliability—A full ANOVA model for the estimation of true and error variance. Angle Orthodontist 57: 168–175

Butcher G W, Stephens C D 1981 The reflex optical plotter—a preliminary report. British Dental Journal 151: 304–305

Hing N R 1989 The accuracy of computer generated prediction tracings. International Journal of Oral and Maxillofacial Surgery 18: 148–151

Houston W J B 1982 A comparison of the reliability of measurement of cephalometric radiographs by tracing and direct digitization. Swedish Dental Journal Supplement 15: 99–103

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

Konstiantos K A, O'Rielly M T, Close J 1994 The validity of the prediction of soft tissue profile changes after Le Fort I osteotomy using the dentofacial planner computer software. American Journal of Orthodontics and Dentofacial Orthopedics 105: 241–249

Lowey M N 1993a 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

Lowey M N 1993b The development of a new method of cephalometric and study cast mensuration with a computer controlled video image capture system: Part II: Study cast mensuration. British Journal of Orthodontics 20: 315–331

Mackay F, Bottomley J, Semb G, Roberts C 1994 Dentofacial form in the five-year-old child with unilateral cleft lip and palate. Cleft Palate Craniofacial Journal 31: 372–375

McCutcheon M J, Fletcher S G, Hasegawa A 1977 Video-scanning system for measurement of lip and jaw motion. Journal of the Acoustic Society of America 61: 1051–1055

Morrant D G 1992 A method of assessing surgical outcome in unilateral cleft lip and nose repair using standardized video recordings. [Thesis] Department of Oral Health and Development, University of Manchester, Manchester

Neely B J M, Bradley D P 1964 A rating scale for evaluation of video tape recorded X-ray studies. Cleft Palate Journal 1: 88–94

Phillips C, Greer J, Vig P, Matteson S 1984 Photocephalometry: Errors of projection and landmark location. American Journal of Orthodontics 86: 233–243

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

Sarver D M, Johnston M W 1990 Video imaging: Techniques for superimposition of cephalometric radiography and profile images. International Journal of Adult Orthodontics and Orthognathic Surgery 5: 241–248

Sarver D M, Johnston M W 1993 Orthognathic surgery and aesthetics: Planning treatment to achieve functional and aesthetic goals. British Journal of Orthodontics 20: 93–100

Sarver D M, Johnston M W, Matukas V J 1988 Videoimaging for planning and counselling in orthognathic surgery. Journal of Oral and Maxillofacial Surgery 46: 939–945

Shaw W C *et al*. 1992 A six centre international study of treatment outcome in patients with clefts of the lip and palate. Part 5. General discussion and conclusions. Cleft Palate Journal 29: 413–418

Smith N J D 1987 Risk assessment: The philosophy underlying radiation protection. International Dental Journal 37: 43–51.

Speculand B, Butcher G W, Stephens C D 1988 Three-dimensional measurement: the accuracy and precision of the Reflex Metrograph. British Journal of Oral and Maxillofacial Surgery 26: 265–275

Takahashi I *et al*. 1989 Application of video surgery to orthodontic diagnosis. International Journal of Adult Orthodontics and Orthognathic Surgery 4: 219–222

Wall B F, Kendall G M 1983 Collective doses and risks from dental radiology in Great Britain. British Journal of Radiology 56: 511–516