European Journal of Orthodontics 23 (2001) 569–578

The clinical significance of error measurement in the interpretation of treatment results

A. Kamoen*, L. Dermaut* and R. Verbeeck**
Departments of *Orthodontics and **Dental Material Sciences, Ghent University, Belgium

SUMMARY The aims of this investigation were to determine the errors involved in cephalometric landmark identification and to link these to the interpretation of treatment results. Fifty cephalograms were randomly selected from patient files and the following were determined. (i) Accuracy of the digitizer—single tracing digitized on five occasions on each of 10 different positions on the digitizer by one observer. (ii) Intra- and inter-observer digitizing error—35 landmarks on the same tracing (on a fixed position) digitized on five occasions by each of four judges. (iii) Intra- and inter-observer tracing error—five separate tracings of 10 different cephalograms by four judges.

There were no significant differences in the variances of the co-ordinates for each landmark between the different positions on the digitizer (mean variance *x*-axis 0.07 mm and *y*-axis 0.08 mm). (ii) One-way ANOVA showed no significant intra- or inter-observer differences in digitation. (iii) Levene's test for homogeneity of variance showed significant differences in the co-ordinates of different landmarks and between the same landmarks on different cephalograms. Two-way ANOVA showed significant differences between observers for the same landmark that were greater than the intra-observer differences.

The results indicate that tracing accuracy is a limiting factor in cephalometry. The variance of each landmark is dependent on the quality of the cephalogram. Inter-observer differences were greater than intra-observer effects and these were random, rather than systematic errors. Minimal error estimation calculations enable discrimination between treatment results and measurement errors.

Introduction

For the interpretation of results following orthodontic and orthopaedic treatment, lateral cephalograms are often used. Both in growth studies, as well as in clinical trials, the reported changes are rather small. Adequate information on the error of the method is therefore of major importance. Some authors test intra-observer error variance using the formula of Dahlberg (1940), but this is only an estimation of one investigator's own error variance. Links between this error of the method and the interpretation of treatment results are generally not mentioned. In every research project, however, it is important to reduce the error of the method as much as possible in relation to the rather small measured changes in an attempt to draw valid conclusions from these results. Only a few and rather vague guidelines have been published. According to Midtgård *et al*. (1974) and Battagel (1993), the error variance should be less than 3 per cent of the total variance. Midtgård *et al*. (1974) stated that it was quite impossible to contain the error variance to less than 10 per cent of the total variance. Baumrind and Frantz (1971b) mentioned that the observed difference as a result of therapy in an individual patient should be at least twice the standard deviation of the estimated error.

In statistical books, exact guidelines concerning links between the error of the method and the interpretation of treatment results have never been proposed (Swinscow, 1981; Houston, 1983b; Kirkwood, 1988; Riegelman and Hirsh, 1989; Bailar and Mosteller, 1992). Although very

valuable, it seems that this link is not easy to make.

Most cephalometric studies report small changes due to therapy. It is important that these are appropriately evaluated. Measurement errors must be diminished, but cannot be eliminated and, where therapeutic changes are small, measurement errors may contribute significantly to assess differences. There is a need for a methodology that evaluates the contribution of measurement error to treatment change. There is no consistency on this aspect in the literature.

Although sources of measurement errors have been previously determined, studies referring to the effect of the different composing errors tested by different observers are scarce.

Error due to repositioning of the patient in the cephalostat

Duplicate radiographs should be taken, as patient repositioning may include an important factor of error (Cooke and Wei, 1991). However, this is not ethical for clinical practice and every day research.

Error due to the radiographic procedure used

Radiographs are reported to be more reliable than digital lateral cephalograms (Jackson *et al.*, 1985; Macrì and Wenzel, 1993).

Attempts have also been made to increase the resolution of images by xeroradiography. Although this seems to give a higher precision in landmark location on cephalograms (Ruppenthal *et al.*, 1991), it is an expensive technique and not routinely used. There is also a considerably higher risk of image artefacts when one is not familiar with the technique due to insufficient operator training (Gratt *et al.*, 1985).

Error due to the measuring technique

In the past, all measurements were made by means of callipers. Computational methods that are now available are more accurate (Darryl and Frances, 1991) and are used to facilitate data processing.

Error due to landmark definition and identification

The 'definition' of landmarks is also an important factor in avoiding errors, but there can never be more exactness than agreement about the definition of the reference points (Houston *et al.*, 1986).

The error involved in landmark identification is without doubt the most important. Landmarks that are placed on anatomically formed edges or creases are easy to identify, whereas those that are placed on curves with wide radii show proportionally greater errors of measurement (Baumrind and Frantz, 1971a).

Error due to digitizer inaccuracy

The digitizer table may also include some error in the duplication of results. Positional changes of tracings on digitizers may possibly lead to some error due to less accuracy at the borders of the digitizer tablet. This source of error however, can be easily determined.

Intra- and inter-observer errors

Apart from the above factors, there is also the 'observer-factor': between two or more replicate measurements made by one or several investigators, there will always be some variance of the results due to operator inconsistency. Testing the inter-observer reliability aims to identify systemic errors of observers and to compare the measuring accuracy of one person to others. Intra-observer reliability testing reports on the measuring consistency of one person on different occasions.

A number of studies have been carried out to test the reliability of cephalometric measurements. Baumrind and Frantz (1971a) investigated the reliability of landmark identification, but no distinction was made between intra- and inter-observer errors. Hatton and Grainger (1958) also tested the reliability of cephalograms, but that study included 3-year-old children. Haynes and Chau (1993), who used Delaire's analysis, stated that there is much variance in the localization of some landmarks and proposed further research on conventional cephalometrics.

Most studies use only two investigators (e.g. Stabrun and Danielsen, 1982) and a limited number of landmarks, because often only landmarks used in one specific cephalometric analysis are considered.

In general, most investigations only concentrate on one or a few aspects, which are of importance for the reliability of results. This investigation endeavoured to extend published research by using a larger number of cephalograms, more investigators and more landmarks to make a combination of the different sources of error discussed above. Some of the landmarks used in this investigation were derived from the studies of Baumrind and Frantz (1971a) and Houston (1983a), but others have been added. This was undertaken in an attempt to test the reliability of landmark recording for the most commonly used cephalometric analyses. An attempt was also made to link the estimation of cephalometric errors to the interpretation of clinical results.

The aims of this study were to:

1. determine landmark reproducibility at 10 different positions of the digitizer tablet to establish possible inaccuracies of the digitizer;
2. determine intra-observer digitizing error of four different judges;
3. determine inter-observer digitizing error of the same four judges;
4. determine intra-observer tracing error for a sample of 50 cephalograms;
5. determine inter-observer tracing error for a sample of 10 cephalograms;
6. propose a procedure that can be followed to link the error of the method to the interpretation of treatment results from clinical trials.

Materials and methods

From the patient files of the orthodontic department at the University of Ghent, 50 lateral cephalograms were randomly selected (blind selection from more than 2000 files). All radiographs were taken with the same X-ray device (Panex EC, J. Morita Corporation, Parma, Italy). On each of these radiographs, four reference points were marked according to Baumrind and Frantz (1971a). Three were used to define the co-ordinate system. For this reason a Plexiglass plate with the four reference points, marked by means of small holes (diameter 0.1 mm), was constructed to enable transfer of the relative position of the four reference points exactly to each lateral cephalogram. Conventional cephalograms obtained during normal clinical practice were used as duplicate radiographs could not be ethically justified. Landmark definitions were taken from Riolo *et al*. (1974) to avoid dependence on any specific cephalometric analysis.

Table 1 shows the different cephalometric points used.

Table 1 The variances of the 35 landmarks along the *x*- and *y*-axes expressed in millimetres.

Landmark	SD *x*-axis (mm)	SD *y*-axis (mm)
Menton	0.5	0.2
Gnathion	0.3	0.3
Pogonion (Pg)	0.3	0.7
Point B	0.4	1.1
Infradentale	0.3	0.4
Lower incisor incisal edge	0.3	0.3
Upper incisor incisal edge	0.2	0.3
Supradentale	0.3	0.4
Point A	0.6	1.1
Anterior nasal spine	1.2	0.5
Upper incisor apex	0.7	1.1
Upper incisor lingual bony contact point	0.6	0.7
Lower incisor lingual bony contact point	0.6	0.6
Lower incisor apex	1.0	1.0
Lower molar mesial cusp tip	0.9	0.4
Upper molar mesial cusp tip	0.6	0.4
Gonion	0.5	0.5
Gonial intersection	0.5	0.5
Basion	1.5	1.2
Articulare posterior	0.6	0.6
Articulare anterior	0.8	0.7
Condylion	1.1	1.7
Sella Turcica	0.4	0.3
Ethmoid registration point	0.9	0.6
Nasion	0.4	0.7
Frontomaxillary nasal suture	1.3	1.0
Orbitale	1.4	1.3
Inferior zygoma	1.1	1.4
Pterygo-maxillary fissure, superior	1.0	1.4
Pterygo-maxillary fissure, inferior	0.6	1.6
Posterior nasal spine	1.1	0.5
Premolar mesial contact point	1.4	0.4
Porion	1.0	1.3
Anterior Downs point	0.3	0.2
Posterior Downs point	0.5	0.3

Accuracy of the digitizer tablet (aim 1)

The accuracy of the digitizer table (DIGI-PAD controller Type 5A, GTCO Corporation, USA) that was used was determined for different areas of the digitizer tablet. A tracing with the four reference points and the 35 marked landmarks was placed on the tablet. For 10 different positions equally divided over the whole surface of the digitizer tablet (Figure 1), five series of measurements of the same tracing were made each time by the same investigator to determine possible differences in the input of the landmark co-ordinates on the different locations. Variability on each position and between the different positions was calculated by superimposition of each of the five readings using the reference points.

Intra- and inter-observer digitizing errors (aims 2 and 3)

The same tracing paper was placed at a central position on the digitizer tablet and digitized five times by four different investigators (senior residents) to detect possible intra- and inter-observer variance in entering the digitized co-ordinates for data processing.

Intra-observer tracing error (aim 4)

The intra-observer tracing error was calculated for each of the 35 landmarks used in this study.

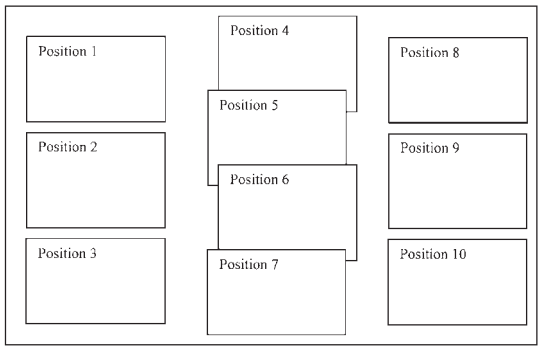

Figure 1 Different positions of the tracing paper on the digitizer tablet to check the consistency of the digitizer table.

Therefore, 50 selected radiographs were traced five times by each of the different investigators and the co-ordinates of all the reference points were placed into the digitizer at the same central position on the tablet by the same investigator.

Inter-observer tracing error (aim 5)

To determine the inter-observer tracing error, 10 lateral cephalograms were randomly selected from the original 50 radiographs and traced five times each by four investigators. The first investigator then digitized these tracings. Not only were possible differences between the readings of the different investigators measured (inter-observer tracing error), but also each investigator's intra-observer tracing variance was calculated. Thus, not the intra-observer tracing error of one person, but of all the investigators who participated in this project could be detected.

Methodology for interpreting measurement error contribution to assessment of treatment results (aim 6)

To link the error of the method to the interpretation of treatment results, an overall estimation of the tested errors was used and a proposal will be made to test the clinical significance of this error in the interpretation of treatment results. The width of 95 per cent confidence intervals of single occasion measurements of some angles and distances that make part of four commonly used cephalometric analysis (Steiner, 1953; Downs, 1956; Tweed, 1969; Ricketts, 1970) was calculated from the results obtained. Changes as a result of therapy are presented in clinical studies by means of cephalometric data before and after treatment. Some of these changes can partly be contributed to normal growth. Only few control groups are available today since ethical considerations rarely permit the gathering of material from untreated patients to allow comparison of treatment with normal growth changes. Therefore, craniofacial growth data published in atlases are most commonly used for comparison, but only average values of normal growth with their standard deviations for each age are reported. Growth

changes between two observation periods can be obtained by subtracting the average values between two age levels for most cephalometric data. A realistic estimation of the standard deviations is not an easy task, since most of these atlases are composed of material gathered from a mixed longitudinal and cross-sectional sample.

Results

Each of the landmarks was recorded into an *x*–*y*-axes co-ordinate system.

Digitizing differences between different positions on the digitizer tablet

When considering the *x* and *y* co-ordinates for the 35 landmarks, Levene's test for homogeneity of variances (Levene, 1960) showed no significant difference ($P \leq 0.05$) among the variances of the co-ordinates of each of the landmarks at the different locations of the digitizer tablet for either of the axes. This means that the variance of each point was independent of the position on the digitizer tablet.

The variance was also independent of the given point, meaning that the variance was comparable for each of the 35 landmarks. The standard deviation based on the pooled data was 0.07 mm for the *x*-axis and 0.08 mm for the *y*-axis, which was well within the given tolerance of the digitizer table (0.1 mm).

Although not important, for practical reasons further measurements were carried out on a central position of the digitizer tablet in the area of position five and six (Figure 1).

Intra- and inter-observer digitizing error

A one-way ANOVA analysis of the *x* and *y* co-ordinates for each of the landmarks, to determine the consistency in the digitizing procedure of each of the four investigators who participated in this study, showed no significant differences among the investigators. Two-factor analysis also revealed no interaction between operator and occasion (*x*-axis, $P = 0.0847$; *y*-axis, $P = 0.0771$), meaning that all the readings were independent of the given measurement. There were also no systematic errors involved.

Intra-observer tracing error

Levene's test for homogeneity of variances showed significant differences between the variances of the co-ordinates of most of the landmarks for each of the 50 cephalograms. The variances of some of the landmarks are shown in Table 1 and, as an example, Figure 2 shows the plots of some of the most commonly used cephalometric landmarks. When considering these variances, it is apparent that some landmarks have a uniform variance in all directions, but others vary more in one direction. For example, point B varies more along the *y*- than along the *x*-axis. This dependency of the variance on the direction may markedly influence the accuracy of the derived quantities, i.e. distances and angles.

The variance of each landmark is also dependent on the cephalogram, in other words some tracings can be carried out with a higher consistency. This is dependent on the quality of the cephalogram.

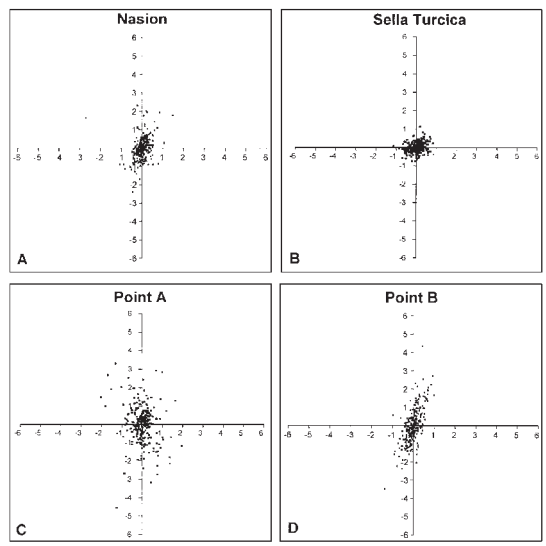

Figure 2 Plots of some of the most commonly used cephalometric landmarks: (A) nasion, (B) sella turcica, (C) Point A, (D) Point B expressed in millimetres.

Inter-observer tracing error

Ten randomly selected cephalograms, traced five times each by four independent judges, were evaluated to test inter-observer tracing error. The *x* and *y* co-ordinate of each point was different on each cephalogram, which is due to a normal biological variation of each landmark among patients. Two-way ANOVA analysis showed significant differences between the judges for the tracing procedure, both for the *x* and *y* co-ordinates of each landmark. No systematic errors were made by any single judge, but there was a considerable inter-observer tracing error involved in landmark recording, which was higher than the intra-observer tracing error and also dependent on the considered landmark.

The clinical significance of the overall error in the interpretation of treatment results

As a second part of this study, the errors of some angles and distances, which make part of four commonly used cephalometric analyses (Steiner, 1953; Downs, 1956; Tweed, 1969; Ricketts, 1970), were calculated as an example. These errors, derived from the above-mentioned errors of the separate points, are listed in Table 2. In most studies, the changes of cephalometric variables as a result of therapy are rather small. The degree of error of these variables is sometimes important. Therefore, there is a need to compare the registered changes and the error.

The values of the error of the angles and distances include a confidence interval of 95 per cent (Table 2). More important are the errors of difference between two measurements of the same variable within a period of time (e.g. before and after treatment).

An attempt to differentiate the therapeutically-induced changes from the error of the method will be proposed. First, the error values for the differences between two measurements of the same angle or distance (before and after treatment) were calculated. The error value on the other hand is also dependent on the sample size. Table 3 lists the estimations of the minimal overall error in cephalometric studies for four sample sizes (n = 20, 30, 40 and 80). The

Table 2 The error of some commonly used cephalometric variables from different cephalometric analyses.

Analysis	CI (95%)
Steiner (1953)	
SNA angle (°)	0.7
SNB angle (°)	0.5
ANB angle (°)	0.3
Upper incisor to NA (mm)	0.9
Upper incisor to NA angle (°)	1.2
Lower incisor to NB (mm)	0.7
Lower incisor to NB angle (°)	1.0
Pogonion to NB plane (mm)	1.8
Upper incisor to lower incisor (°)	1.6
Occlusal plane to S–N (°)	0.7
Go–Gn to S–N (°)	0.6
	Error
Ricketts (1970)	
Facial angle (°)	0.7
x–*y*-axes angle (°)	0.5
Facial contour (mm)	1.0
Lower incisor to A–Pg line (mm)	0.8
Upper incisor incisal edge 1A–Pg (mm)	1.1
Lower incisor to A–Po line (°)	1.1
	Error
Tweed (1969)	
FMA (Frankfort–mandibular plane angle) (°)	0.8
FMIA (Frankfort plane–mandibular incisor angle) (°)	1.4
IMPA (Incisor–mandibular plane angle) (°)	1.4
	Error
Downs (1956)	
Facial angle (°)	0.7
Angle of convexity (°)	0.8
AB plane angle (°)	0.7
Mandibular plane angle (°)	0.8
y-axis (°)	0.8
Cant of occlusal plane (°)	0.8
Interincisal angle (°)	1.6
Lower incisor to occlusal plane (°)	1.1
Lower incisor to mandibular plane (°)	1.1
Upper incisor to AP plane (mm)	0.9

therapeutic effect (TE) after clinical treatment was calculated as follows:

$$TE = \Delta_{clin} - \Delta_{growth} \quad \text{(eqn 1)}$$

where Δ_{clin} represents the observed change of the specific cephalometric variable from the start to the end of the treatment and Δ_{growth}, the change due to normal growth during that period.

Table 3 Estimation of the minimal overall error in cephalometric studies for four sample sizes.

Analysis	Δ_{errors}	Critical difference (99%)			
		n = 20	n = 30	n = 50	n = 80
Steiner (1953)					
SNA angle	0.86	0.8	0.6	0.5	0.4
SNB angle	0.61	0.6	0.4	0.3	0.3
ANB angle	0.54	0.5	0.4	0.3	0.2
Upper incisor to NA (mm)	1.28	1.2	0.9	0.7	0.5
Upper incisor to NA angle	1.78	1.6	1.3	1.0	0.7
Lower incisor to NB (mm)	0.89	0.8	0.6	0.5	0.4
Pogonion to NB plane	2.80	2.5	2.0	1.5	1.2
Upper incisor to lower incisor	2.06	1.9	1.5	1.1	0.9
Occlusal plane to S–N	0.85	0.8	0.6	0.5	0.4
Go–Gn to S–N	0.71	0.6	0.5	0.4	0.3
Ricketts (1970)					
Facial angle (difference between male/female)	1.33	1.2	0.9	0.7	0.6
x–*y*-axes angle	0.69	0.6	0.5	0.4	0.3
Facial contour	1.51	1.4	1.1	0.8	0.6
Lower incisor to A–Pg line	1.13	1.0	0.8	0.6	0.5
Upper incisor to UIE A–Pg	1.70	1.5	1.2	0.9	0.7
Lower incisor to A–Po line	1.41	1.3	1.0	0.8	0.6
Tweed (1969)					
FMA (Frankfort–mandibular plane angle)	1.35	1.2	1.0	0.7	0.6
FMAI (Frankfort plane–mandibular inc. angle)	1.82	1.6	1.3	1.0	0.8
IMPA (Incisor–mandibular plane angle)	2.51	2.3	1.8	1.3	1.0
Downs (1956)					
Facial angle	1.33	1.2	0.9	0.7	0.6
Angle of convexity	1.18	1.1	0.8	0.6	0.5
AB plane angle	0.95	0.9	0.7	0.5	0.4
Mandibular plane angle	1.35	1.2	1.0	0.7	0.6
y-axis	1.36	1.2	1.0	0.7	0.6
Cant of occlusal plane	1.27	1.2	0.9	0.7	0.5
Inter-incisal angle	2.06	1.9	1.5	1.1	0.9
Lower incisor to occlusal plane	1.35	1.2	1.0	0.7	0.6
Lower incisor to mandibular plane	1.37	1.2	1.0	0.7	0.6
Upper incisor to AP plane	1.76	1.6	1.3	0.9	0.7

According to the error propagation theory, the variance of TE is given by:

$$\sigma^2_{TE} = \sigma^2_{clin} + \sigma^2_{growth} \qquad \text{(eqn 2)}$$

as Δ_{clin} and Δ_{growth} are independent quantities. The latter also implies that σ_{clin} and σ_{growth} contain at least an evaluation error σ_o. Hence, the minimum value of

$$\sigma^2_{TE} = 2\sigma^2_o \qquad \text{(eqn 3)}$$

An adequate estimate for σ_o is given by s_o as determined in the present study, i.e. $\sigma_o \approx s_o$ for the treatment to be effective, the difference given by equation 1 must differ significantly from zero. This implies that the value obtained for TE must be at least greater than the estimated value for σ_{TE}. Taking into account the number of experiments resulting in the value of TE, this critical value is given by:

$$\Delta_{crit} = t(0.01, n\text{–}1)\,\frac{\sqrt{2}s_o}{\sqrt{n}} \qquad \text{(eqn 4)}$$

at the 99 per cent confidence level.

The 99 per cent significance level can be considered as a more stringent estimation of the error involved. In growing patients, the measured clinical difference before and after

treatment must be reduced by the amount of growth expected during that period of therapy (see, for example, growth atlases) and that difference must be compared to the critical difference values in Table 3. By doing so, it can be concluded that the final therapeutic effect might or might not exceed the error of the method.

Depending on the sample size, the observed difference in a clinical study should exceed the critical differences listed in Table 3. If the observed differences are smaller than the critical ones, they could be contributed to error rather than to the effects of therapy.

Discussion

Generally, in cephalometric studies, most authors include the error of their method. They report their overall error, but in general they do not link this error to the interpretation of treatment results. In this study, an attempt has been made to make a comprehensive estimation of cephalometric errors for some commonly used cephalometric landmarks. A methodology is suggested for evaluating the contribution of measurement error to the estimation of the therapeutic effects of treatment.

In the first part of this study, an attempt was made to analyse the different contributing sources of error involved in the analysis of lateral cephalograms. Whilst studies concerning method errors have previously been carried out, the present investigation additionally evaluated the interaction of multiple observers with the different sources of error.

Baumrind and Frantz (1971a) considered many different sources in error analysis, but by superimposition of all their readings, no distinction was made between the intra- and inter-observer tracing errors. The error for the different landmarks they measured is dependent on the error of sella and nasion, because the *x*-axis was defined as the best estimation for these two points. Midtgård *et al*. (1974) studied the reproducibility of 15 landmarks. They placed the roentgenograms on top of each other with both the contours of the anterior and posterior bases of the cranium and the contours of the ala magna of the sphenoid bone as the reference plane and, therefore, did not use an independent co-ordinate system. In the present investigation, an independent error registration was undertaken using a separate *x* and *y* co-ordinate system. Houston *et al*. (1986) stated that the greatest contribution to the error variance is from the tracings. The between radiographs variance is generally small and inconsistent, but no significant information about differences was reported in their study. Battagel (1993) suggested that the error of measurement is of importance and concluded that Dahlberg's estimation is mathematically the soundest method to evaluate measurement error. In the present study, all the contributing factors that make part of the whole measurement error were assessed to determine their individual contribution in the whole measurement error. The results show that the errors involved in the digitizing procedure are minimal in comparison with those in the tracing procedure. This is in agreement with many other studies on error determination (Baumrind and Frantz, 1971a; Midtgård *et al*., 1974; Houston, 1983a). The amount of error is different for each considered landmark: the smaller the error in the determination of the relevant landmarks, the smaller the error involved in the angles or distances of a system of analysis. The different calculated overall errors of the angles and distances are listed in Table 2, and might be used as a general estimation for other studies.

The estimations in Table 3 can be used to evaluate the treatment outcome of clinical studies. If the observed average values of a clinical study are smaller than the reported error, it cannot be concluded that the observed effect is due to therapy. By introducing the proposed system, a link can be made between the error values in cephalometrics and the small therapeutic changes, which are commonly reported in clinical studies. Since only minimal error values can be measured, prudence needs to be exercised, as it is probable that the clinical error values are larger. These minimal error values are based on the fact that, for example, a patient who is situated before treatment on the left-hand side of the normal distribution curve of the sample will also be situated on the same side after

treatment. However, it is also possible that this patient will be situated at the right hand side of the curve after treatment. Therefore, these minimal error values may need to be doubled for clinical interpretations.

To confirm the error values to the observed cephalometric changes during therapy, the results of an earlier study (Dermaut *et al*., 1992) were re-evaluated. Changes of SNA and SNB (Table 4) were found to be statistically different from normal growth within the same time period, tested by the Student's *t*-test, but no mention was made of the error of the method. By linking the error of the method to the therapeutic effects as suggested above, the observed treatment changes were still found to be larger than the minimal error values (see last column, Table 4). This means that the observed differences could be attributed to therapy. This proposed procedure to link the therapeutic effect of treatment to the error of the method is based on a minimal estimation of the error. The values for 'therapeutic effect' should, therefore, at least exceed these minimal error values.

Conclusions

1. The accuracy of the digitizer tablet was independent of the considered position with an error of less than 0.1 mm as suggested by the manufacturer.
2. The digitizing error was independent of the judges.
3. Tracing accuracy was dependent on the considered landmark and was found to be the most important source of error.
4. Inter-observer reliability revealed no systematic errors.
5. By comparing treatment outcomes of clinical studies with the minimal error values, an interpretation can be made as to whether the observed differences are due to therapy, rather than to the error of the method.

Address for correspondence

Professor L. Dermaut
Kliniek voor mond-, tand- en kaakziekten
De Pintelaan 185
B-9000 Gent
Belgium

Acknowledgements

The authors would like to thank Guy De Pauw for technical advice and Geert Dermaut for technical help.

Table 4 Comparison of the cephalometric findings reported in an earlier study (Dermaut *et al*., 1992) with the proposed procedure.

Cephalometric measurements Diff. T1–T2	Total experimental group (n = 78)		Control group Class II (n = 60)		Treatment effect compared with normal growth $\bar{x}_1 - \bar{x}_2$	t	TE compared with error (n = 80)	
	$\bar{x}_1$	SD	$\bar{x}_2$	SD				
Δ SNA	–0.8°	1.2	+0.2°	3.3	–1.0°	$P \leq 0.05$	>0.4	$P \leq 0.05$
Δ SNB	+0.9°	1.2	0.0°	3.1	+0.9°	$P \leq 0.05$	>0.3	$P \leq 0.05$

+, Increase in angulation; –, decrease in angulation; TE, therapeutic effect.
Total experimental group: 78 patients with a severe Class II malocclusion treated with a headgear activator (van Beek, 1982).
Control group: 60 non-treated Class II patients, selected at random from the cases reported by Riolo *et al*. (1984).

References

Bailar J C, Mosteller F 1992 Medical uses of statistics, 2nd edn. NEJM Books, Boston

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

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

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

Cooke M S, Wei S H Y 1991 Cephalometric errors: a comparison between repeat measurements and retaken radiographs. Australian Dental Journal 36: 38–43

Dahlberg G 1940 Statistical methods for medical and biological students. Interscience Publications, New York

Darryl N D, Frances M 1991 Reliability of cephalometric analysis using manual and interactive computer methods. British Journal of Orthodontics 18: 105–109

Dermaut L R, van den Eynde F, De Pauw G 1992 Skeletal and dento-alveolar changes as a result of headgear activator therapy related to different vertical growth patterns. European Journal of Orthodontics 14: 140–146

Downs W B 1956 Analysis of the dentofacial profile. Angle Orthodontist 26: 191–212

Gratt B M, Sickles E A, Littman R I 1985 Comparison of dental xeroradiography and conventional film techniques for the frequency and significance of image artefacts. Oral Surgery, Oral Medicine, Oral Pathology 60: 546–552

Hatton M E, Grainger R M 1958 Reliability of measurements from cephalograms at the Burlington Orthodontic Research Center. Journal of Dental Research 37: 853–859

Haynes S, Chau N Y 1993 Inter- and intra-observer identification of landmarks used in the Delaire analysis. European Journal of Orthodontics 15: 79–84

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

Houston W J B 1983b Walther's orthodontic notes, 4th edn. Wright PSG, Bristol

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

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

Kirkwood B R 1988 Essentials of medical statistics. Blackwell Scientific Publications, Oxford

Levene H 1960 Robust tests for equality of variances. Contribution to probability and statistics. Stanford University Press, Stanford

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

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

Ricketts R M 1970 Analysis—the interim. Angle Orthodontist 40: 129–137

Riegelman R K, Hirsch R P 1989 Studying a study and testing a test. How to read the medical literature. Little Brown and Company, Boston

Riolo M L, Moyers R E, McNamara Jr J A, Hunter W S 1974 An atlas of craniofacial growth, Monograph No. 2, Craniofacial Growth Series. Center for Human Growth and Development, University of Michigan, Ann Arbor

Ruppenthal T, Doll G, Sergl H G, Fricke B 1991 Vergleichende Untersuchung zur Genauigkeit der Lokalisierung Kephalometrischer Referenzpunkte bei Anwendung digitaler und konventioneller Aufnahmetechnik. Fortschritte der Kieferorthopädie 52: 289–296

Stabrun A E, Danielsen K 1982 Precision in cephalometric landmark identification. European Journal of Orthodontics 4: 185–196

Steiner C C 1953 Cephalometrics for you and me. American Journal of Orthodontics 39: 729–755

Swinscow T D V 1981 Elementaire statistiek. Stafleu's Wetenschappelijke uitgeversmaatschappij B.V. Alphen aan de Rijn, Brussel

Tweed C H 1969 The diagnostic facial triangle in the control of treatment objectives. American Journal of Orthodontics 55: 651–667

van Beek H 1982 Overjet correction by a combined headgear and activator. European Journal of Orthodontics 4: 279–290