European Journal of Orthodontics 34 (2012) 263–275
doi:10.1093/ejo/cjq144
Advance Access Publication 8 march 2011

Consistency and precision of landmark identification in three-dimensional cone beam computed tomography scans

Will Schlicher*, Ib Nielsen*, John C. Huang*, Koutaro Maki**, David C. Hatcher* and A. J. Miller*

*Division of Orthodontics, Department of Orofacial Sciences, School of Dentistry, University of California at San Francisco, USA and **Department of Orthodontics, Showa University, Tokyo, Japan

Correspondence to: Professor A. J. Miller, Division of Orthodontics, Department of Orofacial Sciences, School of Dentistry, University of California at San Francisco, San Francisco, CA 94143-0438, USA. E-mail: art.miller@ucsf.edu

SUMMARY The purpose of this study was to quantify the consistency and precision of locating three-dimensional (3D) anatomic landmarks. The hypotheses tested are that these landmarks have characteristic and variable error patterns associated with their type and location. The consistency and precision of nine orthodontists identifying 32 landmarks of 19 patients were quantified. The cone beam computed tomography (CBCT) data were acquired using a Hitachi CB MercuRay system. Prior to the study, all examiners were calibrated with respect to the definitions of the landmarks and on the use of the software program (Dolphin) for identifying the landmarks. In addition, a reference guide was provided that had the definitions and sample images of the landmarks. Data were collected in spreadsheets as *x*, *y*, and *z* co-ordinates and statistically analysed to determine the mean and standard deviation (SD). The mean location for a given landmark on a given patient served as the reference point. The mean of the distances from the reference point was used as the consistency, while the SD of this mean was used as a measure of precision. The error in the *x*, *y*, and *z* planes was calculated in order to determine the specific characteristics of each landmark.

The consistency in landmark location and precision did not differ significantly among the nine examiners. Sella turcica was the most consistently (0.50 mm) and most precisely (0.23 mm) identified anatomic landmark. The most inconsistent landmark was porion-right (2.72 mm) and the most imprecise landmark was orbitale-right (1.81 mm). Due to the lack of even distribution of the errors, careful use of these landmarks for analysis purposes is needed.

Introduction

The capability to view the craniofacial complex in three-dimensions has made cone beam computed tomography (CBCT) an invaluable tool for clinicians and researchers. While three-dimensional (3D) superimpositions, visual treatment objectives, and facial analyses may, in the future, obviate the need for precise landmark location, at the present time, they are critical to the process of quantifying the size, position, and shape of the cranium, maxilla, mandible, and dentition. This conversion from the two-dimensional (2D) image to three dimensions raises the question of how well individuals can locate many of the same conventional landmarks used in standard cephalometric analysis when given access to a 3D digital representation of their patients (Lou *et al.*, 2007; de Oliveira *et al.*, 2009).

Landmark error

Landmark identification error is an important question that numerous authors have attempted to answer since cephalometric analysis was introduced in orthodontics by Broadbent (1931). Baumrind and Frantz (1971a,b) reported the distribution of landmark selection by nine orthodontic residents. Three conclusions by these authors are important issues that need resolution in an era of 3D data: the errors in landmark identification are too great to ignore, the magnitude of error varies greatly from landmark to landmark, and the distribution of errors for most landmarks is not an error *per se* but, rather, demonstrates the difficulty in precisely locating the particular landmark. Other studies have also found that there are characteristic patterns of landmark identification distribution (Stabrun and Danielsen, 1982; Tng *et al.*, 1994; Trpkova *et al.*, 1997) and errors that cannot be ignored (Richardson, 1966; Stabrun and Danielsen, 1982; Haynes and Chau, 1993; Tng *et al.*, 1994; Trpkova *et al.*, 1997; Perillo *et al.*, 2000).

A number of studies have reported on the intra- and inter-observer variations in landmark locations. A meta-analysis of six studies on landmark identification error concluded that errors in the *x*-axis of 0.59 mm and in the *y*-axis of 0.56 mm are 'acceptable levels of accuracy' (Trpkova *et al.*, 1997). The conclusion is that significant variation exists around each of the commonly used landmarks in

cephalometrics. In order to best control the error, the characteristic patterns of error need to be better understood so that the analysis does not depend upon the specific landmark, which may introduce clinically significant errors.

Error in postero-anterior cephalometrics

The same study designs that considered error in landmark identification in lateral headfilms have been used on postero-anterior (PA) headfilms (Leonardi *et al.*, 2008). One such study that evaluated landmark error was undertaken by Major *et al.* (1994). Using 33 dry skulls and 25 patients without noticeable asymmetries, they studied 52 landmarks and found inter-examiner reliability to be between 0.31 and 4.70 mm. There were four examiners, and the skull and patient data were never combined into one data set. In this case, 0.31 mm was the error in the transverse/horizontal axis for the landmark IPL (i.e. the crest of the alveolus between the mandibular central incisors) while the 4.70 mm of error was in the vertical axis for the landmark CS (i.e. the most superior aspect of the condyle). Although they reported the data for the x and y co-ordinates, no attempt was made to quantify the 2D distance using the Pythagorean theorem. If this had been done for IPL and CS, then the true error in landmark identification among all examiners was actually 0.83 and 4.98 mm, respectively. A final finding by Major *et al.* (1994) was that certain landmarks possess significantly different amounts of error when skull and patient landmark errors are compared. Studies that look only at skulls do not represent the error that could be expected clinically since the soft tissue affects the contrast of the image.

CBCT craniofacial analysis

A basic way of approaching a 3D analysis is to propose a reference structure as the means of comparing distances of landmarks from this reference point among patients. Lagravere and Major (2005) and Lagravere *et al.* (2008) proposed a point midway between the foramen spinosum on the floor of the cranium. This point was found to be reliably located. An initial attempt by Park *et al.* (2006) at skeletal analysis using 3D analysis was based upon relationships between various planes. While Park *et al.* (2006) also suggested a number of simple linear and angular measurements, without established 3D norms, these values are of limited diagnostic value. Cephalometrics has progressed since 1931, and CBCT provides information that requires studies to test the various hypotheses (Kumar *et al.*, 2007, 2008; Periago *et al.*, 2008; Stratemann *et al.*, 2008; Moerenhout *et al.*, 2009). The primary aim of the present research was to quantify the consistency and precision of locating 3D anatomic landmarks using CBCT-generated volumetric images and slices.

Subjects and methods

The Committee on Human Research, as the institutional review board, approved the research project (H8933071801).

Patients

Nineteen patients (six males and 13 females, range 18–35 years of age, mean and 1 SD of 21.1 ± 7.9 years of age), who presented to the Orthodontic Clinic between July 2004 and July 2007 for treatment were randomly selected. All patients accepted for treatment were screened with a CBCT scan, which replaced the traditional 2D lateral headfilm. The same settings and techniques were used when taking all 19 scans. No patients with anatomic anomalies or obvious asymmetries were included since this would alter or obscure certain landmarks.

CBCT/landmark identification software

A Hitachi CB MercuRay (Hitachi Medico Technology, Tokyo, Japan) operated by the same certified technician was used for all patient scans. With the patient sitting upright, a rotating source/detector gantry captured a volumetric image of the patient's head. A 10 second scan acquired 512 images in a 12 inch diameter spherical volume with 0.2–0.376 mm^3 voxels in high-resolution mode with 12 bits/voxel (2^{12} = 4096 shades of grey). The version of the system used in this study was a scalable 12 inch charged-coupled device detector. The X-ray source was generated with a wattage of 120 kVp and a current of 15 mA for each 10 second scan.

Digital imaging and communications in medicine data sets were then imported into Dolphin Imaging 10.1 (Dolphin, Chatsworth, California, USA) in order to identify anatomic landmarks using the 3D data. No unique personal identifiers were used as part of the 3D file to prevent any bias and to maintain the privacy of the patients. Each patient was assigned a random identifier so that examiners could retrieve the file when they identified the defined landmarks. Once the data sets were imported into Dolphin Imaging, the head positions were standardized. Creating uniformity was expected to allow for easier landmark identification. Because the windows that display the three planes (sagittal, vertical, and transverse) depend upon the operator's defined orientation, this variable was controlled by orientating all patients the same way. A line connecting sella and nasion corrected the pitch (x-axis), a line parallel to the inferior surface of the sphenoid bone at the antero-posterior position of sella corrected roll (z-axis), and a line parallel to the anterior border of the glenoid fossa at the level of sella and the inferior border of the sphenoid bone corrected the yaw (y-axis).

The nine residents identified the same 32 landmarks in 19 different patient files in Dolphin Imaging. The Dolphin Imaging 3D module allowed placement of landmarks using any one of the four available views (volumetric, sagittal,

transverse, and vertical; Figure 1). The software continually updated the location of the landmark in the other three views as it was moved in any single view. This feature allowed refined placement of landmarks so that each dimension was accounted for prior to moving on to the next landmark (Figures 1 and 2). The views that displayed the three planes each showed a 0.3 mm thick slice, which offers a significant advantage over 2D films that superimpose the entire craniofacial complex. The examiners all began with the same patient and followed the same order of patients. Once they completed locating the 32 landmarks, the co-ordinate data were copied into Excel (Microsoft, Redmond, Washington, USA) spreadsheets and saved using a specific file name: all files included the patient number and examiner's name.

Examiners

The nine examiners were second- and third-year orthodontic residents. All examiners attended two 1 hour training sessions. In the two sessions, they were taught how to navigate through the Dolphin Imaging 3D module and how to place the landmarks. They were also given the opportunity to locate each of the landmarks and ask questions regarding their interpretation of the definitions provided. By the end of the second session, all examiners had found each landmark at least once. They were then given 6 months to locate the landmarks on the 19 patient files. At all times, the examiners had access to a manual with directions for opening patient images, landmark definitions, and captured screenshots of the landmarks (Schlicher, 2008).

Landmarks

Definitions for 3D anatomic landmarks are few since orthodontists have classically utilized the 2D headfilm. Many of the definitions were first defined by Solow (1966). In all cases, the definitions were checked with the 3D images, and, if needed, one author (WS) added the definition describing the transverse dimension. The final definitions were edited and confirmed with another investigator (IN; Tables 1 and 2).

Statistics

In order to assess landmark identification consistency and precision, there must be a point about which all others are compared. Because there was no gold standard for identifying the absolute position of each landmark in live patients (i.e. directly identifying a site on the bone), the average co-ordinates of all the examiners for each landmark served as the centroid for each particular landmark. From this centroid, the distribution of examiners' landmarks was assessed. Firstly, the distances from the centroid were calculated using the Pythagorean theorem that included all three dimensions in the x, y, and z planes. For a given patient

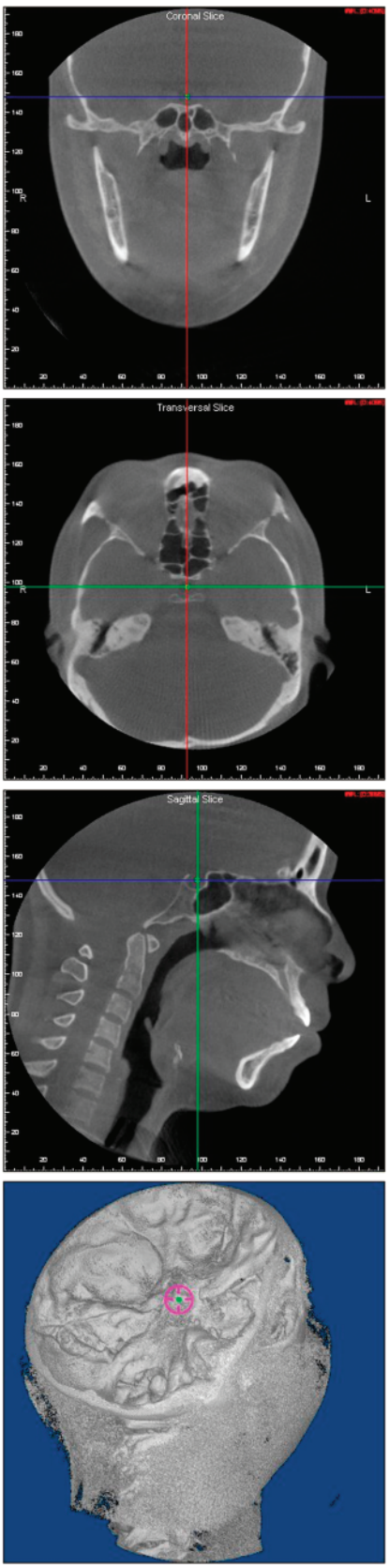

Figure 1 Example of the views seen when identifying sella using the Dolphin Imaging program.

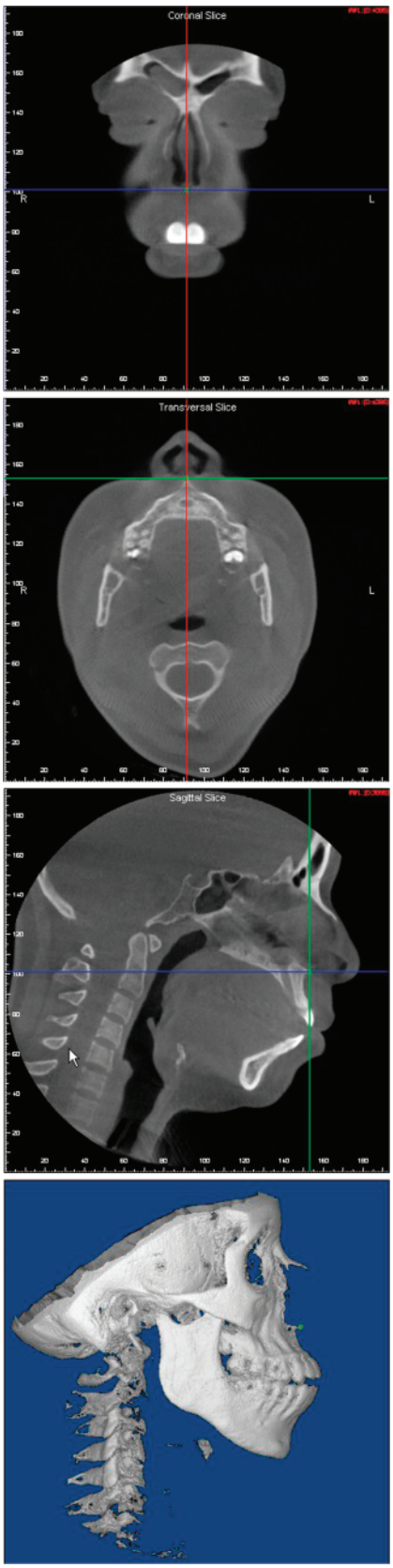

Figure 2 Example of the views seen when identifying anterior nasal spine using the Dolphin Imaging program.

and landmark, a mean distance from the centroid for that specific landmark was calculated to assess how well examiners agreed upon the location of that landmark. If a gold standard were available as the true location of the landmarks, then this study would be about accuracy and precision. Due to the impossibility of a gold standard for an anatomical location on a live patient (Stratemann *et al.*, 2008; Brown *et al.*, 2009), this study actually quantified the consistency and precision of landmark identification by multiple examiners. The consistency for a landmark was the mean of the measurements (total of all measurements/number of observations) of how far the landmarks were from the centroid by all the examiners. The SD of these distances was used to assess the precision of landmark identification. In essence, the inter-observer error of the method of using CBCT planes and volumetric images was measured simultaneously to define a landmark position. The study used the mean and SD to define the central tendency of this data.

For each landmark, the variation in each plane (i.e. *x*, *y*, and *z*) was calculated in a similar manner. For each dimension, the centroid co-ordinate value was compared with the values of the examiners. In order to assess changes in consistency and precision throughout the study, trend lines with accompanying correlation coefficients were drawn through scatterplots. These scatterplots had the patient number on the *x*-axis and the amount of consistency or precision on the *y*-axis.

Results

Individual inter-examiner performance

Prior to calculating the consistency and precision of landmark identification, an examination of the raw data and inter-examiner performance were evaluated. Before pooling all the data for analysis of individual landmark consistency and precision, it was important to know if the examiners performed equally. To do this, their consistencies for each patient were compared. Once the mean location was calculated in the *x*, *y*, and *z* planes for all examiners, the overall distance from that mean could be determined for each examiner. The mean 'distance from mean' across all 32 landmarks for a given patient represented the examiner's performance for that patient. These data, plotted to visualize each examiner's performance throughout the study across 19 patients, showed that the residents' performances were similar, allowing pooling of the data for subsequent analysis (Figure 3). The examiners differed for three of the 19 patients primarily due to some of the examiners having a few landmarks that were markedly different from the mean for all the examiners for that patient.

In evaluating the raw data, the decision was made to remove from the data set outliers and any identified landmarks that were technical errors (i.e. choosing the next

landmark instead of the correct one, hitting the cursor incorrectly), which were always greater than 2 SDs from the mean landmark location. The rationale for this decision was that large outliers were not due to a misunderstanding of the landmark definition or an inability to locate the landmark within the 3D image but rather a technical error in the use of the Dolphin Imaging 3D module. Justification for this decision came not only from looking at the millimetric deviation from the mean location but also from keeping track of those errors that were greater than 2 SDs from the mean. This decision allowed the study of the precision and consistency of defining the landmarks based on the examiners trying to define the actual position than on a study determining how the examiners used the Dolphin program.

Table 1 Order of landmark identification. L, left; R, right.

Order of landmark identification	Landmark
1	Sella
2	Basion
3	Articulare
4	Nasion
5	ANS
6	Point A
7	Point B
8	Pogonion
9	Gnathion
10	Menton
11	L-glenoid fossa
12	L-condylion
13	L-porion
14	L-gonion
15	L-orbitale
16	L-infraorbital foramen
17	R-infraorbital foramen
18	L-maxillary central incisor crown tip
19	L-maxillary central incisor root apex
20	R-maxillary central incisor crown tip
21	R-maxillary central incisor root apex
22	L-mandibular central incisor crown tip
23	L-mandibular central incisor root apex
24	R-mandibular central incisor crown tip
25	R-mandibular central incisor root apex
26	R-orbitale
27	R-gonion
28	R-porion
29	R-condylion
30	R-glenoid fossa
31	R-maxillary cant point
32	L-maxillary cant point

How identifying landmarks changed as the study progressed was also evaluated. If gross errors greater than 2 SDs were the result of lapses in utilization of all the benefits offered by Dolphin Imaging, then tracking the number of these errors as the study progressed would give an idea of how well the residents began to understand and utilize the program's features. While the trend in Figure 4 shows a slight negative correlation and, thus, suggests an improvement in the use of the program, the Pearson's correlation coefficient was only $r = 0.068$ which is weak, suggesting that the errors did not change with use of the program. The outliers of patients 1 and 12 helped to make the trend with a negative slope.

Landmark consistency and precision

The average consistency across all landmarks (32), patients (19), and examiners (nine) was 1.64 mm (Figure 5). Thus, for 5472 data points, the examiners were, on average, only 1.64 mm from the mean location of a given landmark. The

Table 2 Three-dimensional landmark definitions.

Landmark	Definition
Sella	The geometric centre of the sella turcica
Nasion	The most anterior and median point along the frontonasal suture (Solow, 1966)
Articulare	A constructed point at the predicted intersection of the inferior surface of the sphenoid bone (cranial base) with the posterior border of the mandible when the patient is viewed from the side
Basion	The most inferior point on the anterior border of the foramen magnum; the most postero-inferior point on the clivus (Solow, 1966)
ANS	The apex of anterior nasal spine (Solow, 1966)
Point A	A midline point in the deepest concavity along the anterior contour of the maxilla
Point B	The deepest point on the anterior contour of the lower alveolar arch (Solow, 1966)
Glenoid fossa (right and left)	The most superior point of the glenoid fossa
Condylion (right and left)	The most supero-posterior point on the condylar head
Porion (right and left)	The most superior point of the external acoustic meatus located laterally at the point when the meatus is entirely encircled in bone
Gonion (right and left)	A point at the intersection of the mandibular plane and the ramus plane; a point on the bony contour of the gonial angle determined by bisecting the tangent angle
Orbitale (right and left)	The deepest point of the infraorbital margin
Infraorbital foramen (right and left)	The anterior opening of the foramen
Pogonion	The most prominent point of the chin
Menton	The lowest point on the symphysis
Gnathion	A point at the intersection of the facial and mandibular planes
Maxillary cant point (right and left)	The most superior and lateral point along the oral surface of the palatine bones at the antero-posterior position of the maxillary first molars

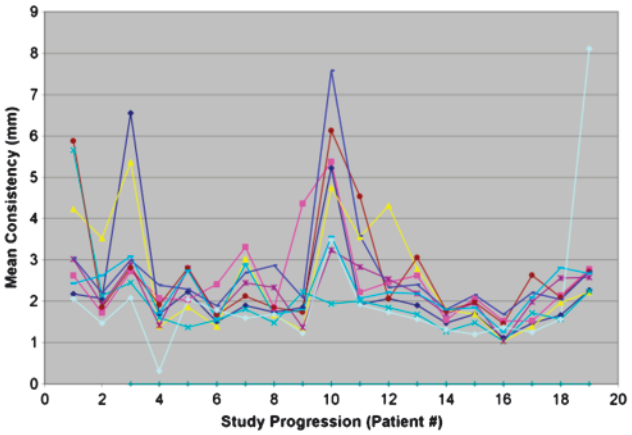

Figure 3 Consistency of each of the nine examiners for each patient as defined by the mean of the linear distance by each examiner from each of the centroids for all 32 landmarks on that patient.

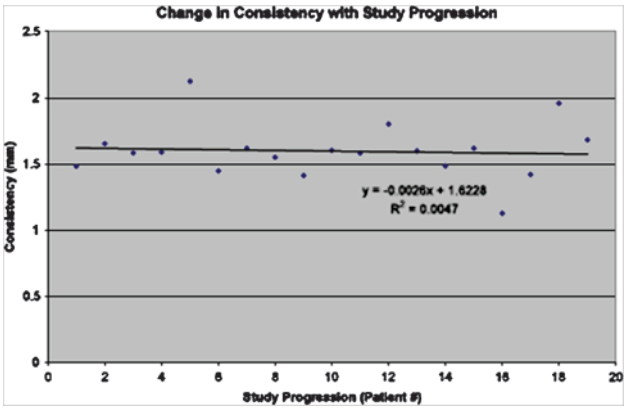

Figure 4 Correlating consistency with landmark identification experience as examiners progressed through 19 patients.

average precision (SD of the deviation) about these 5472 data points was 0.87 mm. Therefore, 66 per cent (1 SD) of all landmarks fell within 0.87 mm of one another. When consistency was tracked throughout the study, there was no significant improvement in how well landmarks were identified. The mean consistency and precision were oversimplifications of the data. The differences between the landmarks began to reveal potential explanations for the variations. In general, midline structures and landmarks formed by acute angles were more consistently identified than bilateral structures and landmarks along broad curves (Figure 5).

Individual landmark

Since landmark identification in this study was in three dimensions, consistency can be described not only by the total distance from the mean but also by the *x*, *y*, and *z* distances from the mean landmark. By breaking down the overall consistency into its three component axes, a more thorough evaluation could be carried out (Table 3). Ranking the landmarks by how consistent they were identified overall as well as by their component axis provided an analysis of the effect of three dimensions. Generally speaking, a landmark that was ranked highly among one axis was also ranked highly among the others. There were a number of landmarks that did not follow this trend. For example, left glenoid fossa was ranked 27th (*x*-axis), 7th (*y*-axis), and 21st (*z*-axis). This type of distribution suggests that the definition and/or the anatomy are partially at fault for the poor ranking in two of the three axes. Table 3 shows that midline structures and landmarks formed by acute angles were more easily identified.

Sella

Sella was the most consistently identified landmark with an overall consistency of 0.50 mm. Figure 6a follows the performance of the examiners throughout the study and allows closer examination of which dimension(s) most affected any inconsistencies. In the case of sella, while there was slight improvement in the overall consistency (light trend line: 0.0092 mm per patient), this was due to the more significant improvement in the consistency in the *x*-axis (dark trend line: 0.196 mm per patient; $r = 0.60$). This improvement is not likely to be clinically significant, but does show that, as a group, the examiners were better able to centre the landmark transversely in the sella tursica at the end of the study.

Basion

The overall consistency for basion was 0.85 mm (rank = 4th) and similar to sella. While the improvement was not clinically significant, the reason for the alteration in overall consistency was due to a change in the *x*-axis. There was not a strong correlation coefficient for any of the trend lines.

Articulare

Articulare is an unusual 3D landmark due to its previous definition in lateral cephalometrics. The steps required to locate the new definition required utilizing all four visualization windows in the computer display. In most instances, the landmark was first located using a lateral view of the reconstructed volumetric image and placing the landmark at the intersection of the posterior border of the mandible and the cranial base. The landmark was then dragged to the midline so that there would be consistency in all three dimensions. The landmark's overall consistency suffered (1.81 mm; rank = 20th) as a result of the multistep process. When the consistency was broken down into its component axes, there were discrepancies in the ranks ($x = 13$th, $y = 25$th, and $z = 23$rd). The location of the landmark in the *y*- and *z*-axes required more manipulation of the images than the *x*-axis, which only

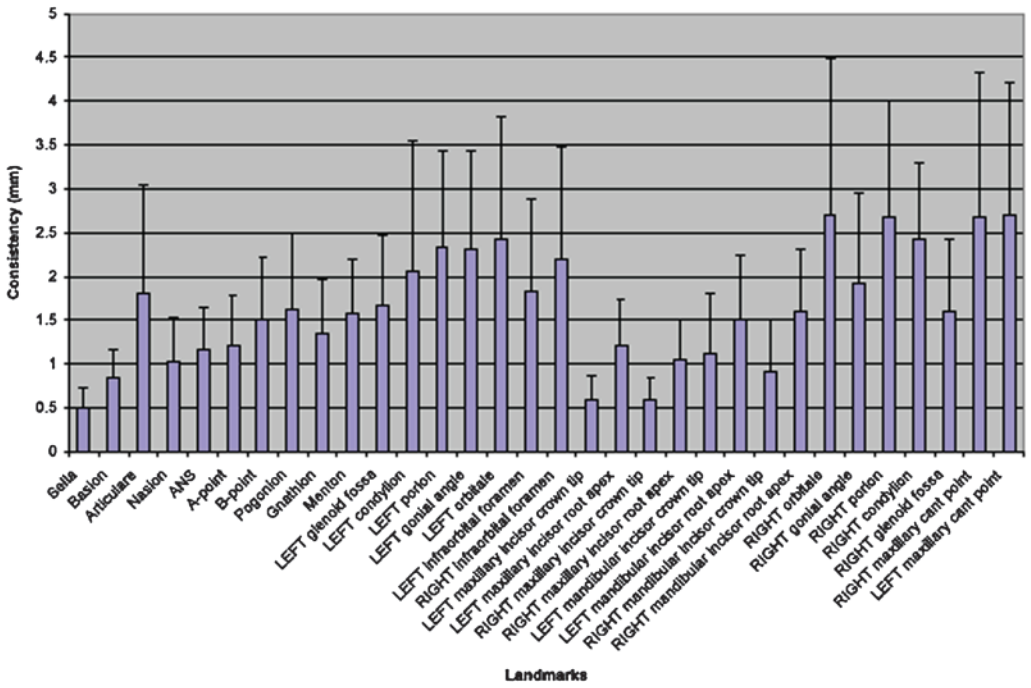

Figure 5 Overall consistency and precision: each bar is the cumulative consistency for all nine residents across all 19 patients. The error bars are 1 standard deviation from the mean consistency and are the measure of precision.

required the operator to place the landmark in the midline. Although there were only weak correlations (0.34–0.50), there was improvement in the consistency in all dimensions.

Nasion

Nasion, along with most of the other midline structures in this study, had an overall consistency of 1.02 mm (rank = 6th). The *y*-axis (vertical) was the most difficult axis for the examiners to agree upon (Figure 6b). Although the *y*-axis rank was 16th, the consistency was only 0.62 mm and, thus, not likely to be clinically significant. The good consistency was due to nasion being located at the intersection of two clearly identifiable structures located along the patient's midline.

Anterior nasal spine

In a typical lateral headfilm, anterior nasal spine (ANS) can be difficult to identify since the X-rays easily penetrate the thin and pointy structure leaving that area of the film exposed. With an overall consistency of 1.15 mm (rank = 9th), ANS is a structure that is much easier to identify with CBCT imaging. While the identification error was low, the distribution of the error was in keeping with previous research on landmarks. The *x*- and *y*-axes were consistently identified (0.47 and 0.36 mm) while the *z*-axis was consistently ranked 17th (0.76 mm). The difficulty with this landmark was not in the vertical or transverse dimensions but in the sagittal dimension. Even with CBCT technology, there appears still to be some blurring of the very tip of ANS, which leads to greater error. The improvement in overall consistency during the study was due solely to improvement in *z*-axis consistency. While the improvement in overall accuracy had a weak correlation, the *z*-axis improvement of 0.023 mm per patient had a moderate correlation ($r = 0.57$).

Point A

As a curved structure, the largest amount of error came from the axis parallel to the broadest curve. The large difference in consistency between the three axes illustrates this point. The consistency in the vertical dimension was ranked 27th (1.02 mm) while that of the *x*- and *z*-axes were ranked 5th (0.47 mm) and 6th (0.34 mm), respectively. These results are in keeping with the hypothesis that curved structures are more difficult to consistently identify.

Point B

Similar to point A, point B lies along a broad curve without clear anatomic boundaries. All measurements were less consistent than point A with the greatest error also in the *y*-axis (0.87 mm; rank = 23rd). There was no improvement in landmark identification, which indicates that the error was not due to misunderstanding of the definition or technical difficulty in locating the landmark.

Chin points—pogonion, gnathion, menton

Pogonion, gnathion, and menton are all located very close to one another along the anterior, antero-inferior, and

Table 3 Landmarks ranked by overall consistency.

Landmark	Overall consistency (mm)	*x*-axis consistency (mm)	*y*-axis consistency (mm)	*z*-axis consistency (mm)
Sella	0.50 ± 0.24 (1)	0.14 ± 0.13 (1)	0.31 ± 0.23 (4)	0.23 ± 0.16 (1)
Left maxillary incisor crown tip	0.58 ± 0.28 (2)	0.39 ± 0.32 (3)	0.23 ± 0.16 (2)	0.24 ± 0.15 (2)
Right maxillary incisor crown tip	0.59 ± 0.25 (3)	0.39 ± 0.31 (4)	0.17 ± 0.14 (1)	0.31 ± 0.21 (3)
Basion	0.85 ± 0.32 (4)	0.33 ± 0.25 (2)	0.35 ± 0.26 (5)	0.32 ± 0.28 (4)
Right mandibular incisor crown tip	0.91 ± 0.60 (5)	0.54 ± 0.43 (10)	0.37 ± 0.26 (8)	0.37 ± 0.26 (8)
Nasion	1.02 ± 0.50 (6)	0.48 ± 0.35 (8)	0.62 ± 0.49 (16)	0.33 ± 0.20 (5)
Right maxillary incisor root apex	1.05 ± 0.46 (7)	0.56 ± 0.44 (11)	0.63 ± 0.38 (18)	0.36 ± 0.31 (7)
Left mandibular incisor crown tip	1.13 ± 0.69 (8)	0.50 ± 0.41 (9)	0.46 ± 0.35 (11)	0.53 ± 0.43 (11)
ANS	1.15 ± 0.49 (9)	0.47 ± 0.35 (7)	0.36 ± 0.25 (6)	0.76 ± 0.52 (17)
Point A	1.20 ± 0.59 (10)	0.47 ± 0.36 (5)	1.07 ± 0.60 (27)	0.34 ± 0.24 (6)
Left maxillary incisor root apex	1.20 ± 0.53 (11)	0.47 ± 0.36 (6)	0.71 ± 0.40 (19)	0.38 ± 0.30 (9)
Gnathion	1.35 ± 0.61 (12)	0.67 ± 0.44 (15)	0.72 ± 0.38 (20)	0.78 ± 0.59 (19)
Left mandibular incisor root apex	1.50 ± 0.73 (13)	0.58 ± 0.46 (12)	0.76 ± 0.61 (21)	0.70 ± 0.58 (16)
Point B	1.50 ± 0.72 (14)	0.65 ± 0.43 (14)	0.87 ± 0.63 (23)	0.55 ± 0.39 (13)
Menton	1.58 ± 0.62 (15)	0.69 ± 0.45 (16)	0.38 ± 0.28 (9)	1.20 ± 0.73 (29)
Right glenoid fossa	1.59 ± 0.85 (16)	1.06 ± 0.82 (25)	0.31 ± 0.23 (3)	0.88 ± 0.59 (22)
Right mandibular incisor root apex	1.60 ± 0.71 (17)	0.72 ± 0.48 (19)	0.78 ± 0.45 (22)	0.79 ± 0.59 (20)
Pogonion	1.63 ± 0.87 (18)	0.70 ± 0.51 (17)	1.23 ± 0.95 (29)	0.50 ± 0.36 (10)
Left glenoid fossa	1.66 ± 0.82 (19)	1.11 ± 0.78 (27)	0.37 ± 0.35 (7)	0.82 ± 0.70 (21)
Articulare	1.81 ± 1.22 (20)	0.64 ± 0.52 (13)	0.92 ± 0.68 (25)	0.93 ± 1.11 (23)
Left infraorbital foramen	1.83 ± 1.05 (21)	0.92 ± 0.66 (21)	0.92 ± 0.84 (24)	0.94 ± 0.75 (25)
Right gonial angle	1.91 ± 1.05 (22)	0.71 ± 0.53 (18)	1.37 ± 1.03 (31)	0.77 ± 0.54 (18)
Left condylion	2.06 ± 1.49 (23)	1.04 ± 0.73 (24)	0.52 ± 0.49 (15)	1.50 ± 1.37 (30)
Right infraorbital foramen	2.19 ± 1.28 (24)	1.03 ± 0.79 (23)	1.25 ± 1.03 (30)	0.99 ± 0.73 (27)
Left gonial angle	2.31 ± 1.12 (25)	0.86 ± 0.59 (20)	1.78 ± 1.10 (32)	0.96 ± 0.58 (26)
Left porion	2.33 ± 1.11 (26)	2.20 ± 1.19 (30)	0.49 ± 0.35 (13)	0.62 ± 0.45 (15)
Right condylion	2.42 ± 0.87 (27)	1.25 ± 0.82 (28)	0.46 ± 0.38 (12)	1.07 ± 0.83 (28)
Left orbitale	2.43 ± 1.39 (28)	2.09 ± 1.49 (29)	0.62 ± 0.40 (17)	0.59 ± 0.42 (14)
Right maxillary cant point	2.67 ± 1.65 (29)	1.00 ± 0.69 (22)	1.08 ± 0.78 (28)	1.83 ± 1.58 (32)
Right porion	2.68 ± 1.32 (30)	2.40 ± 1.30 (32)	0.50 ± 0.40 (14)	0.55 ± 0.41 (12)
Right orbitale	2.69 ± 1.79 (31)	2.37 ± 1.84 (31)	0.43 ± 0.40 (10)	0.94 ± 0.85 (24)
Left maxillary cant point	2.70 ± 1.51 (32)	1.08 ± 0.72 (26)	1.00 ± 0.84 (26)	1.70 ± 1.36 (31)

The first number in each cell is the mean consistency for that landmark across all nine assessors and all 19 patients. This value is followed by the value of 1 standard deviation from the mean consistency. The number in the parentheses is that landmark's rank for each column. For example, sella (grey cell) has a consistency of 0.31 mm in the *y*-axis dimension with a SD of 0.23 mm. Sella's *y*-axis consistency is ranked 4th.

inferior surfaces of the chin, respectively. The pattern of error distribution seen with points A and B was also evident with these landmarks. The point on the most anterior surface of the chin, pogonion had less error in the *x*- and *z*-axes (0.70 and 0.50 mm) than in the *y*-axis (1.23 mm). Although the trend line correlation coefficients were weak to moderate ($r = 0.22–0.60$), there was a slight tendency for pogonion to be identified with less consistency as the study progressed. Gnathion, just inferior to pogonion, was the most consistently identified of the three points along the chin (1.35 mm; rank = 12th). The error among the axes was evenly distributed. As the most antero-inferior point of the anterior mandible, the definition and anatomy of gnathion leaves little room for interpretation. The same cannot be said for menton where again one axis was identified with less consistency than the others. Whereas points A and B and pogonion are along curves that are orientated vertically, menton is along a curve on the inferior surface of the anterior mandible. Thus, the dimension with the greatest error was the *z*-axis (1.20 mm; rank = 29th).

Glenoid fossa

The glenoid fossa was never part of lateral cephalometric analysis due to the inability to visualize it on a 2D film. As a potential landmark for analysing cranial base asymmetries, the glenoid fossa consistency was in the middle of the 32 landmarks. The right and left consistency did not differ significantly (1.59 and 1.66 mm). Identifying the most superior point was not a problem for the examiners as evidenced by the good *y*-axis ranking of 3rd (right) and 7th (left). There was less agreement as to the transverse and sagittal locations of the most superior point of the fossa.

Condylion

Without the problem of superimposition of multiple structures on a classic 2D film, condylion was thought to be a landmark that would be much easier to identify using CBCT. The most superior point on a structure that is not a perfectly rounded dome proved to be difficult to consistently identify. There was a better consistency on the left

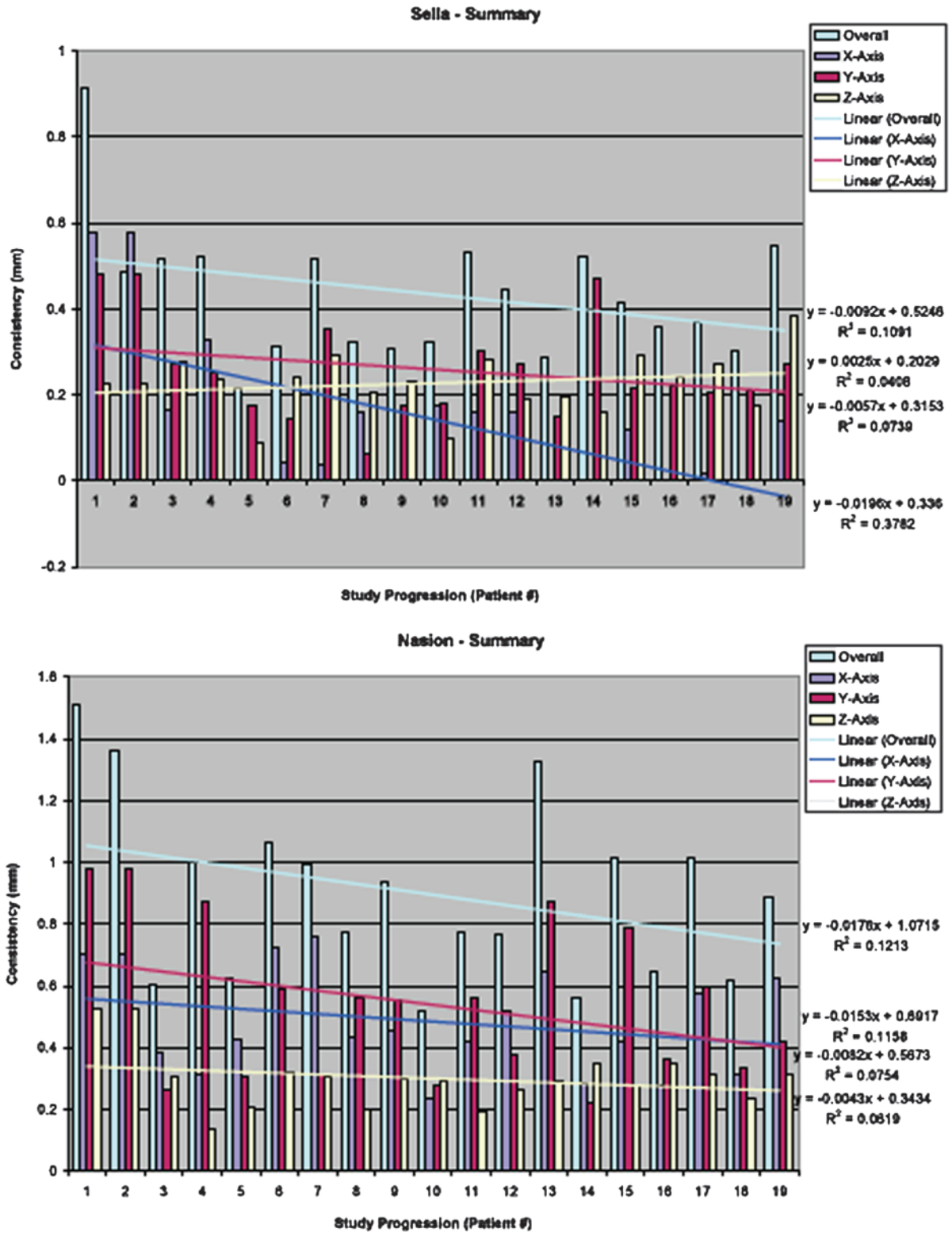

Figure 6 Examination of consistency for each of the three axes, as well as the overall change, with lines indicating trends in consistency for a given axis as the nine examiners' values were combined for each successively completed patient proceeding from patient 1 to patient 19. (a) sella and (b) nasion.

(2.06 mm; rank 23rd) than on the right (2.42 mm; rank 27th) side. Similar to glenoid fossa, the *y*-axis was the source of the least amount of error (left: 0.52 mm and right: 0.46 mm). Even when outliers were removed from the sample, there continued to be significant variation in the consistency from patient to patient. There was no significant

improvement or worsening of how consistently condylion was indentified.

Porion

The definition of porion did not include a description of its transverse location within the external auditory meatus. This proved to cause significant identification difficulty. Overall, porion appeared to have significant identification error (left: 2.33 mm; rank 26th and right: 2.68 mm; rank 30th). This is one point where the vast majority of the overall inconsistency was due to errors in one axis. The vertical and sagittal axes showed consistency while the transverse axis was ranked 30th on the left side and 32nd on the right side. The effect of the *x*-axis error on the overall consistency was noted. These results suggest that porion should be used for analysis of vertical and sagittal measurements (cants, Frankfort horizontal, etc.) but not for any analysis in the transverse dimension (left versus right asymmetries).

Gonion

Gonion is located along a broad curve on the posterior mandible, and its location is open to interpretation. This landmark is easily identified on a traditional lateral headfilm so the only advantage with 3D imaging is that the left and right gonial angles can be both identified and used for new analyses. The vertical dimension was ranked the lowest at 31 (left: 1.37 mm) and 32 (right: 1.78 mm). This bilateral similarity further suggests that the reason for the inconsistency in the *y*-axis was due to either the lack of clarity in the definition and/or the anatomy of the landmark. Of all landmarks located along curves, orbitale and gonion were on the broadest curves.

Orbitale

Orbitale has traditionally been a landmark used to establish the Frankfort horizontal reference line. A number of measurements are made from angles formed by Frankfort horizontal and other dental and skeletal lines. With the addition of the transverse dimension, this landmark also has the potential to be used for analysing canting of the orbits. The *x*-axis was largely responsible for its poor ranking among the other landmarks (left: 28th and right: 31st). The consistency for the *y*- and *z*-axes was less than 1 mm suggesting that orbitale is suitable for establishing Frankfort horizontal and also for evaluating canting of the orbits. However, the significant transverse inconsistency prevents its use for analysing differences in the transverse position of the orbits (i.e. hyper- and hypotelorism).

Infraorbital foramen (infraorbitale)

Fine structures that are difficult or impossible to visualize on lateral headfilms are now visible with CBCT images. Foramens are one such anatomic feature that could be used in craniofacial analysis. The reason for studying the consistency of infraorbital foramen was that it could be an alternative for orbitale. While the consistency was in the lower third (left: 1.83 mm, rank 21st and right: 2.19 mm, rank 24th), the axes all contributed equally. As a discrete anatomic point, the errors were similar in each dimension. This landmark was identified with greater consistency than the traditional orbitale landmark.

Maxillary cant points/palatal recess

The orientation and width of the palate are often difficult to assess on PA lateral headfilms. A common method to analyse transverse discrepancies between the maxilla and mandible has been to compare the position of the maxillary J-points with the mandibular antegonion notches. The J-points are part of a complex 3D structure that would be very difficult to consistently identify using CBCT imaging. The maxillary cant points allow measurement of the width of the maxillary arch while providing a means to analyse canting. Unfortunately, they were inconsistently identified (left: 2.70 mm, rank 32nd and right: 2.67 mm, rank 29th). The *x*- and *y*-axes are the critical dimensions if these points are to be used for transverse measurements and canting analysis. Within these dimensions, the consistency was ranked poorly (22nd–28th) but measured between 1.00 and 1.08 mm. There was an insignificant change in the consistency overall and also for each axis throughout the study. With the SD in the *x*-axis of 0.70 mm, the width of the palate could be measured 95 per cent of the time to within 2.80 mm.

Dental landmarks

The midpoints along the incisal edges and root apices of the maxillary and mandibular incisors were identified. The root apex of the mandibular incisors is often difficult to identify on a traditional headfilm due to superimposition of the canine, lateral incisor, and central root apices. In addition, since the most prominent incisor is measured on a lateral headfilm, it is difficult to know which root apex belongs to the most prominent incisal edge. A third source of error when identifying incisors on lateral headfilms is that rotated incisors can appear more proclined since the mesial or distal surface is more anterior than the middle of the incisal edge. CBCT imaging allows unadulterated views of these structures.

For a given tooth, the incisal edge was always identified more consistently than its respective root apex. The maxillary incisal edge and root apex measurements were all less than the similar mandibular measurements. Among the incisal edge measurements, the *x*-axis had the most variation. The variation was more evenly distributed among the root apices.

Discussion

CBCT imaging allows full examination of the craniofacial complex without superimposition of complex structures on a traditional 2D film (van Vlijmen *et al.*, 2009). A systematic review of landmarks used data from spiral CT scans (Kragskov *et al.*, 1997). This present study using CBCT demonstrated some additional factors. While certain landmarks exhibited some improvement throughout the present study, the general trend did not show significant improvement in landmark identification. The amount of time required to locate the 32 landmarks decreased from approximately 25 minutes to approximately 10 minutes. A possible reason for the lack of improvement is that while the examiners' speed increased, it was to the detriment of their identification consistency. Another likely possibility is that the definitions and anatomy leave a certain amount of interpretation that will always result in differences among individuals. A more recent study evaluating landmark identification from CBCT-generated data of 12 subjects prior to surgery showed that three judges, evaluating the same patient three times, demonstrated excellent intra- and inter-observer reliability with correct training and calibration (de Oliveira *et al.*, 2009).

Anatomy and landmark consistency

The primary aim of this research was to quantify the consistency of landmark identification using CBCT image sets. As in previous investigations using 2D lateral headfilms, the findings showed that each landmark has a unique quantity and distribution of error. While the specifics for each landmark have been presented above, some general trends were evident.

Midline landmarks

Except for pogonion and articulare, every midline landmark was identified with greater consistency than all bilateral skeletal structures. One possible reason for this difference is familiarity with the sagittal slice window in the 3D module. The window in the 3D module that displays the sagittal plane appears similar to a lateral headfilm once the midline of a patient is identified. In fact, the midline structures are clear in the sagittal plane window because the display is that of a 0.3 mm thick slice without the superimposition of multiple structures on to a single 2D film.

Another possible explanation for the consistency and precision among midline landmarks is the order of landmark identification. Midline structures were grouped together, which may have artificially improved the consistency and precision. The reason is that once sella was identified, the *x*-axis did not need to be drastically adjusted for subsequent landmarks. Therefore, the examiners were able to identify the other midline landmarks by only working with the other two axes. While the calibration training stressed the importance of checking all three dimensions prior to moving on to the next landmark, follow-up discussions with the examiners and evaluation of the raw data showed that many did not adjust the *x*-axis after identifying sella.

A third explanation is how the patient files were loaded into Dolphin Imaging's 3D module. The windows that display the sagittal, coronal, and transverse planes depend upon how the patient is orientated prior to identifying landmarks. So that the examiners would have windows that were accurate representations of the these planes, all patients were orientated so that as the examiners worked their way through the landmarks there would be little required change in the *x*-axis. How orientation affects landmark identification is an interesting question in the age of 3D data but was not an aim of this study. The likely reality regarding midline structure consistency is that all three of these explanations were involved but to different degrees for each resident.

Bilateral landmarks

Familiarity and anatomy are responsible for the poorer performance in locating bilateral landmarks. While many of the bilateral landmarks in this study are routinely identified on a lateral headfilm, the average clinician is not acutely aware of each landmark's 3D location. For example, orbitale has been classically defined as the most inferior point along the orbital rim when viewed on a lateral headfilm. The problem is that the definition lacks a transverse component. The definition in this study had the landmark placed on the most antero-inferior border of the orbit. Because orbital is roughly circular, once the most inferior point of the orbit is located, there can only be one transverse location. The combination of 'most anterior and most inferior' in the definition ensures that only one transverse location is the true landmark position. The complexity of redefining landmarks and changing the way clinicians view radiographs is more significant for bilateral landmarks. In contrast with midline landmarks, unfamiliarity with the radiographic appearance of bilateral landmarks made them less consistently located. Another explanation is that the bilateral landmarks are located along broad curves. Quite simply, the more broad the curve, the more difficult it is for the eye to capture this most prominent point or depression. For these reasons and possibly others, as a group, the bilateral landmarks showed less overall consistency than the midline landmarks.

Left versus right

An unanticipated finding was that landmarks on the left side were identified with greater consistency and precision than the same landmarks on the right side. Except for the glenoid fossa and palatal recess, all other bilateral skeletal structures showed this left–right difference. This study

involved a time commitment of roughly 20–30 minutes per patient from each assessor. To accurately identify the landmarks, the examiners had to concentrate not only on each patient's image but also on a new definition of many of the landmarks. That being said, the most likely reason for this left–right difference was frustration with the amount of time spent identifying 32 landmarks. Another reason suggested by one resident was that time was limited during clinical hours to complete these tasks. If they were near the end of a landmark set when a patient arrived, they would end up hurrying to finish which would lead to less consistent landmark identification. The combination of impatience with attempts to find time during busy clinical hours to complete the 19 patients may have caused the slight but consistent differences between left and right landmarks.

Discordant axes

The characteristic *x*, *y*, and *z* patterns should have implications in their uses in 3D analyses. Every landmark with very different *x*, *y*, and *z* rankings were found along curves. The dimension that showed the greatest error was parallel to the curve. For example, menton is located on the most inferior point along the inferior surface of the anterior mandible. The consistency in the vertical plane was excellent while the assessors had difficulty agreeing upon the antero-posterior location of the landmark. The same principle can be illustrated by pogonion. The curvature on the anterior surface of the mandible is broad in the vertical and transverse planes. The deepest point along this surface (*z*-axis) is relatively easy to identify while the vertical (*y*-axis) position is open to interpretation. Because pogonion is a midline structure, there was a relatively good consistency in the transverse plane (*x*-axis). The principle whereby the greatest error is in the dimension with the broadest curves holds for all the landmarks that had differences in the *x*-, *y*-, and *z*-axes rankings: points A and B, menton, right/left glenoid fossa, pogonion, articulare, right/left gonial angle, right/left porion, right/left condylion, and right/left orbitale.

Applications of findings

The major weakness of this study was that there was no comparison of 3D CBCT landmark identification and 2D lateral cephalometric landmark identification (Lagravere and Major, 2005; de Oliveira *et al.*, 2009). Only general comparisons with previous research can be made given the uniqueness of this study. One such comparison is the identification error of dental landmarks. The incisal edges of the incisors continue to be consistently identified. However, CBCT appears to offer a significant advantage when it comes to identifying the incisor root apices. Baumrind and Frantz (1971b) reported the 2D mean error for the lower incisor root apex to be 1.74 ± 0.59 mm while in the present study, the 3D mean error for the lower incisor root apices was 1.50 ± 0.73 and 1.60 ± 0.71 mm. While there was slightly less precision, the present study included four more examiners than that of Baumrind and Frantz (1971a,b). For the question of which method is a better choice for landmark identification to be answered, a study using both methods needs to be conducted. From the findings of the present research, landmark identification in three dimensions was excellent considering the worst performing landmarks had overall consistency of less than 3 mm after being identified as much as 171 times.

The most tangible findings in this study relate to the development of 3D analyses (Cevidanes *et al.*, 2005, 2006, 2007; Haney *et al.*, 2010; Stratemann *et al.*, 2010). If a landmark is to be used to evaluate a certain dimension, then it should be shown to have relatively good consistency and precision (Schlicher, 2008). Therefore, when planning which landmarks are to be used to analyse a certain dimension, the consistency should be known for that dimension. A simple first step towards deciding on landmarks would be to organize the landmarks by rank and then decide the amount of consistency to be tolerated for analysis. In the near future, 3D superimpositions will require landmarks for aligning of structures, and only the most accurate should be used. Structural superimpositions, similar to that currently used with lateral headfilms, will be developed. Subsequent studies should include direct comparison between lateral cephalometrics and 3D landmark identification, evaluation of the effect that 3D landmark identification error has on standard cephalometric measurements, and development of novel analyses that comprehensively evaluate the craniofacial complex (measurements and superimpositions).

Conclusions

1. There was no significant improvement in the overall consistency or precision in landmark identification as the examiners completed 19 patients.
2. Sella was the landmark with the best consistency at 0.50 ± 0.23 mm.
3. Left maxillary cant/palatal recess was the landmark with poorest consistency at 2.70 ± 1.51 mm.
4. Landmarks on curves continue to have more errors than those with clear anatomic delineations.
5. While overall consistency is important in understanding the general error in a landmark's identification error, an understanding of the individual axis consistency is essential for appreciating how error can affect craniofacial measurements in three dimensions.
6. Utilization of landmarks should take into account the distribution of error in each dimension so that the most accurate measurements can be made.

References

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

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

Broadbent B 1931 A new x-ray technique and its application in orthodontia. Angle Orthodontist 51: 93–114

Brown A A, Scarfe W C, Scheetz J P, Silveira A M, Farman A G 2009 Linear accuracy of cone beam CT-derived 3D images. Angle Orthodontist 79: 150–157

Cevidanes L H, Styner M A, Proffit W R 2006 Image analysis and superimposition of 3-dimensional cone-beam computed tomography models. American Journal of Orthodontics and Dentofacial Orthopedics 129: 611–618

Cevidanes L H *et al.* 2005 Superimposition of 3D cone-beam CT models of orthognathic surgery patients. Dentomaxillofacial Radiology 34: 369–375

Cevidanes L H *et al.* 2007 Three-dimensional cone-beam computed tomography for assessment of mandibular changes after orthognathic surgery. American Journal of Orthodontics and Dentofacial Orthopedics 131: 44–50

de Oliveira A E, Cevidanes L H, Phillips C, Motta A, Burke B, Tyndall D 2009 Observer reliability of three-dimensional cephalometric landmark identification on cone-beam computerized tomography. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics 107: 256–265

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

Haney E *et al.* 2010 Comparative analysis of traditional radiographs and cone beam computed tomography volumetric images in the diagnosis and treatment planning of maxillary impacted cuspids. American Journal of Orthodontics and Dentofacial Orthopedics 137: 590–597

Kragskov J, Bosch C, Gyldensted C, Sindet-Pedersen S 1997 Comparison of the reliability of craniofacial anatomic landmarks based on cephalometric radiographs and three-dimensional CT scans. Cleft Palate-Craniofacial Journal 34: 111–116

Kumar V, Ludlow J B, Mol A, Cevidanes L 2007 Comparison of conventional and cone beam CT synthesized cephalograms. Dentomaxillofacial Radiology 36: 263–269

Kumar V, Ludlow J, Cevidanes L H S, Mol A 2008 *In vivo* comparison of conventional and cone beam CT synthesized cephalograms. Angle Orthodontist 78: 873–879

Lagravere M O, Major P W 2005 Proposed reference point for 3-dimensional cephalometric analysis with cone-beam computerized tomography. American Journal of Orthodontics and Dentofacial Orthopedics 128: 657–660

Lagravere M O, Carey J, Toogood R W, Major P W 2008 Three-dimensional accuracy of measurements made with software on cone-beam computed tomography images. American Journal of Orthodontics and Dentofacial Orthopedics 134: 112–116

Leonardi R, Annunziata A, Caltabiano M 2008 Landmark identification error in posteroanterior cephalometric radiography. A systematic review. Angle Orthodontist 78: 761–765

Lou L, Lagravere M O, Compton S, Major P W, Flores-Mir C 2007 Accuracy of measurements and reliability of landmark identification with computed tomography (CT) techniques in the maxillofacial area: a systematic review. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics 104: 402–411

Major P W, Johnson D E, Hesse K L, Glover K E 1994 Landmark identification error in posterior anterior cephalometrics. Angle Orthodontist 64: 447–454

Moerenhout B A, Gelaude F, Swennen G R, Casselman J W, Van Der Sloten J, Mommaerts M Y 2009 Accuracy and repeatability of cone-beam computed tomography (CBCT) measurements used in the determination of facial indices in the laboratory setup. Journal of Craniomaxillofacial Surgery 37: 18–23

Park S H, Yu H S, Kim K D, Lee K J, Baik H S 2006 A proposal for a new analysis of craniofacial morphology by 3-dimensional computed tomography. American Journal of Orthodontics and Dentofacial Orthopedics 129: 600 e623–e634

Periago D R, Scarfe W C, Moshiri M, Scheetz J P, Silveira A M, Farman A G 2008 Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. Angle Orthodontist 78: 387–395

Perillo M *et al.* 2000 Effect of landmark identification on cephalometric measurements: guidelines for cephalometric analyses. Clinical Orthodontic Research 3: 29–36

Richardson A 1966 An investigation into the reproducibility of some points, planes, and lines used in cephalometric analysis. American Journal of Orthodontics 52: 637–651

Schlicher W 2008 Consistency and precision of landmark identification in three dimensions. Thesis. University of California at San Francisco

Solow B 1966 The pattern of craniofacial associations. Acta Odontologica Scandinavica 24: 26–36

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

Stratemann S A, Huang J C, Maki K, Hatcher D C, Miller A J 2010 Evaluating the human mandible using cone beam computed tomography. American Journal of Orthodontics and Dentofacial Orthopedics 137: S58–S570

Stratemann S A, Huang J C, Maki K, Miller A J, Hatcher D C 2008 Comparison of cone beam computed tomography imaging with physical measures. Dentomaxillofacial Radiology 37: 80–93

Tng T T, Chan T C, Hägg U, Cooke M S 1994 Validity of cephalometric landmarks. An experimental study on human skulls. European Journal of Orthodontics 16: 110–120

Trpkova B, Major P, Prasad N, Nebbe B 1997 Cephalometric landmarks identification and reproducibility: a meta analysis. American Journal of Orthodontics and Dentofacial Orthopedics 112: 165–170

van Vlijmen O J, Maal T J, Berge S J, Bronkhorst E M, Katsaros C, Kuijpers-Jagtman A M 2009 A comparison between two-dimensional and three-dimensional cephalometry on frontal radiographs and on cone beam computed tomography scans of human skulls. European Journal of Oral Sciences 117: 300–305