Original Article

Accuracy and Landmark Error Calculation Using Cone-Beam Computed Tomography–Generated Cephalograms

Dan Grauer[a]; Lucia S. H. Cevidanes[b]; Martin A. Styner[c]; Inam Heulfe[d]; Eric T. Harmon[e]; Hongtu Zhu[f]; William R. Proffit[g]

ABSTRACT

Objective: To evaluate systematic differences in landmark position between cone-beam computed tomography (CBCT)–generated cephalograms and conventional digital cephalograms and to estimate how much variability should be taken into account when both modalities are used within the same longitudinal study.

Materials and Methods: Landmarks on homologous cone-beam computed tomographic–generated cephalograms and conventional digital cephalograms of 46 patients were digitized, registered, and compared via the Hotelling T^2 test.

Results: There were no systematic differences between modalities in the position of most landmarks. Three landmarks showed statistically significant differences but did not reach clinical significance. A method for error calculation while combining both modalities in the same individual is presented.

Conclusion: In a longitudinal follow-up for assessment of treatment outcomes and growth of one individual, the error due to the combination of the two modalities might be larger than previously estimated. (*Angle Orthod.* 2010;80:286–294.)

KEY WORDS: Cephalogram; Cone-beam CT; Error; Procrustes; Accuracy; CBCT

INTRODUCTION

The advent of cone-beam computed tomography (CBCT) for craniofacial imaging provides volumetric information that allows development of virtual three-dimensional (3-D) models that can be quite valuable in locating impacted teeth, visualizing the temporomandibular joints, and diagnosing asymmetries in complex craniofacial patients.[1] Although new applications such as 3-D cephalometrics are developing rapidly, cephalograms are still necessary for comparison to existing databases,[2] and while 3-D registration and superimposition of CBCT data is being developed,[3] sequential cephalograms provide an easy clinical method for assessing growth and treatment changes. In order to be able to compare the new modalities with our current databases, algorithms have been created to extract information from the CBCT image and to simulate a conventional lateral cephalogram, P-A cephalogram, and panoramic projection. Previous in vitro and in vivo studies comparing both conventional cephalograms and CBCT-extracted cephalograms reported some statistically significant differences that did not reach clinical significance.[4–7]

The aims of this in vivo study were (1) to evaluate any systematic differences in landmark position between CBCT-generated cephalograms and conventional digital cephalograms, using an optimization method to superimpose sets of landmarks, and (2) to estimate how much variability should be taken into

[a] Postdoctoral Fellow, Department of Orthodontics, School of Dentistry, University of North Carolina, Chapel Hill, NC.
[b] Assistant Professor, Department of Orthodontics, School of Dentistry, University of North Carolina, Chapel Hill, NC.
[c] Assistant Professor, Department of Computer Science, School of Arts and Sciences, University of North Carolina, Chapel Hill, NC.
[d] Research Assistant, Department of Orthodontics, School of Dentistry, University of North Carolina, Chapel Hill, NC.
[e] Research Assistant, Department of Computer Science, School of Arts and Sciences, University of North Carolina, Chapel Hill, NC.
[f] Associate Professor, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC.
[g] Kenan Professor, Department of Orthodontics, School of Dentistry, University of North Carolina, Chapel Hill, NC.
Corresponding author: Dr Dan Grauer, Department of Orthodontics, University of North Carolina School of Dentistry, Chapel Hill, NC 27599-7450
(e-mail: grauerd@dentistry.unc.edu)

Accepted: June 2009. Submitted: March 2009.

 DOI: 10.2319/030909-135.1

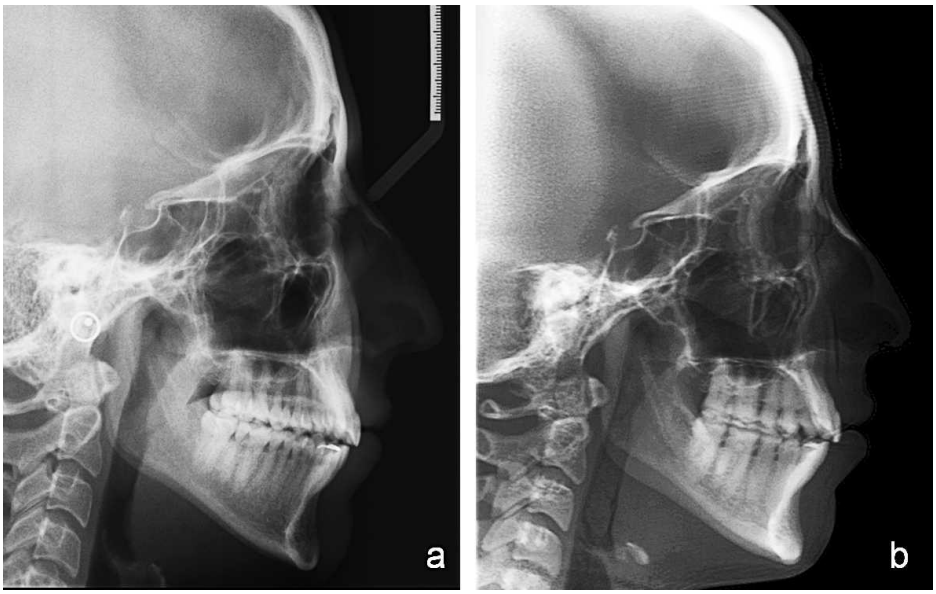

Figure 1. Different aspect of a conventional digital cephalogram (a) and a CBCT-generated cephalogram. (b) Note the difference in contrast and structure superimposition. For the digital cephalogram (JPEG file, 1360 × 2045; 8-bit; Proline, Planmeca, Helsinki, Finland); for the CBCT-generated cephalogram (16 × 22–cm large field of view, primary/axial image type, 1500/5000 window center/width, 400/400 rows/columns; iCAT, Imaging Sciences International).

account when combining conventional and synthetic cephalograms within the same longitudinal study.

MATERIALS AND METHODS

Records of consecutive patients who had radiographic examination at a radiology clinic between January 2005 and August 2006 were screened. Those for whom both a digital cephalogram (Planmeca, Helsinki, Finland) and a CBCT of the head (iCAT, Imaging Sciences International, Hatfield, Pa) had been obtained were selected. Initial inclusion criteria for this study were a medium- or full-field of view that allowed visualization of both the cranial base and the face and a patient age between 17 and 46 years. Records of 46 patients were available and included in the sample.

Creation of a Synthetic Cephalogram

CBCT images were converted into DICOM files and were rendered anonymous by an algorithm included in the iCAT software. Images were loaded into Dolphin 3D (version 2.3 beta) (Dolphin Imaging, Chatsworth, Calif). Threshold filters were set for optimal visualization of the soft and hard tissues.

Images were reoriented to align the cranium relative to the tridimensional coordinate system of Dolphin 3D (version 2.3 beta). Orbits were oriented parallel to the horizontal plane in the frontal view. In the sagittal view the cranium was rotated along the long axis so that the key ridges and orbits were aligned. A cranial view was used to confirm the correct head rotation by aligning the intracranial medial structures with the default coordinate system. Once the virtual 3-D models were aligned, synthetic cephalograms were created. The magnification factor was set to 7.5%, the typical magnification for midline structures with a 60-inch distance from radiation source to the midline with conventional cephalometrics, to simulate the magnification in conventional digital cephalograms. The images were enhanced for better visualization by fine tuning of the contrast and brightness options and were saved as JPEG files (Figure 1).

Cephalogram Tracing

Both conventional and synthetic cephalograms were loaded into Dolphin (version 9.1; Dolphin Imaging) and traced by a single operator. When landmarks were difficult to locate the operator was instructed to change the contrast, gamma, and brightness setting of the image until structures could be visualized. Whenever bilateral structures were not aligned, or when the difference in magnification was obvious between left and right structures, the operator chose the midpoint between the two structures. Cephalograms were verified for anatomic contour and landmark identification by a second operator. Fifteen cephalograms were selected from the sample and were retraced three times, with at least 24 hours in between tracing sessions. Intraclass correlation coefficients were above 0.9 for all landmarks both for x and y coordinates.

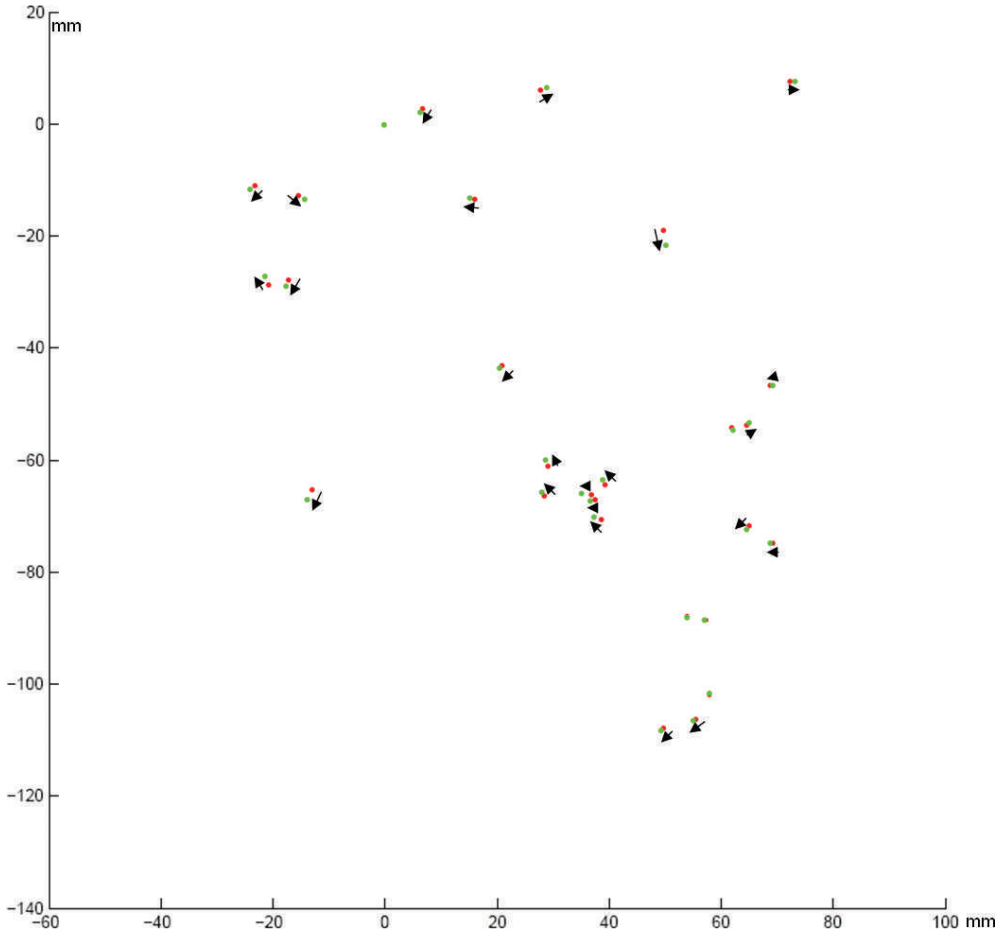

Figure 2. Landmarks located in the CBCT-generated cephalogram (red) have been registered via Procrustes method to the landmarks located on the conventional digital cephalogram (green). Difference vectors are depicted.

Registration Method

The two sets of landmarks belonging to each patient were registered in order to combine landmarks from both modalities into the same coordinate system. The following landmarks were used in the registration process: nasion, orbitale, ethmoid reg, sella ant, sella, articulare, pns, ans, a pt, menton, gnathion, pogonion, b pt, gonion, and porion.

In order to register the landmarks identified on the synthetic cephalogram to the ones belonging to the conventional digital cephalogram, rigid Procrustes registration was employed. Landmark coordinates were exported from Dolphin (version 9.1) into MathLab Software (The MathWorks Inc, Boston, Mass). First, the centers of gravity across all measurements were computed in each set of patient landmarks, both for the conventional and synthetic cephalograms. The centers of gravity of the conventional cephalogram landmarks and the synthetic cephalogram landmarks were superimposed. This process minimizes the translation differences between homologous landmarks while considering all the landmarks in the set. Secondly, an objective function that equals the sum of square distances between the landmark pairs was created. By minimizing this objective function, the best fit relative to the rotation of the two sets of landmarks was obtained.

Measurement

Average difference vector. The residual distances for each patient between homologous landmarks belonging to the two cephalogram modalities were calculated as vectors and will be referred to as ''difference vectors'' (Figure 2). The average difference at each landmark between synthetic and conventional cephalograms was calculated by averaging difference vectors from all patients. This difference will be referred to as the ''average difference vector'' (Table 1).

Average difference length. The absolute length of the individual difference vector is referred to as the ''difference length.'' Based on these length values, we then computed the ''average difference length'' via standard geometric averaging see (Table 1).

Table 1. Landmarks, Average Difference Vectors (X and Y Components and Module), Significance, False-Discovery Rate Method Correction, and Average Difference Lengths. Statistical Significance Was Established at 0.01 (Measurements are in mm)

Landmark	Average Difference Vector (ADV)		(ADV) Magnitude	*P* Value	*P* Value (FDR)	Average Difference Length (ADL)	SD[a] (ADL)
	Average X	Average Y					
Nasion	0.10	0.23	0.25	.595	.617	0.70	1.94
Orbitale	−0.07	0.38	0.39	.017	.067	1.26	1.88
Pterygo-maxillary fissure	0.01	−0.16	0.16	.638	.638	1.29	2.22
Ethmoid registration	0.11	−0.37	0.38	.122	.289	0.67	2.24
Sella anterior	−0.07	0.11	0.13	.415	.553	0.61	2.11
Sella	−0.09	0.01	0.09	.567	.611	0.51	2.03
Basion	−0.42	−0.49	0.64	.004	.031	1.18	2.50
Articulare	−0.19	−0.14	0.24	.124	.289	0.81	1.87
Condylion	0.18	−0.36	0.40	.212	.361	1.23	2.18
Posterior nasal spine	−0.25	−0.11	0.27	.048	.139	0.55	2.24
Anterior nasal spine	−0.48	−0.11	0.49	.001	**.007**	0.82	2.10
A pt	0.12	−0.03	0.12	.175	.350	0.65	1.79
Upper incisor incisal tip	0.34	−0.14	0.37	.000	**.003**	0.58	2.17
Upper incisor root apex	−0.05	−0.17	0.18	.172	.350	0.69	1.91
Upper first molar mesial contact	−0.13	−0.09	0.16	.459	.584	0.90	1.88
Upper first molar mesial cusp	0.05	−0.10	0.11	.539	.603	0.90	2.04
Upper first molar distal contact	−0.03	−0.17	0.17	.499	.603	0.79	2.36
Menton	−0.06	0.18	0.19	.219	.361	0.69	1.80
Gnathion	0.18	0.17	0.24	.266	.414	0.58	2.26
Pogonion	0.24	0.30	0.38	.007	.037	0.62	2.39
B pt	0.22	−0.45	0.50	.001	**.007**	0.99	1.61
Lower incisor incisal tip	0.19	−0.16	0.25	.015	.067	0.55	1.96
Lower incisor root apex	0.13	−0.08	0.15	.368	.516	0.72	1.85
Lower first molar mesial contact	0.20	−0.26	0.33	.046	.139	0.91	2.04
Lower first molar mesial cusp	0.23	−0.02	0.23	.333	.491	1.00	1.99
Lower first molar distal contact	0.14	−0.26	0.29	.050	.139	0.93	1.90
Gonion	−0.02	−0.20	0.21	.534	.603	0.94	2.32
Porion	0.28	0.02	0.28	.208	.361	1.04	2.10

[a] SD indicates standard deviation, FDR indicates False-Discovery Rate method.

Plotting. In order to visualize the difference vectors around each landmark, these vectors were transposed onto an arbitrarily selected landmark set (Figure 3). In order to visualize the envelope of landmark location probability, we plotted the average difference length (and two standard deviations) around each one of the landmarks (Figure 4).

Statistical Analysis

Statistical analyses were performed using SAS (version 9.1; SAS Institute Inc, Cary, NC). The hypothesis of interest was that there was no systematic difference between the two modalities at each landmark. We calculated the Hotelling T^2 statistic for the difference vectors between each pair of homologous landmarks in order to formally assess any systematic difference between the two modalities. To account for multiple comparisons across all landmarks, the false-discovery rate method was used.[8]

If the two modalities were to be used in a longitudinal study, the estimate of the measurement error has to account for the bias and variability derived from the use of two different modalities. Furthermore, to measure a distance on the cephalogram between two different landmarks, the envelope of error for both landmarks has to be considered.

In order to calculate the bias and variability of the measurement errors obtained from the use of the two modalities at each landmark (see Appendix), we used a two-step process. First, we calculated the difference vectors for all subjects and then computed the sample covariance matrix of these difference vectors. Second, we used the Gaussian random vector with a mean of zero and the half of the estimated covariance matrix to characterize measurement errors from both modalities.

To estimate the bias and variability of the distance between any two landmarks obtained from the use of the two modalities, we calculated the difference between the measured location difference vectors obtained from the two modalities and estimated their sample covariance matrix. Then, we can use the Gaussian random vector with a mean of zero and the half of the estimated covariance matrix to characterize measurement errors of location difference vectors between any two landmarks from both modalities.

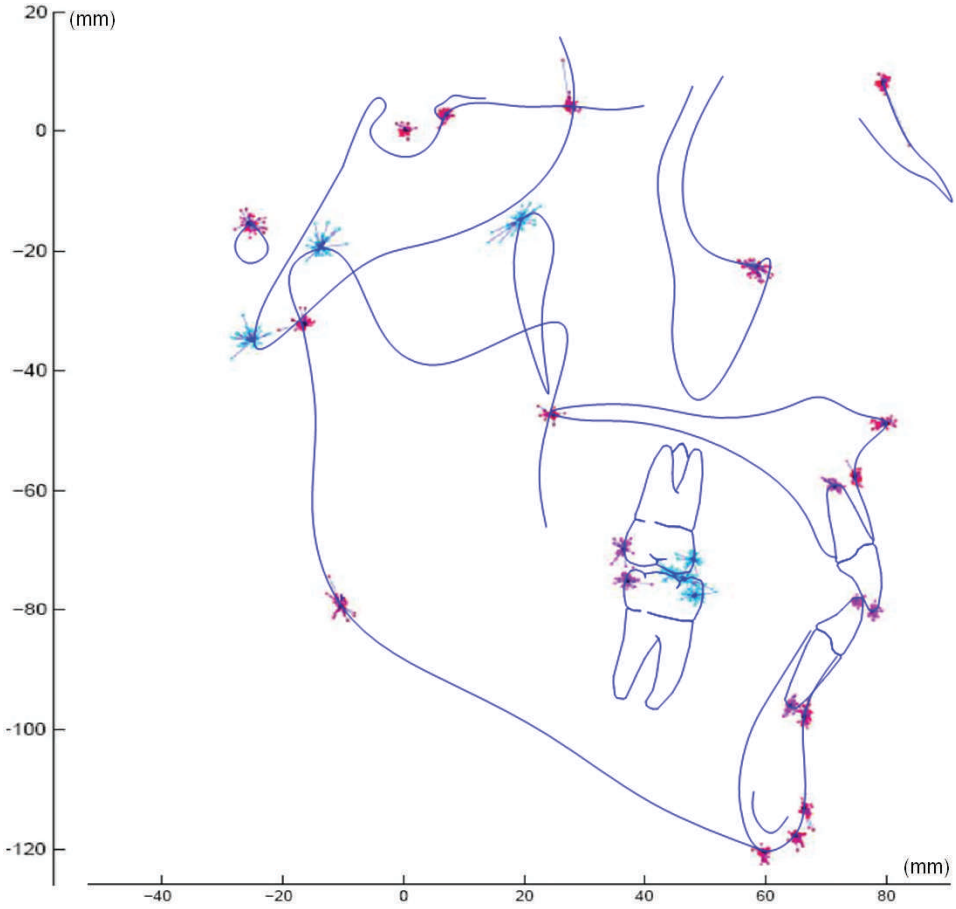

Figure 3. Difference vectors are grouped by landmark on a cephalogram tracing. The envelope of error—or difference between modalities—can be visualized. (Red and purple landmarks were used in the registration process; blue landmarks were only plotted.)

DISCUSSION

Registration Process

The Procrustes registration process is necessary to avoid an uneven distribution of error (differences) across landmarks. In order to compute the differences between modalities, homologous sets of landmarks have to be combined in the same coordinate system. Most studies simply compare absolute linear or angular measurements between modalities. These methods do not allow for establishment of directionality or discrimination between envelopes of landmark location probability.[4–7,9] Combining homologous sets of landmarks through an arbitrary coordinate center introduces bias.

The most frequent arbitrary coordinate center is centered in sella, with a horizontal plane described by a line 6° inferiorly rotated from sella-nasion plane. However, small differences in the locations of the landmarks that compose the coordinate system will have a great impact on the relative locations of landmarks located at a distance from the center of coordinates. The use of this arbitrary coordinate system to describe the relative coordinates of landmarks across modalities could lead to errors. Studies using the sella as the arbitrary coordinate center find their greater differences at mandibular structures or related measurements that are located far away from the coordinate system center.[10] In our method, the registration of homologous sets of landmarks and establishment of envelopes of landmark location probability did not depend on a single landmark but rather on a set of landmarks distributed uniformly across the head and face anatomy.

Sources of Variability

Main sources of variability that could affect our results are variability due to landmark identification and variability due to head orientation and alignment of x-ray emitter.

Landmark identification. The variability due to landmark identification displays characteristic patterns described by Baumrind and Frantz.[11] The systematic error in landmark identification affects both modalities, and it is likely that the net effect on the difference between modalities is negligible. In terms of landmark

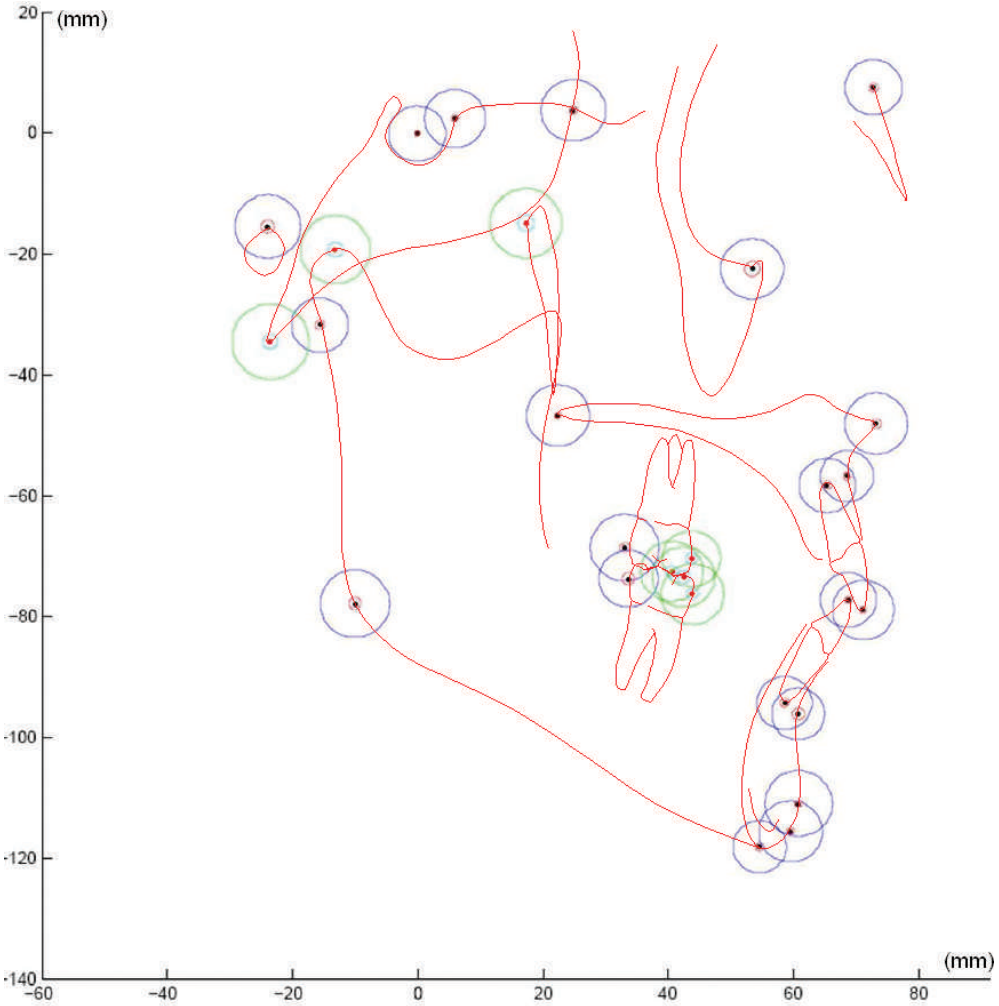

Figure 4. Difference lengths depicted as average plus three standard deviations are plotted on a cephalogram tracing. (Purple landmarks were used in the registration process; green landmarks were only plotted.)

identification, general findings in this study are in agreement with in vitro studies by Kumar et al[6] and Moshiri et al.[9] These studies measured dry skulls, and it is important to note that landmark identification is slightly more complex when soft tissue is present. The general aspect of a CBCT synthetic cephalogram is different from that of a conventional digital cephalogram (Figure 1). Landmark identification was easier in the synthetic cephalograms. Some landmarks that often lack the adequate contrast for an easy identification in conventional digital cephalograms were easily recognized because of the higher difference in contrast in the synthetic cephalograms.

Head orientation and alignment of x-ray emitter. Some of the differences found between homologous landmarks could be related to different head orientation. Malkoc et al[12] have found that linear and angular measurements on lateral cephalograms change from 16.1% to 44.7% with 14° of head rotation. Positioning of the patient inside the Planmeca cephalostat depends on the technician's skill, and that introduces another factor for which we cannot control.

The patient's anatomy also affects head positioning in the cephalostat. When the ears are used as a reference, we assume that the patient is relatively symmetric and that his/her ears are at the same level. In asymmetric patients this could create a head positioning error. Once the image is acquired, no corrections can be made to the roll and yaw of the head. Conversely, when a synthetic cephalogram is created the operator can easily manipulate the DICOM three-dimensionally to orient the head until bilateral structures are matching. The operator is able to see through the skull and match the position of para-medial structures. The position of the anatomical structures inside the field of view of the CBCT, in terms of rotation and translation, does not influence the accuracy of the measurements.[13] In this study, while creating the

synthetic cephalograms, no effort was made to replicate the position of the patient's head obtained in the conventional cephalograms.

Another source of projection errors is the misalignment of the x-ray emitter focal spot, which affects the conventional cephalogram machines. Even though we are certain that our x-ray unit was calibrated periodically, the fact that the cephalograms were obtained over a period of 18 months implies that the alignment of the x-ray source may have not been constant throughout the whole period. In an ex vivo study, Lee et al[14] reported that this type of misalignment could cause systematic error in the interpretation of facial asymmetry in PA cephalograms. That could be the case for conventional digital cephalograms too.

Dry Skull and In Vivo Studies

The accuracy and precision of measurements with CBCT have been assessed by several studies.[13,15,16] Ludlow et al[17] concluded that measuring in both reconstructed panoramic projection and in the 3-D volume through the stack of slices provides accurate measurements of mandibular anatomy. Lascala et al[18] reported a slight underestimation in linear measurements compared with direct measurements with a caliper used on skulls.

Our results are in agreement with ex vivo studies that have compared the accuracy and reliability of CBCT-generated cephalograms using skulls. Kumar et al[6] concluded that with dry skulls CBCT is comparable to conventional cephalometry in terms of precision and accuracy. In a recent article Moshiri et al[9] reported that CBCT-extracted cephalograms were, on average, more accurate than conventional digital lateral cephalograms when compared using direct measurement on skulls as a gold standard. In both studies, linear measurements of the mandible differed between the conventional and the CBCT synthetic cephalograms.

The findings from in vivo studies that assess differences in modalities are more directly comparable to our results. Recent in vivo studies have compared measurements between conventional cephalograms and CBCT-generated cephalograms and have concluded that even though some differences were found, they were not statistically or clinically significant.[4,5,7] These studies compared absolute measurements between modalities independently of landmarks' absolute coordinates. Given that there is no systematic error in landmark location between modalities, it is expected that the average differences in measurements reported between modalities would be centered around zero. When applied to an individual, the error in landmark location between modalities (or difference vector) could be much greater than the population average. When the two modalities are utilized in a longitudinal study of the same individual and when linear or angular measurements are computed, the reported error should include the envelope of landmark location probability at both landmarks (and at three landmarks if it is an angular measurement).

Table 2. Difference Between Modalities for Four Linear Measurements. Mean Difference, Standard Deviation (SD), and Percentiles (Measurements are in mm)

Length[a]	Mean Difference	SD	Percentile					
			10%	25%	40%	60%	75%	90%
ANS-me	0.90	0.49	0.32	0.53	0.71	0.96	1.21	1.56
N-Me	1.25	0.80	0.38	0.65	0.90	1.31	1.70	2.38
Co-Gn	1.37	0.73	0.53	0.83	1.09	1.48	1.80	2.36
Co-ANS	1.32	0.70	0.50	0.79	1.06	1.42	1.71	2.25

[a] ANS: Anterior Nasal Spine, Me: Menton, N: Nasion, Co: Condylion, Gn: Gnation.

With the method presented here, by calculating the envelope of landmark location probability around each landmark we can estimate the mean increase in error while measuring linear distances (Table 2). For instance, according to our method, if both modalities were used to calculate the distance between condylion and gnathion in an individual, the error could be as high as or higher than 2.36 mm (one out of 10 cases would display an error greater than 2.36 mm). This has an obvious impact when one is measuring small changes in mandibular length between time points. With our method, the error in measurement for any combination of two landmarks can be computed, and angular measurements can be analyzed similarly. In longitudinal follow-up for assessment of treatment outcomes and growth of one individual, the error due to combination of the two modalities might be larger than previously estimated.

In agreement with previous reports, the average difference in our study is below clinical significance. In longitudinal studies, when both modalities are used in the same individual, we should consider that the error of the method could produce clinically significant differences. This is especially the case when the variables measured display small incremental differences with growth. CBCT-generated cephalograms could be used as a diagnostic tool, but when assessing treatment outcomes at different times for one individual, the variability between modalities makes it advisable to obtain sequential records with the same modality.

RESULTS

The average differences in location between homologous landmarks in both modalities are shown in Table I and Figure 2 as the average difference vector and average difference length. In order to compare

difference vectors between patients, all sets of difference vectors around each landmark were transposed to an arbitrary center of coordinates and plotted (Figure 3). Most landmarks displayed a circular array of difference vectors. The average difference length and two standard deviations were also transposed to an arbitrary center of coordinates and plotted (Figure 4), which illustrates landmark location probability.

The distribution of the difference vectors was centered around zero for most landmarks, and there was no systematic difference between the two modalities. After adjustment for multiple comparisons via the false-discovery rate method (Table II), only three landmarks (ANS, MxI and B) showed a statistically significant difference, and even for these landmarks the magnitude of the differences did not reach clinical significance (0.5 mm).

CONCLUSIONS

- There is no systematic error when we compare average homologous landmark coordinates in conventional digital cephalograms and CBCT-generated cephalograms.
- In longitudinal studies, when both modalities are used in the same individual, the error of the method could produce clinically significant differences.

ACKNOWLEDGMENTS

We thank Drs Bob Scholz, Ceib Phillips, and David Hatcher for their help and support. Supported by NIDCR DE017727 and DE018962.

REFERENCES

1. Ludlow JB, Davies-Ludlow LE, Brooks SL. Dosimetry of two extraoral direct digital imaging devices: NewTom cone beam CT and Orthophos Plus DS panoramic unit. *Dentomaxillofac Radiol*. 2003;32:229–234.
2. Hunter WS, Baumrind S, Moyers RE. An inventory of United States and Canadian growth record sets: preliminary report. *Am J Orthod Dentofacial Orthop*. 1993;103:545–555.
3. Cevidanes LH, Styner MA, Proffit WR. Image analysis and superimposition of 3-dimensional cone-beam computed tomography models. *Am J Orthod Dentofacial Orthop*. 2006;129:611–618.
4. Cattaneo PM, Bloch CB, Calmar D, Hjortshoj M, Melsen B. Comparison between conventional and cone-beam computed tomography–generated cephalograms. *Am J Orthod Dentofacial Orthop*. 2008;134:798–802.
5. Kumar V, Ludlow J, Soares Cevidanes LH, Mol A. In vivo comparison of conventional and cone beam CT synthesized cephalograms. *Angle Orthod*. 2008;78:873–879.
6. Kumar V, Ludlow JB, Mol A, Cevidanes L. Comparison of conventional and cone beam CT synthesized cephalograms. *Dentomaxillofac Radiol*. 2007;36:263–269.
7. van Vlijmen OJ, Berge SJ, Swennen GR, Bronkhorst EM, Katsaros C, Kuijpers-Jagtman AM. Comparison of cephalometric radiographs obtained from cone-beam computed tomography scans and conventional radiographs. *J Oral Maxillofac Surg*. 2009;67:92–97.
8. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc Ser*. 1995;57:289–300.
9. Moshiri M, Scarfe WC, Hilgers ML, Scheetz JP, Silveira AM, Farman AG. Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. *Am J Orthod Dentofacial Orthop*. 2007;132:550–560.
10. Ludlow JB, Gubler M, Cevidanes L, Mol A. Precision of cephalometric landmark identification: CBCT vs conventional cephalometric views. *Am J Orthod Dentofacial Orthop*. In press.
11. Baumrind S, Frantz RC. The reliability of head film measurements. 1. Landmark identification. *Am J Orthod*. 1971;60:111–127.
12. Malkoc S, Sari Z, Usumez S, Koyuturk AE. The effect of head rotation on cephalometric radiographs. *Eur J Orthod*. 2005;27:315–321.
13. Kobayashi K, Shimoda S, Nakagawa Y, Yamamoto A. Accuracy in measurement of distance using limited cone-beam computerized tomography. *Int J Oral Maxillofac Implants*. 2004;19:228–231.
14. Lee KH, Hwang HS, Curry S, Boyd RL, Norris K, Baumrind S. Effect of cephalometer misalignment on calculations of facial asymmetry. *Am J Orthod Dentofacial Orthop*. 2007; 132:15–27.
15. Marmulla R, Wortche R, Muhling J, Hassfeld S. Geometric accuracy of the NewTom 9000 cone beam CT. *Dentomaxillofac Radiol*. 2005;34:28–31.
16. Mischkowski RA, Pulsfort R, Ritter L, et al. Geometric accuracy of a newly developed cone-beam device for maxillofacial imaging. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2007;104:551–559.
17. Ludlow JB, Laster WS, See M, Bailey LJ, Hershey HG. Accuracy of measurements of mandibular anatomy in cone beam computed tomography images. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2007;103:534–542.
18. Lascala CA, Panella J, Marques MM. Analysis of the accuracy of linear measurements obtained by cone beam computed tomography (CBCT-NewTom). *Dentomaxillofac Radiol*. 2004;33:291–294.

APPENDIX

Statistical Details

To estimate the bias and variability of the measurement errors obtained from the use of the two modalities at each landmark, we employed a two-step process. First, at the l-th landmark we assumed that $m_i^{(1)}(l) = \mu_i(l) + \varepsilon_i^{(1)}(l), m_i^{(2)}(l) = \mu_i(l) + \varepsilon_i^{(2)}(l)$, where $\mu_i(l)$ denotes the true location of the l-th landmark and where $m_i^{(1)}(l)$ and $m_i^{(2)}(l)$ represent the measurements obtained from the two modalities, respectively. Assuming that measurement errors $\varepsilon_i^{(1)}(l)$ and $\varepsilon_i^{(2)}(l)$ are independent Gaussian random vectors with mean zero and covariance $\Sigma(l)$, we can estimate $\Sigma(l)$ as follows: (1) calculate the difference vectors $m_i^{(1)}(l) - m_i^{(2)}(l)$ for all subjects and then compute the sample covariance matrix $S_i^{(1,2)}(l)$ of these difference vectors; (2) use

$S_i^{(1,2)}(l)/2$ as a consistent estimate of $\Sigma(l)$. Finally, we can use the Gaussian random vector with mean zero and covariance $S_i^{(1,2)}(l)/2$ to characterize measurement errors from both modalities.

Second, we estimated the bias and variability of the distance between any two landmarks obtained from the use of the two modalities. Specifically, we assumed that

$$m_i^{(1)}(l_1) - m_i^{(1)}(l_2) = \mu_i(l_1) - \mu_i(l_2) + \varepsilon_i^{(1)}(l_1) - \varepsilon_i^{(1)}(l_2),$$

$$m_i^{(2)}(l_1) - m_i^{(2)}(l_2) = \mu_i(l_1) - \mu_i(l_2) + \varepsilon_i^{(2)}(l_1) - \varepsilon_i^{(2)}(l_2),$$

where $\mu(l_1) - \mu(l_2)$ denotes the true location difference between the l_1-th and l_2-th landmarks and where $m_i^{(k)}(l_1) - m_i^{(k)}(l_2)$ for $k = 1, 2$ represents the measured location difference vector obtained from the two modalities. Assume that measurement error difference vectors $\varepsilon_i^{(1)}(l_1) - \varepsilon_i^{(1)}(l_2)$ and $\varepsilon_i^{(2)}(l_1) - \varepsilon_i^{(2)}(l_2)$ are independent Gaussian random vectors with mean zero and covariance $\Sigma(l_1, l_2)$. Similar to estimating $\Sigma(l)$, we can use the half of the sample covariance matrix of $m_i^{(1)}(l_1) - m_i^{(1)}(l_2) - m_i^{(2)}(l_1) + m_i^{(2)}(l_2)$, denoted by $S_i^{(1,2)}(l_1, l_2)/2$, to consistently estimate $\Sigma(l_1, l_2)$. Then, we can use the Gaussian random vector with mean zero and covariance $S_i^{(1,2)}(l_1, l_2)/2$ to characterize measurement errors of location difference vectors between any two landmarks from both modalities. Finally, we can estimate the bias and variability of the measurement error of the distance between any two landmarks from both modalities.