Accuracy of linear measurements using three imaging modalities: two lateral cephalograms and one 3D model from CBCT data

Pisha Pittayapat***, Michael M. Bornstein****, Thaís Sumie Nozu Imada****, Wim Coucke******, Ivo Lambrichts****** and Reinhide Jacobs*

*OIC, OMFS IMPATH Research Group, Department of Imaging & Pathology, Faculty of Medicine, University of Leuven and Oral & Maxillofacial Surgery, University Hospitals Leuven, Leuven, Belgium, **Department of Radiology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand, ***Department of Oral Surgery and Stomatology, School of Dental Medicine, University of Bern, Bern, Switzerland, ****Department of Stomatology, Bauru School of Dentistry, University of São Paulo (FDB/USP), São Paulo, Brazil, *****Department of Clinical Biology, Scientific Institute of Public Health, Brussels, Belgium, ******Biomedical Research Institute, Laboratory of Morphology, Hasselt University, Campus Diepenbeek, Diepenbeek, Belgium

Correspondence to: Pisha Pittayapat, OIC, OMFS IMPATH Research Group, Department of Imaging & Pathology, Faculty of Medicine, University of Leuven and Oral & Maxillofacial Surgery, University Hospitals Leuven, Kapucijnenvoer 7, Leuven 3000, Belgium. E-mail: p.pittayapat@gmail.com

Summary

Background: The aim of this study was to evaluate the accuracy of linear measurements on three imaging modalities: lateral cephalograms from a cephalometric machine with a 3 m source-to-mid-sagittal-plane distance (SMD), from a machine with 1.5 m SMD and 3D models from cone-beam computed tomography (CBCT) data.

Methods: Twenty-one dry human skulls were used. Lateral cephalograms were taken, using two cephalometric devices: one with a 3 m SMD and one with a 1.5 m SMD. CBCT scans were taken by 3D Accuitomo® 170, and 3D surface models were created in Maxilim® software. Thirteen linear measurements were completed twice by two observers with a 4 week interval. Direct physical measurements by a digital calliper were defined as the gold standard. Statistical analysis was performed.

Results: Nasion–Point A was significantly different from the gold standard in all methods. More statistically significant differences were found on the measurements of the 3 m SMD cephalograms in comparison to the other methods. Intra- and inter-observer agreement based on 3D measurements was slightly better than others.

Limitations: Dry human skulls without soft tissues were used. Therefore, the results have to be interpreted with caution, as they do not fully represent clinical conditions.

Conclusions: 3D measurements resulted in a better observer agreement. The accuracy of the measurements based on CBCT and 1.5 m SMD cephalogram was better than a 3 m SMD cephalogram. These findings demonstrated the linear measurements accuracy and reliability of 3D measurements based on CBCT data when compared to 2D techniques. Future studies should focus on the implementation of 3D cephalometry in clinical practice.
Introduction

A cephalometric analysis is a key element in orthodontic diagnostics. First introduced by Hofrath in Germany and Broadbent in the USA, this radiographic technique has been widely accepted as a standard tool for orthodontic treatment planning (1, 2). Traditionally, the technique is performed on a 2D lateral cephalogram, which does not represent the full dimensions of the human face and also has disadvantages such as geometric distortion and superimposition of anatomical structures. In the past, a cephalogram with a long distance between X-ray source and mid-sagittal plane of the patient’s face (3–4 m) was used. This type of machine allows more parallel X-ray beams, leading to less magnification of the images and possibly less radiation dose to the patient, when paired with sensitive image receptors (3–6). Today, most of the machines on the market combine panoramic and cephalometric radiographic options within one single device. The design of these machines is more compact, which allows for a 1.5 m distance between the X-ray source and the mid-sagittal plane of the patient’s face (7).

In recent years, 3D imaging modalities, especially cone-beam computed tomography (CBCT), have played an important role in dentistry because of lower radiation doses compared to the multi-slice CT and their availability in the dental offices (8). In orthodontics, 3D images have overcome the obstacle of 2D images by allowing orthodontists to visualize craniofacial structures without superimposition and distortion (9–11). Several publications have shown that the accuracy of 3D measurements is good or even superior to the measurements performed on lateral cephalograms (12–15). However, no investigation has directly compared the measurements from both left and right sides of the images. Furthermore, no study has compared measurements on images from a traditional cephalometric device with a long source-to-mid-sagittal distance (SMD).

The aim of this study was to evaluate the accuracy of linear measurements on three different imaging modalities for cephalometric analysis: lateral cephalograms from a cephalometric device with 3 m SMD, lateral cephalograms from a device with a 1.5 m SMD, and 3D models from cone-beam computed tomographic data.

Materials and methods

Sample

In total, 21 dry human skulls with present upper and lower first incisors and first molars were collected from the Department of Anatomy, Hasselt University, Diepenbeek, Belgium. Mandibles were attached to the skulls by taping around them starting from the temporal area of one side to the other. The occlusion was fixed at the maximum intercuspation. The study protocol (reference number: ML6960, BE322201010078) was approved by the UZ Leuven Medical Ethics Committee.

Imaging modalities

Three sets of radiographic images, two different types of lateral cephalograms and one CBCT, were acquired. First, lateral cephalograms of the dry skulls were taken by a 3 m SMD cephalometric machine with DX104 Comet tube (COMET, 3175 Flamatt, Switzerland; 70 kVp, 32–40 mA) at the University of Bern, Bern, Switzerland. Phosphor imaging plates, size 24 × 30 cm (Digora PCT system, Soredex, Tuusula, Finland) were used as image receptors. Second, lateral cephalograms were taken on the same samples by a digital cephalometric device with 1.5 m SMD equipped with complementary metal oxide semiconductor (CMOS) sensor (Cranex® 3D, Soredex; 81 kVp, 10 mA, 16 seconds), which has a CMOS receptor system. The dry skulls were placed in both devices and fixed with ear rods. The Frankford horizontal plane was adjusted to be parallel to the floor. Left and right sides were recorded by the main operator according to the anatomical structures without placing any radiopaque markers on the skulls during image acquisition. Last, CBCT scans were taken on the same skulls with a CBCT device (3D Accuitomo® 170, J. Morita, Kyoto, Japan) using the largest field of view (FOV): diameter 170 × height 120 mm (High-Fidelity/Hi-Fi mode: 90 kVp, 154 mA, voxel size: 0.25 mm). A 1.7 mm thick copper filter was attached to the machine during image acquisition to simulate soft tissue attenuation.

The two sets of lateral cephalograms were exported and stored in TIFF. The radiographs were then imported to Adobe® Photoshop CS4 (Adobe Systems Incorporated, San José, California, USA) and prepared for observation. A letter ‘L’ was placed on each image (both 1.5 m SMD group and 3 m SMD group) close to the angle of mandible to indicate the left side (Figure 1). CBCT data were exported to Digital Imaging and Communications in Medicine (DICOM) then imported to Maxilim® software (Medicim NV, Sint-Niklaas, Belgium). 3D surface models were created for all samples.

Cephalometric measurements

Ten cephalometric landmarks (Table 1) resulting in a total of 13 cephalometric linear measurements were included in this study (Table 2). Linear measurements including lateral landmarks were performed on both right and left sides.

The measurements of 2D lateral cephalometric groups, 1.5 m SMD group and 3 m SMD group, were done on Adobe® Photoshop CS4 (Figure 2). The digital cephalograms were calibrated by means of visible rulers and ear rods in the images. For the 3D group, all measurements were performed on Maxilim® software (Figure 2).

Figure 1. A letter ‘L’ was placed on the image close to the angle of the mandible to indicate the left side. (A) A lateral cephalogram from a cephalometric device with 1.5 m source-to-mid-sagittal-plane distance (SMD). (B) A lateral cephalogram from a cephalometric device with 3 m SMD. (C) A photograph of the same mandible as on the images.
The inter- and intra-observer variability of the cephalometric methods was evaluated by means of linear mixed models. The gold standard was taken as an explanatory variable, the observer as a random factor, and the measurements obtained by the different observers as dependent variables.

**Results**

Summary statistic was performed and the results are shown in Table 3. The biggest deviation can be observed on Go–Me (Table 3).

**Comparison with the gold standard**

3D measurements showed statistically significant differences ($P < 0.05$) from the gold standard for N–A and SmN–Go left. A statistically significant difference ($P < 0.05$) of N–A measurement was found for the 1.5 m SMD group. For lateral cephalogram with 3 m SMD, statistically significant differences ($P < 0.05$) were observed for N–A, N–B, N–Me, Go–Me right, Go–Me left, and Go–Co left measurements (Table 4).

**Comparison between cephalometric techniques**

When comparing 3D measurements with measurements on both 2D cephalograms, statistically significant differences ($P < 0.05$) were found for all measurements except N–ANS. When comparing between measurements on the 1.5 m SMD group and 3 m SMD group, all measurements were statistically significantly different ($P < 0.05$).

**Observer agreement**

Inter- and intra-observer variability of the 3D measurements was expressed as a percentage of coefficients of variability (CV). Inter-observer variability of 3D measurements (5.7 per cent) was lower than the other methods (6.3 per cent for 3 m SMD and 6.1 per cent for 1.5 m SMD cephalograms), which could be interpreted as a higher reproducibility for the measurements using CBCT images. For intra-observer variability, CV of 3D was shown to be between 2.4 and 3 per cent and also lower than the 2D methods (2.9–6 per cent for 3 m SMD and 3.6–4.1 per cent for 1.5 m SMD cephalograms). Therefore, the intra-observer agreement of 3D measurements was better than the two 2D measurements from both cephalometric devices.

**Discussion**

In the present study, the accuracy of linear measurements using three different types of imaging modalities for cephalometry was assessed
and comparisons of measurements among the techniques were performed. Although there are several publications that have evaluated the accuracy and compared measurements between 2D and 3D imaging techniques, there is, to the best of our knowledge, no publication in English that included cephalograms from a 3 m source-to-image-receptor distance.

Dry human skulls were used as in vitro subjects in this study to account for the fact that several different imaging modalities had to be taken on the same samples, and thus, it was unethical to use patients for this type of study. Although the skulls could not represent real human anatomy including soft tissues, this model offered some advantages. Direct measurements on hard tissue were possible unlike using real human subjects, and these were later used as gold standard. A 1.7 mm thick copper filter was used to mimic soft tissue attenuation during image acquisition to prevent any overexposures.

Cephalometric landmarks selected for this study included midline landmarks and lateral landmarks, both on the right and left side. Although only one measurement will be used for a lateral cephalogram in clinical situations, measurements of both sides were used in the present investigation in order to directly compare 3D with 2D measurements. No fiducial marker was placed prior to image acquisition. This was done in order to mimic the real clinical situation, as landmark identification is one of the variables affecting the intra- and inter-observer agreement. With marker placement in an experiment, this clinical observer bias might be largely eliminated.

Considering the imaging modalities used in the present study, 2D and 3D images are different in nature. 2D imaging systems, in this study the lateral cephalograms, are based on projecting shadows of anatomical structures on the image receptors. In 2D, structures aligned obliquely to the image receptor will result in distorted shadows on lateral cephalograms. These shadows are usually measured and used for cephalometric analysis. In this study, the effect of the SMD distance of two devices was evaluated, yet another factor that might influence magnification of 2D cephalometric radiographs is the distance from the mid-sagittal plane to the image receptor. In the current experimental set-up, both machines had a very similar 15 cm distance from the mid-sagittal plane to the image receptor, minimizing this secondary magnification bias.

On the other hand, for the 3D modality (CBCT), the image data were acquired and quantified in voxels, forming a realistic volume, which is definitely different from the 2D projection. This is one of the biggest advantages of 3D over the 2D imaging as it can capture

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Gold standard</th>
<th>3 m SMD group</th>
<th>1.5 m SMD group</th>
<th>3D group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave, SD, Min, Max</td>
<td>Ave, SD, Min, Max</td>
<td>Ave, SD, Min, Max</td>
<td>Ave, SD, Min, Max</td>
<td>Ave, SD, Min, Max</td>
</tr>
<tr>
<td>N–ANS</td>
<td>44.95, 3.66, 37.70, 50.94</td>
<td>43.99, 3.55, 36.45, 51.13</td>
<td>44.48, 3.59, 36.75, 52.09</td>
<td>45.42, 3.66, 37.40, 52.30</td>
</tr>
<tr>
<td>N–A</td>
<td>50.15, 4.51, 41.33, 58.10</td>
<td>48.10, 3.70, 40.28, 56.74</td>
<td>48.97, 3.84, 40.44, 55.82</td>
<td>49.98, 4.02, 41.90, 57.30</td>
</tr>
<tr>
<td>N–Me</td>
<td>86.78, 7.95, 70.49, 99.55</td>
<td>83.74, 7.12, 68.38, 97.20</td>
<td>85.06, 7.53, 69.77, 99.24</td>
<td>85.35, 7.70, 69.30, 100.00</td>
</tr>
<tr>
<td>N–Me</td>
<td>103.97, 10.32, 86.08, 122.47</td>
<td>102.39, 9.80, 84.14, 121.29</td>
<td>103.49, 10.05, 84.90, 122.65</td>
<td>104.43, 9.94, 85.70, 122.60</td>
</tr>
<tr>
<td>ANS–Me</td>
<td>60.74, 7.64, 47.91, 73.66</td>
<td>60.00, 7.35, 46.04, 73.98</td>
<td>60.81, 7.60, 47.18, 74.61</td>
<td>60.46, 7.48, 47.00, 74.70</td>
</tr>
<tr>
<td>ANS–PNS</td>
<td>48.07, 2.68, 43.57, 53.85</td>
<td>44.98, 2.89, 39.88, 53.56</td>
<td>47.12, 2.89, 41.74, 53.03</td>
<td>47.28, 2.85, 41.70, 53.30</td>
</tr>
<tr>
<td>Ba–PNS</td>
<td>40.73, 3.46, 34.73, 47.42</td>
<td>42.02, 6.02, 34.79, 73.53</td>
<td>41.93, 3.73, 34.32, 52.04</td>
<td>41.08, 3.41, 35.20, 47.90</td>
</tr>
<tr>
<td>SmN–Go right</td>
<td>38.13, 5.31, 27.92, 47.31</td>
<td>36.66, 6.25, 25.03, 67.02</td>
<td>36.56, 5.02, 26.28, 47.08</td>
<td>37.39, 5.14, 27.20, 47.60</td>
</tr>
<tr>
<td>SmN–Go left</td>
<td>38.81, 5.20, 28.25, 48.83</td>
<td>37.51, 6.09, 24.85, 67.02</td>
<td>38.96, 5.93, 27.85, 61.24</td>
<td>37.70, 4.94, 27.60, 47.70</td>
</tr>
<tr>
<td>Go–Co right</td>
<td>51.54, 6.00, 37.86, 59.44</td>
<td>48.12, 6.10, 36.74, 59.60</td>
<td>49.45, 5.81, 36.65, 61.04</td>
<td>50.56, 6.01, 35.10, 60.50</td>
</tr>
<tr>
<td>Go–Co left</td>
<td>52.32, 6.65, 38.69, 64.37</td>
<td>49.12, 5.94, 36.74, 62.89</td>
<td>51.51, 6.67, 37.34, 66.47</td>
<td>50.16, 6.31, 33.90, 61.70</td>
</tr>
<tr>
<td>Go–Me right</td>
<td>77.71, 6.43, 64.52, 88.83</td>
<td>62.80*, 5.61, 49.65, 75.97</td>
<td>64.17*, 5.86, 52.84, 75.92</td>
<td>77.86, 6.94, 62.80, 92.20</td>
</tr>
<tr>
<td>Go–Me left</td>
<td>76.70, 6.43, 63.42, 88.82</td>
<td>61.89*, 5.43, 49.89, 75.18</td>
<td>63.56*, 6.19, 49.66, 75.95</td>
<td>77.86, 7.10, 62.80, 95.70</td>
</tr>
</tbody>
</table>

* Measurements > 5 mm different from the gold standard.
structures with their real dimensional relationship. As a result, 3D image data, representing anatomical structures without any geometric distortion, can be measured. The linear distances in this study were defined as direct distances on 3D models, not orthogonal distances or distances created by projecting a 3D structure on a plane, which is the principle of 2D lateral cephalography. The purpose of this study design was to compare the measurements on 2D and 3D imaging modalities using their full capacity. Thus, a comparison of measurements by creating 2D projections from 3D data was avoided. Therefore, measurements with lateral landmarks are expected to exhibit the most pronounced differences between 2D and 3D imaging, which was demonstrated by Go–Me values on lateral cephalograms that significantly deviated from the gold standard and 3D measurements in the present study.

For the accuracy evaluation, it was found that the accuracy of measurements on 3 m SMD cephalograms was lower than the other two groups with six measurements exhibiting statistically significant differences when compared to the gold standard (Table 3). There was no English publication found to directly compare the results of the present study, but it was found in a few publications that the accuracy of 3D measurements was better than measurements on lateral cephalograms (16–18). In 2010, Varghese et al. (16) published results on the accuracy of CT and digital cephalometric measurements. The results showed that the accuracy of CT measurements was better than the 2D lateral cephalograms. Olmez et al. (17) found that there were no significant differences between the computer-assisted 3D and physical measurements, while the 2D measurements showed significant differences when compared to the physical measurements. Gribel et al. (18) investigated the accuracy and reliability of measurements on lateral cephalograms and CBCT. No statistically significant difference was found between CBCT measurements and the gold standard. However, for the lateral cephalograms, all measurements were statistically significantly different from the gold standard (18).

N–A was the only measurement that was statistically significantly different from the gold standard for all types of imaging techniques. This can be explained from previous studies published on landmark identification. It has been demonstrated in other investigations that point A (Table 1) was less reliable in terms of landmark identification (19–22). The position of the landmark situated on a curved surface such as the concavity of the alveolar process for point A may affect the accuracy of the identification more than a landmark situated on a small pointed area like ANS (19). This surely affected the accuracy of linear measurements in this study, when one of the landmarks was less reliable and more difficult to define or prone to subjectivity. Interestingly, Perillo et al. (20) stated that the lack of precision in identification of landmarks may not, on average, preclude cephalometric diagnosis.

It was speculated at the beginning of the study that measurements involving both a midline landmark (Me) and a lateral landmark (Go) would result in a statistically significant difference when compared to the gold standard. However, the results of the present study showed significant differences only for the 3 m SMD cephalometric group, but not for the 1.5 m SMD group. Table 3 shows deviations of the measurements of the 1.5 m SMD group from the gold standard values, but when the regression model was applied, the results were not statistically significant except for the N–A measurements (Table 4).

The comparison of all techniques with each other showed significant differences in almost all measurements in all pairs of techniques. This could mean that although the measurements were accurate when compared to the gold standard, the measurements were actually significantly different when comparing between the techniques. As shown in Table 3, some measurements including the lateral landmarks were highly deviated among cephalometric techniques, but when compared to the gold standard, the result was not statistically significantly different.

The results of the 3 m SMD cephalometric group were rather unexpected because the system should have provided a less to none magnified lateral cephalogram; thus, the midline landmarks should have been close to those obtained from gold standard physical measurements. One reason that could help explain this circumstance was the quality of the phosphor imaging plate. From the exposure parameter applied to the device, it did not give the image optimal brightness, contrast, and sharpness. This might have affected the landmark identification process. The images were calibrated properly by using the diameter of the ear rod as a reference—so this could be excluded as a possible factor affecting the measurement values.

The results of this study showed that the observer agreement of measurements on 3D models was slightly superior to the agreement...
of measurements on 2D lateral cephalograms, regardless of the type of the cephalometric device. This finding was in line with results published by previous studies. A study done by Griibel et al. (18) showed that measurements on 3D images [intraclass correlation coefficient (ICC) = 0.99] were as reliable as the measurements on 2D images (ICC = 0.98). On the other hand, Damstra et al. found the ICC of the 2D measurements on lateral cephalograms (ICC > 0.97) to be higher than the ICC of 3D measurements (ICC > 0.88), but there was no statistically significant difference between the two methods (23).

In a study by van Vlijmen et al. (24), the result was in the opposite direction to the present study. It was found that the intra-observer reliability of the measurements on the conventional cephalometric radiographs was higher compared with the intra-observer reliability of measurements on the 3D models. The authors suggested two factors possibly affecting the results: the learning curve in 3D tracing and the added third dimension of the image (24). In the present study, an intensive calibration of the observers was performed prior to the measurements, in order to familiarize with the software and landmark definition both in 2D and 3D. The results were therefore improved.

The results of the present investigation showed that inter-observer agreement is lower than the intra-observer agreement. This was expected, as observer performance can be affected by several factors such as background experiences, the familiarity of the observers to the software, and the ability to identify landmarks according to the definitions. A calibration session was conducted prior to the observation to minimize the effect of these factors.

To acquire CBCT images on real patients, dental CBCT examinations should be fully justified over conventional X-ray imaging and dose optimization by FOV collimation and low dose settings should be achieved (8, 25). Large FOV CBCTs should be used only when full indication and justification for the benefit of the patient is applied, as the radiation dose received from the CBCTs is strongly related to FOV size and also dependent on the specific CBCT machine (8). Recent guidelines on orthodontic use of CBCT imaging were published by the American Academy of Oral and Maxillofacial Radiology (26). Furthermore, guidelines and recommendations on CBCT use for dental and maxillofacial radiology have been made available by the European Commission to offer clinicians and orthodontists some guidance and recommendations (27). In general, the selection of radiographic imaging should be based on initial clinical evaluation and must be justified based on individual need without being considered ‘routine’ (26, 28).

Especially when treating children and young adults, the decision to perform a CBCT examination must be based on the patient's history, clinical examination, available radiographic imaging, and the presence of a clinical condition for which the benefits of the diagnosis and/or treatment plan outweigh the potential risks of exposure to radiation (26, 28). Therefore, 3D cephalometric analysis and 3D orthodontic treatment planning should only be performed when their benefits to the patients in specific cases can overcome the radiation risk.

Conclusions

This study has confirmed the knowledge on the accuracy of linear cephalometric measurements of 2D and 3D images. Although the results did not show that 3D measurements were more accurate than the 2D standard digital lateral cephalograms (1.5 m SMD), the results did confirm that 3D measurements were more reliable than measurements on 2D images.

Funding

Interfaculty Council for Development Co-operation (IRO).

Acknowledgements

The authors would like to thank Dr K. Rovaris who participated in the observations and Dr K. Dula for his expert advice.

References