Measuring the airway in 3 dimensions: A reliability and accuracy study

Hakan Ela and Juan Martin Palomo
Ankara, Turkey, and Cleveland, Ohio

Introduction: The aim of the study was to compare the reliability and accuracy of 3 commercially available digital imaging and communications in medicine (DICOM) viewers for measuring upper airway volumes.

Methods: Thirty cone-beam computed tomography scans were randomly selected, and the upper airway volumes were calculated for both oropharynx and nasal passage. Dolphin3D (version 11, Dolphin Imaging & Management Solutions, Chatsworth, Calif), InVivoDental (version 4.0.70, Anatomage, San Jose, Calif), and OnDemand3D (version 1.0.1.8407, CyberMed, Seoul, Korea) were compared with a previously tested manual segmentation program called OrthoSegment (OS) (developed at the Department of Orthodontics at Case Western Reserve University, Cleveland, Ohio). The measurements were repeated after 2 weeks, and the intraclass correlation coefficient was used for the reliability tests. All commercially available programs were compared with the OS program by using regression analysis. The Pearson correlation was used to evaluate the correlation between the OS and the automatic segmentation programs.

Results: The reliability was high for all programs. The highest correlation found was between the OS and Dolphin3D for the oropharynx, and between the OS and InVivoDental for nasal passage volume. A high correlation was found for all programs, but the results also showed statistically significant differences compared with the OS program. The programs also had inconsistencies among themselves.

Conclusions: The 3 commercially available DICOM viewers are highly reliable in their airway volume calculations and showed high correlation of results but poor accuracy, suggesting systematic errors. (Am J Orthod Dentofacial Orthop 2010;137:S50.e1-S50.e9)

The upper airway has long been an area of interest in orthodontics, with topics such as the relationships between facial type and airway, airway shape and volume with growth and development, and the clinician’s potential to modify the airway. However, most studies evaluating the airway have been conducted with 2-dimensional (2D) cephalograms, providing limited data such as linear and angular measurements, for a complex 3-dimensional (3D) structure.

With the introduction of cone-beam computed tomography (CBCT), the 3D diagnosis of the patient became more accessible in dentistry. CBCT has become a well-accepted oral and maxillofacial diagnostic imaging technique in a short time, and this was mainly due to lower radiation exposure and shorter scan acquisition times necessary to obtain an acceptable image compared with conventional computed tomography scans. CBCT technology allows the segmentation and visualization of hollow structures such as the airway in 3 dimensions. Thus, with 3D imaging, we are moving from lengths and angles toward volumes and surface areas. To visualize a CBCT scan, digital imaging and communications in medicine (DICOM) viewer software is necessary. DICOM is the accepted file format for a medical image, and a DICOM viewer allows viewing, measuring, segmenting, and complete analysis of a CBCT scan.

To segment and structure the airway means to delineate and remove all other surrounding structures for a clearer analysis and visualization. The segmentation of the airway can be done either manually or automatically. Manual segmentation requires the operator to delineate the airway slice by slice and then render all data into a 3D volume for analysis. Automatic segmentation can be done by differentiating structures with different density values. This means that, because the airway is radiolucent, the density values for the airway are lower than the values for the surrounding soft tissues, allowing easy and automatic...
differentiation. The density values are called Hounsfield units (HU). Automatic segmentation of the airway is significantly faster and more practical than manual segmentation, but the reliability and the accuracy of the method with commercially available programs have never been tested.

The aim of this study was to compare the reliability and accuracy of 3 commercially available DICOM viewers for measuring upper airway volume. Upper airway volume was divided into oropharynx (OP) and nasal passage (NP) for this study.

MATERIAL AND METHODS

Our experimental protocol was approved by the Case Western Reserve University institutional review board, and all records used in this study were obtained from the patient data base of the Department of Orthodontics. All CBCT images were acquired with a Hitachi CB Mercuray Scanner (Hitachi Medical Systems America, Twinsburg, Ohio) as a routine part of the initial diagnostic records for orthodontic patients. All images were preexisting and taken by using 2mA, 120kVp, and 12-in field of view (F mode) setting. Each patient’s image data consisted of 512 slices, with a slice thickness of 0.377 mm, a resolution of 1024 x 1024 pixels, and 12 bits per pixel (4096 gray scale). Scans in which the defined airway was not clear or fully contained in the image and images containing artifacts were excluded. Each patient signed a consent form allowing the use of orthodontic records for research purposes.

The study sample consisted of 30 randomly selected patients. All DICOM files were loaded to a computer with a Xeon (Intel, Santa Clara, Calif) processor, running Windows XP Professional (Microsoft, Redmond, Wash) operating system. For the commercially available DICOM viewers, the programs tested were Dolphin3D (version 11, Dolphin Imaging & Management Solutions, Chatsworth, Calif), InVivo-Dental (version 4.0.70, Anatomage, San Jose, Calif), and OnDemand3D (version 1.0.1.8407, CyberMed, Seoul, Korea). Dolphin3D (D3D), InVivoDental (IVD), OnDemand3D (OD3D), and a custom-written program with Visual C++ (Microsoft) for the Department of Orthodontics, School of Dental Medicine, at Case Western Reserve University called OrthoSegment (OS) were used to render the NP and OP airway volumes separately.

Amirlak et al.18 used the OS program to test the reliability and accuracy of CBCT images. They used a water displacement technique for comparing the CBCT volumes with actual volumes and found that the manually segmented volumes of the OS program were highly accurate. Based on their results, we used the values from the OS program as the gold standard to which the other results were compared.

For the OP volume, the superior and inferior limits were slightly modified from the limits used by Ogawa et al.14 OP volume was defined as the volume of the pharynx between the palatal plane (ANS-PNS line) extending to the posterior wall of the pharynx and the plane parallel to the palatal plane that passes from the most anteroinferior point of the second cervical vertebrae (Fig 1). The inferior limit of the NP airway was defined as the superior limit of the OP airway, and the superior limit was defined as the last slice before the nasal septum fused with the posterior wall of the pharynx. So, the superior border of the NP was defined on the axial slice first and then it was reflected to the sagittal plane (Fig 2). The described airway volumes were rendered with the D3D, IVD, and OD3D programs, according to their manufacturers’ recommendations (Figs 3-5).18-21 In the OS program, first, the limits of the airway were defined on the midsagittal slice on the sagittal view, and then the airway was painted slice by slice on the axial images.14 After we painted all slices between the defined limits, we rechecked the image for any inconsistencies from the coronal and sagittal aspects. Finally, the painted images were rendered, and a 3D volume was obtained (Fig 6).
measurements for OP airway volume was found with the iVD program (8.77 mm³), and the highest was for the OD3D program (97.2 mm³). The lowest mean difference between the first and second measurements for the nP airway volume was found with the OD3D (70.26 mm³), and the highest was found with the D3D program (515.78 mm³). nP airway volume differences were generally larger compared with the OP airway volume differences.

The results of the paired samples t test and the Pearson correlation coefficients are given on Table II. The highest correlations were found between the OS and D3D programs for OP airway, and between the OS and iVD for nP airway volume. Although there were high correlations between all programs, the results for each program showed statistically significant differences (Table II).

DISCUSSION

There are currently more than 15 third-party DICOM viewers mainly for orthodontics, implantology, and oral and maxillofacial surgery. Although the reliability and accuracy of CBCT machines have been evaluated, testing the reliability of CBCT-related software has not gone further than error studies with a few samples.

Yamashina et al. used a soft-tissue equivalent phantom to evaluate the reliability and accuracy of CBCT in measuring the density values of air, water, and soft tissues. They concluded that the measurement of the air space surrounded by soft tissues was quite accurate, concluding that the airway volume acquired from CBCT is nearly a 1-to-1 representation of the real volume.

In this study, 3 commercially available software programs that use automatic segmentation to calculate
Fig 3. Airway volume rendering with the D3D program: A, OP; B, NP.

Airway volumes were tested. Even though all 3 programs use automatic segmentation, this does not mean that they all use the same methods. Some differences between these programs are described below.

The IVD program allows more control where the user can “sculpt out” the desired airway volume from the rest of the 3D structures and, by adjusting the brightness and opacity values, clean out the unwanted voxels before calculating the final airway volume. The program also lets the user change the threshold values (range of density values displayed) to obtain a solid airway volume (Fig 4).

The automatic segmentation of the OD3D program requires the operator to first identify landmarks directly on the image, representing the density range of the area to be explored. The program does not allow control of the threshold for the area identified. Figure 5 shows a rendering made with the OD3D program. Even though a solid volume is visible on the 3D view, it can also be observed on the axial, sagittal, and coronal slices that the program...
All programs showed lower reliability when measuring NP volumes than for OP volumes. The NP airway is more challenging. With a manual segmentation program, it is harder to define NP areas, where the nasal turbinate and the concha region create an intricate anatomy. This area was also a challenge for the automatic segmentation programs, with most programs missing the nasal turbinates region where the NP airway has narrow spaces.

All images used in this study were preexisting, taken at 2 mA and 120 kVp during a 9.6-second rotation. These settings were chosen to follow the “as low...
would be more accurate, maintaining the already reliable values shown.

Manual segmentation seems to be the method with the greatest accuracy and allows the most operator control. But manual segmentation is also significantly more time-consuming and impractical for clinical use; it takes approximately 1 hour for both OP and NP volume calculations, whereas the same procedure with automatic segmentation can often be done in less than 5 minutes. All automatic segmentation programs compared with the manual segmentation program

Fig 5. Airway volume rendering with the OD3D program: A, OP; B, NP.
showed high correlation and poor accuracy. With this scenario, systematic errors are often suggested. The high correlation suggests that all programs behaved similarly, distinguishing large airways from small airways, but when a value was given, they were not the same among programs.

For automatic segmentation, volume measurements should be done with proper technique and diligence, because the measurement changes depend on the image threshold chosen. As in our study, when proper technique and protocol are followed, reliability of the results is high, but, when comparing the results of 3 commercially available programs, the values had high correlation and poor accuracy, showing sometimes overestimates or underestimates of the actual volumes. If airway volumes are to be used as
diagnostic information, the clinician should expect as reliable and accurate values as when linear or angular measurements are made. There is no norm at this point for airway volumes, perhaps because the airway volume is extremely variable, depending on head posture and breathing stage, but it is not unreasonable to expect consistency among different software packages for the same images.

### CONCLUSIONS

We found that 3 commercially available DICOM viewers are highly reliable in their airway volume calculations and showed high correlation of results, but poor accuracy, suggesting systematic errors.

### REFERENCES

7. Malkoc S, Usumez S, Iseri H. Long-term effects of symphyseal distraction and rapid maxillary expansion on pharyngeal

### Table I. Descriptive statistics of intraoperator means and standard deviations of measurements 1 and 2 of OP, NP, and total volumes for each program and the difference between the 2 measurements

<table>
<thead>
<tr>
<th>Program</th>
<th>OP airway (n = 30)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>NP airway (n = 30)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measurement 1 (mm³)</td>
<td>Measurement 2 (mm³)</td>
<td>Mean (mm³)</td>
<td>Mean (mm³)</td>
<td>Mean (mm³)</td>
<td>Measurement 1 (mm³)</td>
<td>Measurement 2 (mm³)</td>
<td>Mean (mm³)</td>
<td>Mean (mm³)</td>
<td>Mean (mm³)</td>
</tr>
<tr>
<td></td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
<td>Mean (mm³) (both means)</td>
</tr>
<tr>
<td>OrthoSegment</td>
<td>6801.85</td>
<td>2914.02</td>
<td>7411.23</td>
<td>3313.30</td>
<td>6249.07</td>
<td>2634.88</td>
<td>4651.66</td>
<td>1738.51</td>
<td>5919.84</td>
<td>2714.45</td>
</tr>
<tr>
<td>Dolphin3D</td>
<td>6790.76</td>
<td>3055.17</td>
<td>7477.50</td>
<td>3197.68</td>
<td>6240.30</td>
<td>2648.00</td>
<td>4554.46</td>
<td>1751.90</td>
<td>5569.06</td>
<td>2619.94</td>
</tr>
<tr>
<td>InVivoDental</td>
<td>6766.31</td>
<td>2980.83</td>
<td>7444.37</td>
<td>3250.08</td>
<td>6244.69</td>
<td>2631.50</td>
<td>4603.06</td>
<td>1741.03</td>
<td>5744.45</td>
<td>2654.50</td>
</tr>
<tr>
<td>OnDemand3D</td>
<td>71.09</td>
<td>66.28</td>
<td>8.77</td>
<td>97.2</td>
<td>350.78</td>
<td>515.78</td>
<td>71.03</td>
<td>70.26</td>
<td>350.78</td>
<td>515.78</td>
</tr>
<tr>
<td>ICC</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.88</td>
<td>0.97</td>
<td>0.89</td>
<td>0.98</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### Table II. Comparison of segmentation programs with linear regression analysis

<table>
<thead>
<tr>
<th>Program</th>
<th>OP volume (n = 30)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>NP volume (n = 30)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>ICC</td>
<td>Robust SE</td>
<td>t</td>
<td>P (&gt;</td>
<td>t</td>
<td>)</td>
<td>r</td>
<td>ICC</td>
<td>Robust SE</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.85</td>
<td>0.88</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>D3D</td>
<td>6801.85</td>
<td>2914.02</td>
<td>7411.23</td>
<td>3313.30</td>
<td>6249.07</td>
<td>2634.88</td>
<td>4651.66</td>
<td>1738.51</td>
<td>5919.84</td>
<td>2714.45</td>
</tr>
<tr>
<td>IVD</td>
<td>0.94</td>
<td>–521.62</td>
<td>190.62</td>
<td>–2.74</td>
<td>0.010*</td>
<td>0.95</td>
<td>0.85</td>
<td>–383.57</td>
<td>–5.65</td>
<td>0.000†</td>
</tr>
<tr>
<td>OD3D</td>
<td>0.74</td>
<td>–2163.25</td>
<td>382.57</td>
<td>–5.65</td>
<td>0.000†</td>
<td>0.85</td>
<td>0.85</td>
<td>–1236.49</td>
<td>–6.66</td>
<td>0.000†</td>
</tr>
<tr>
<td>OS (constant)</td>
<td>6766.31</td>
<td>551.21</td>
<td>12.28</td>
<td>0.000†</td>
<td>5638.95</td>
<td>7893.67</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>ICC (both measurements)</strong></td>
<td><strong>0.99</strong></td>
<td><strong>0.99</strong></td>
<td><strong>0.99</strong></td>
<td><strong>0.99</strong></td>
<td><strong>0.98</strong></td>
<td><strong>0.88</strong></td>
<td><strong>0.97</strong></td>
<td><strong>0.89</strong></td>
<td><strong>0.98</strong></td>
<td><strong>0.88</strong></td>
</tr>
</tbody>
</table>

r, Pearson correlation coefficient; OS, OrthoSegment; IVD, InVivoDental; OD3D, OnDemand3D; D3D, Dolphin3D.

*P <0.01; †P <0.001.