AJO-DO

Imaging software accuracy for 3-dimensional analysis of the upper airway

André Weissheimer,[a] Luciane Macedo de Menezes,[b] Glenn T. Sameshima,[c] Reyes Enciso,[d] John Pham,[e] and Dan Grauer[f]

Porto Alegre and Rio Grande do Sul, Brazil, and Los Angeles, Calif

Introduction: The aim of this study was to compare the precision and accuracy of 6 imaging software programs for measuring upper airway volumes in cone-beam computed tomography data. **Methods:** The sample consisted of 33 growing patients and an oropharynx acrylic phantom, scanned with an i-CAT scanner (Imaging Sciences International, Hatfield, Pa). The known oropharynx acrylic phantom volume was used as the gold standard. Semi-automatic segmentations with interactive and fixed threshold protocols of the patients' oropharynx and oropharynx acrylic phantom were performed by using Mimics (Materialise, Leuven, Belgium), ITK-Snap (www.itksnap.org), OsiriX (Pixmeo, Geneva, Switzerland), Dolphin3D (Dolphin Imaging & Management Solutions, Chatsworth, Calif), InVivo Dental (Anatomage, San Jose, Calif), and Ondemand3D (CyberMed, Seoul, Korea) software programs. The intraclass correlation coefficient was used for the reliability tests. A repeated measurements analysis of variance (ANOVA) test and post-hoc tests (Bonferroni) were used to compare the software programs. **Results:** The reliability was high for all programs. With the interactive threshold protocol, the oropharynx acrylic phantom segmentations with Mimics, Dolphin3D, OsiriX, and ITK-Snap showed less than 2% errors in volumes compared with the gold standard. Ondemand3D and InVivo Dental had more than 5% errors compared with the gold standard. With the fixed threshold protocol, the volume errors were similar (–11.1% to –11.7%) among the programs. In the oropharynx segmentation with the interactive protocol, ITK-Snap, Mimics, OsiriX, and Dolphin3D were statistically significantly different ($P < 0.05$) from InVivo Dental. No statistical difference ($P > 0.05$) was found between InVivo Dental and OnDemand3D. **Conclusions:** All 6 imaging software programs were reliable but had errors in the volume segmentations of the oropharynx. Mimics, Dolphin3D, ITK-Snap, and OsiriX were similar and more accurate than InVivo Dental and Ondemand3D for upper airway assessment. (Am J Orthod Dentofacial Orthop 2012;142:801-13)

For the last century, the gold standard method for analysis of craniofacial development was cephalometry, with linear and angular measurements performed on lateral headfilms. However, as a 2-dimensional representation of 3-dimensional (3D) structures, lateral headfilms offer limited information about the airways.[1] Information regarding axial cross-sectional areas and overall volumes can only be determined by 3D imaging modalities.[2-9] Medical computed tomography is a 3D imaging modality used in medicine but not as a routine method for airway analysis because of its high cost both financially and in terms of radiation. These drawbacks were overcome with the introduction of cone-beam computed tomography (CBCT). CBCT is becoming a popular diagnostic method for visualizing and analyzing upper airways. Since its introduction in 1998, CBCT technology has been improved, with lower costs, less radiation exposure to patients, and better accuracy in identifying the boundaries of soft tissues and empty spaces (air).[7] Furthermore, CBCT allows for the

[a]Postgraduate student, Department of Orthodontics, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil; visitor scholar, Graduate Orthodontics, Ostrow School of Dentistry, University of Southern California, Los Angeles.
[b]Associate professor, Department of Orthodontics, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.
[c]Associate professor and director, Graduate Orthodontics, Ostrow School of Dentistry, University of Southern California, Los Angeles.
[d]Assistant professor, Clinical Dentistry, Division of Dental Public Health and Pediatric Dentistry, Ostrow School of Dentistry, University of Southern California, Los Angeles.
[e]Biomedical and computer engineer and resident, Graduate Orthodontics, Ostrow School of Dentistry, University of Southern California, Los Angeles.
[f]Assistant professor, Clinical Dentistry, Graduate Orthodontics, Ostrow School of Dentistry, University of Southern California, Los Angeles.
The authors report no commercial, proprietary, or financial interest in the products or companies described in this article.
Supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES Foundation, process number BEX 5335-11-6).
Reprint requests to: André Weissheimer, Pontifícia Universidade Católica do Rio Grande do Sul, Avenida Ipiranga, 6681, Faculdade de Odontologia, Prédio 6 sala 209, Porto Alegre, RS, Brazil CEP 90619-900; e-mail, andre5051@hotmail.com.
Submitted, April 2012; revised and accepted, July 2012.
0889-5406/$36.00

http://dx.doi.org/10.1016/j.ajodo.2012.07.015

assessment of axial cross-sectional areas and volumes of the upper airways. The accuracy and reliability of CBCT for upper airway evaluation have been validated in previous studies,[2,6,8,10-12] and the use of CBCT for airway evaluation has been reported in a systematic review of the literature.[1]

The evaluation of the size, shape, and volume of the upper airway starts by defining the volume corresponding to the airway passages, a process called segmentation. In medical imaging, segmentation is defined as the construction of 3D virtual surface models (called *segmentations*) to match the volumetric data.[13] In other words, it means to separate a specific element (eg, upper airway) and remove all other structures of noninterest for better visualization and analysis. Upper airway segmentation can be either manual or semiautomatic. In the manual approach, the segmentation is performed slice by slice by the user. The software then combines all slices to form a 3D volume. This method is time-consuming and almost impractical for clinical application. In contrast, semiautomatic segmentation of the airway is significantly faster than manual segmentation.[14] In the semiautomatic approach, the computer automatically differentiates the air and the surrounding soft tissues by using the differences in density values (grey levels) of these structures. In some programs, the semiautomatic segmentation includes 2 user-guided interactive steps: placement of initial seed regions in the axial, coronal, and sagittal slices, and selection of an initial threshold.

Image thresholding is the basis for segmentation. When the user determines a threshold interval, it means that all voxels with grey levels inside that interval will be selected to construct the 3D model (segmentation). Lenza et al[7] reported the use of a single threshold value to segment the airway in each patient's CBCT scan. This approach can generate errors, especially in volume analysis, but it is certainly more reproducible than the use of a dynamic threshold. However, there are few studies comparing the results of threshold filtering with different imaging software programs for airway assessment.

A growing number of software programs to manage and analyze digital imaging communications in medicine (DICOM) files are introduced in the market every year. Many of these have incorporated tools to segment and measure the airway. A systematic review of the literature reported 18 imaging software programs for viewing and analyzing the upper airway in CBCT.[1] However, validation studies with a clear study design were performed in only 4 software programs.[10,14] The systematic review suggested that studies assessing the accuracy and reliability of current and new software programs must be conducted before these imaging software programs can be implemented for airway analysis.[1]

The aim of this study was to compare the precision and accuracy of 6 imaging software programs for measuring the oropharynx volume in CBCT images. The primary null hypothesis, that there are no significant differences in airway volume measurements among the 6 imaging software programs, was tested.

MATERIAL AND METHODS

This study was approved by the ethical committee of Pontifical Catholic University of Rio Grande do Sul in Brazil. The records we used were obtained from the patient database of the Department of Orthodontics and consisted of the pretreatment CBCT scans of a preexisting rapid maxillary expansion sample.[15] The sample included 33 growing patients (mean chronologic age, 10.7 years; range, 7.2-14.5 years) with transverse maxillary deficiency and no congenital malformations. Additionally, a custom-made oropharynx acrylic phantom with a known volume was used as the gold standard (Fig 1). The oropharynx acrylic phantom consisted of an air-filled plastic rectangular prism surrounded by water. Water was the medium of choice because it has a similar attenuation value to soft tissue. The dimensions of the outer surface of the phantom were created to simulate the dimensions of a growing patient's neck, and the rectangular prism to simulate the dimensions of the oropharynx. The oropharynx acrylic phantom's dimensions were measured to the nearest 0.01 mm by using digital calipers (model 727; Starret, Itú, São Paulo, Brazil), and the volume was calculated multiplying the base area by the height. Additionally, the oropharynx acrylic phantom's volume was confirmed by using the water weight equivalent. The oropharynx acrylic phantom was filled with distilled water, at 20°C, and the water weight was determined by using a digital scientific scale (model BG1000; Gehaka, São Paulo, São Paulo, Brazil).

Patients and the phantom were scanned with the i-CAT scanner (Imaging Sciences International, Hatfield, Pa) set at 120 kVp, 8 mA, scan time of 40 seconds, and 0.3-mm voxel dimension. The images were reconstructed with a 0.3-mm slice thickness and exported as DICOM files. Any CBCT scans with artifacts distorting the airway borders were excluded from this study. CBCT scans were imported into OsiriX software (version 4.0; Pixmeo, Geneva, Switzerland) for head orientation and definition of the oropharynx's region of interest. The head orientation was performed by using the palatal plane as a reference (ANS-PNS parallel to the global horizontal plane in the sagittal view and perpendicular to the global horizontal plane in the axial view). After head orientation, a tool in OsiriX (Vol cutter) was used

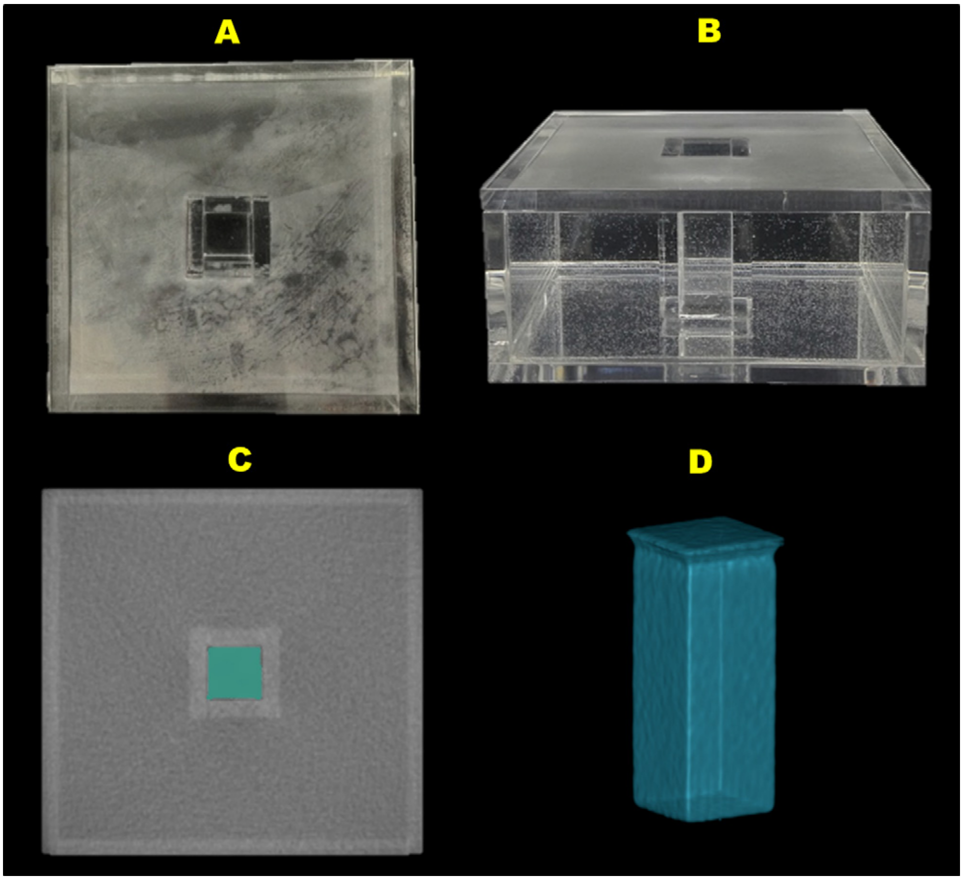

Fig 1. A and **B,** Oropharynx acrylic phantom used as the gold standard; **C,** CBCT image of the oropharynx acrylic phantom in the axial view, with the inner part (prism) containing air as the region of interest for segmentation; **D,** example of a 3D model generated from oropharynx acrylic phantom segmentation.

to select and crop the oropharynx, according to the following references: superior limit, extension of the palatal plane (ANS-PNS) to the posterior wall of the pharynx; inferior limit, a plane parallel to the palatal plane that passes through the most anteroinferior point of the second cervical vertebrae[14]; anterior limit, a perpendicular plane to palatal plane that passes through the posterior nasal spine; posterior limit, posteriorly to the posterior wall of pharynx (Fig 2, *A* and *B*). The regions of noninterest were excluded, and the new volume, containing only the oropharynx region, was exported as the new DICOM file, with the same voxel resolution but a smaller file size (Fig 2, *C*).

Six imaging software programs were used to segment and compute volumes from the CBCT images of the patients' oropharynx and the oropharynx acrylic phantom (Figs 3 and 4). The specifications of each software program are shown in Table I. The oropharynx and the oropharynx acrylic phantom segmentations were performed according to each software manufacturer's recommendations,[16-21] with 2 techniques for semiautomatic segmentation: (1) interactive threshold interval, as used in a previous study,[14] and (2) fixed threshold interval. In the interactive thresholding, the operator selected the best threshold interval based on a visual analysis of the anatomic boundaries of the oropharynx and the oropharynx acrylic phantom in the axial, sagittal, and coronal slices. In the fixed threshold, the threshold interval was fixed (−1000 to −587 grey levels) to test the variability among the software programs. The segmentations with interactive thresholding were performed with Mimics (version 14.12; Materialise, Leuven, Belgium), Dolphin3D (version 11.7; Dolphin Imaging & Management Solutions, Chatsworth, Calif), Ondemand3D (version 1.0.9.1451; CyberMed, Seoul, Korea), OsiriX (version 4.0; Pixmeo), and ITK-Snap (version 2.2.0; www.itksnap.org) (except InVivo Dental [version 5.0; Anatomage, San Jose, Calif]), and those with

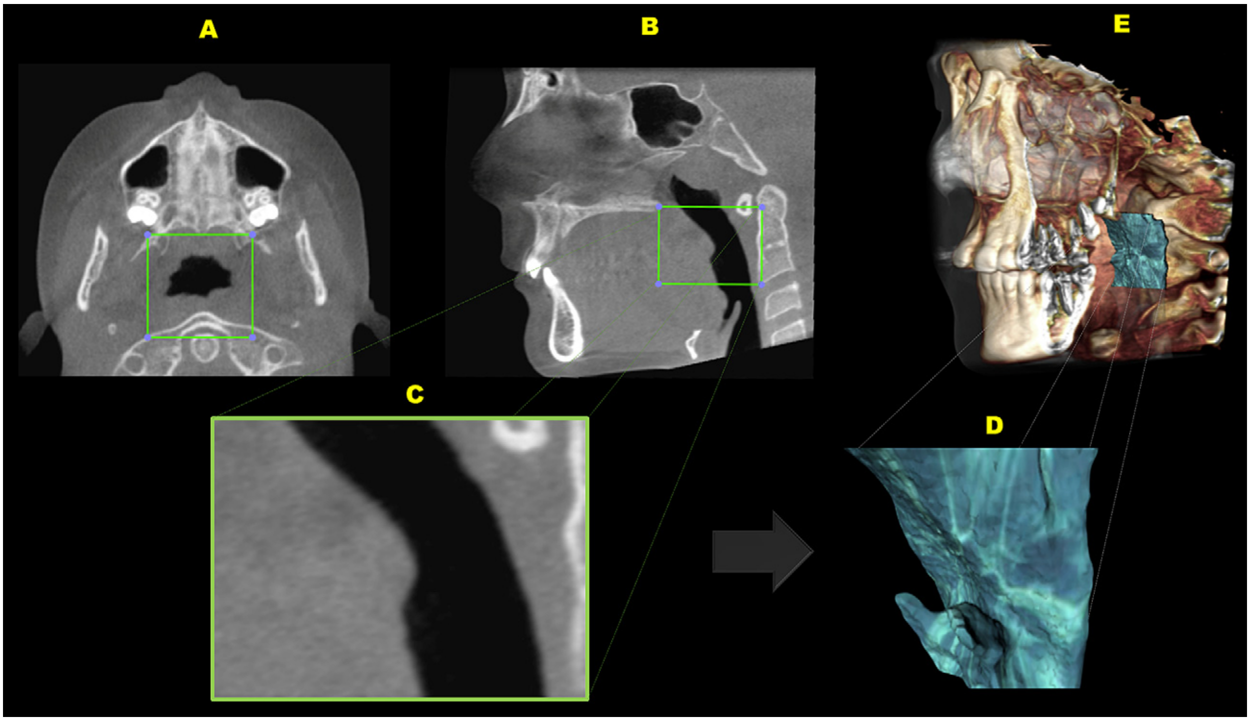

Fig 2. **A** and **B,** CBCT images after head reorientation with OsiriX software, in the axial and sagittal views, respectively, with the oropharynx's region of interest delimited (*green box*) in the axial and sagittal images; **C,** 3D volume exported as a new DICOM file; **D,** from this new volume, the segmentation was performed and the oropharynx's 3D model obtained for volume analysis; **E,** location of the oropharynx's region of interest in the patient's head.

fixed thresholding were InVivo Dental, Mimics, Ondemand3D, OsiriX, and ITK-Snap (except Dolphin3D).

Statistical analysis

The oropharynx and the oropharynx acrylic phantom segmentations and the volume analyses were performed by 1 investigator (A.W.). All measurements were made again 2 weeks later. The reliability of the first and second measurements was evaluated by intraclass correlation coefficient. A total of 680 segmentations were performed with the 6 imaging software programs. Values were imported into an Excel spreadsheet (Microsoft, Redmond, Wash).

The oropharynx acrylic phantom measurements with the 6 imaging software programs failed the Kolmogorov-Smirnov test for normality, so nonparametric tests were used. The oropharynx acrylic phantom segmentations were compared with the gold standard by using the Wilcoxon signed rank test for related samples. The oropharynx volume measurements of the 33 patients, with the 6 imaging software programs and the 2 threshold protocols (interactive and fixed), were normally distributed. Not all imaging software programs offered both interactive and fixed thresholds. A repeated measurements analysis of variance (ANOVA) test was used to compare the Mimics, Dolphin3D, Ondemand3D, OsiriX, and ITK-Snap software programs with the interactive threshold, and a second repeated measurements ANOVA test was used to compare the InVivo Dental, Mimics, Ondemand3D, OsiriX, and ITK-Snap software programs supporting a fixed threshold capability. The Mauchly's test of sphericity showed that the assumption of sphericity was violated; the Greenhouse-Geisser correction was used to complement the ANOVA test. Additionally, post-hoc tests with the Bonferroni correction were used to compare the volume measurements between 2 imaging software programs. Statistical significance was established at $\alpha = 0.05$. The statistical analysis was performed with SPSS software (version 12.0 for Windows; SPSS, Chicago, Ill).

RESULTS

The oropharynx acrylic phantom was used as the gold standard by comparing the actual physical volume with the segmentation volume in the CBCT images. The oropharynx acrylic phantom's physical volume was

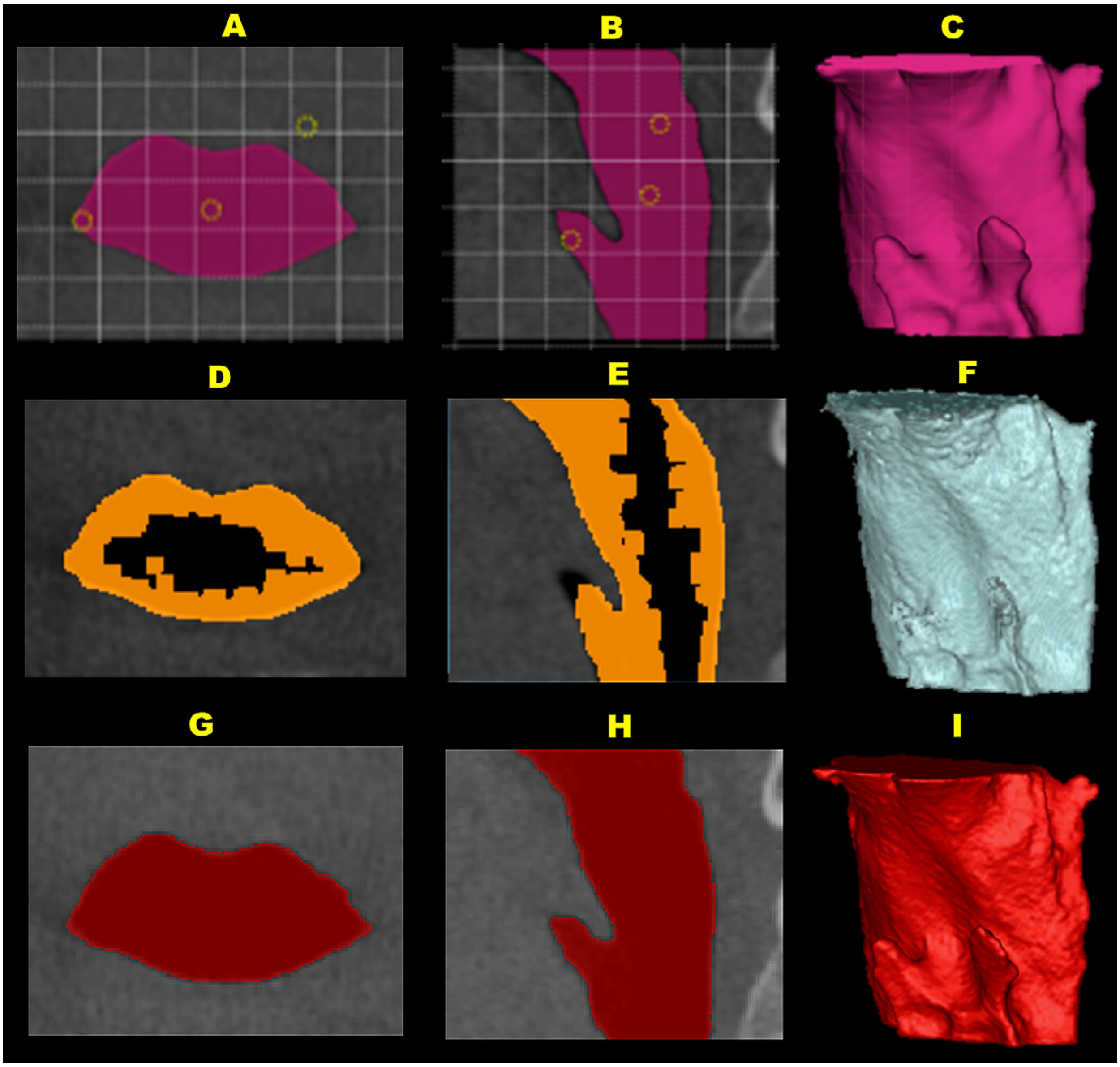

Fig 3. Examples of the oropharynx segmentation process with 3 imaging software programs: **A, D,** and **G,** axial views and **B, E,** and **H,** sagittal views of the oropharynx segmentation processes in the CBCT images; **C, F,** and **I,** 3D models of a patient's oropharynx obtained from segmentations. **A-C,** Segmentation process with Dolphin3D (*pink*) was based on addition and growth of seeds (*yellow circles*) that spread, matched, and filled the oropharynx airway. **D-F,** Segmentation with Ondemand3D (*orange*) showing, in the axial and sagittal views, empty spaces (*black*) without segmentation in some areas and overflowed in other regions. **G-I,** Segmentation with ITK-Snap (*red*). The segmentation matched precisely the oropharynx in the axial and sagittal views.

computed as 9.405 cm^3 with the formula: base area (235.13 mm^2) × height (40 mm). Additionally, the oropharynx acrylic phantom's volume was measured by the water-weight technique. Considering the weight measured (9.40 g) and the standard density of water at 20°C (1 cm^3 of water, 0.9982 g), the volume calculated by the rule of 3 (volume = 9.4/0.9982) was 9.410 cm^3. By using the average of the 2 methods, the oropharynx acrylic phantom's actual physical volume (gold standard) was 9.407 cm^3.

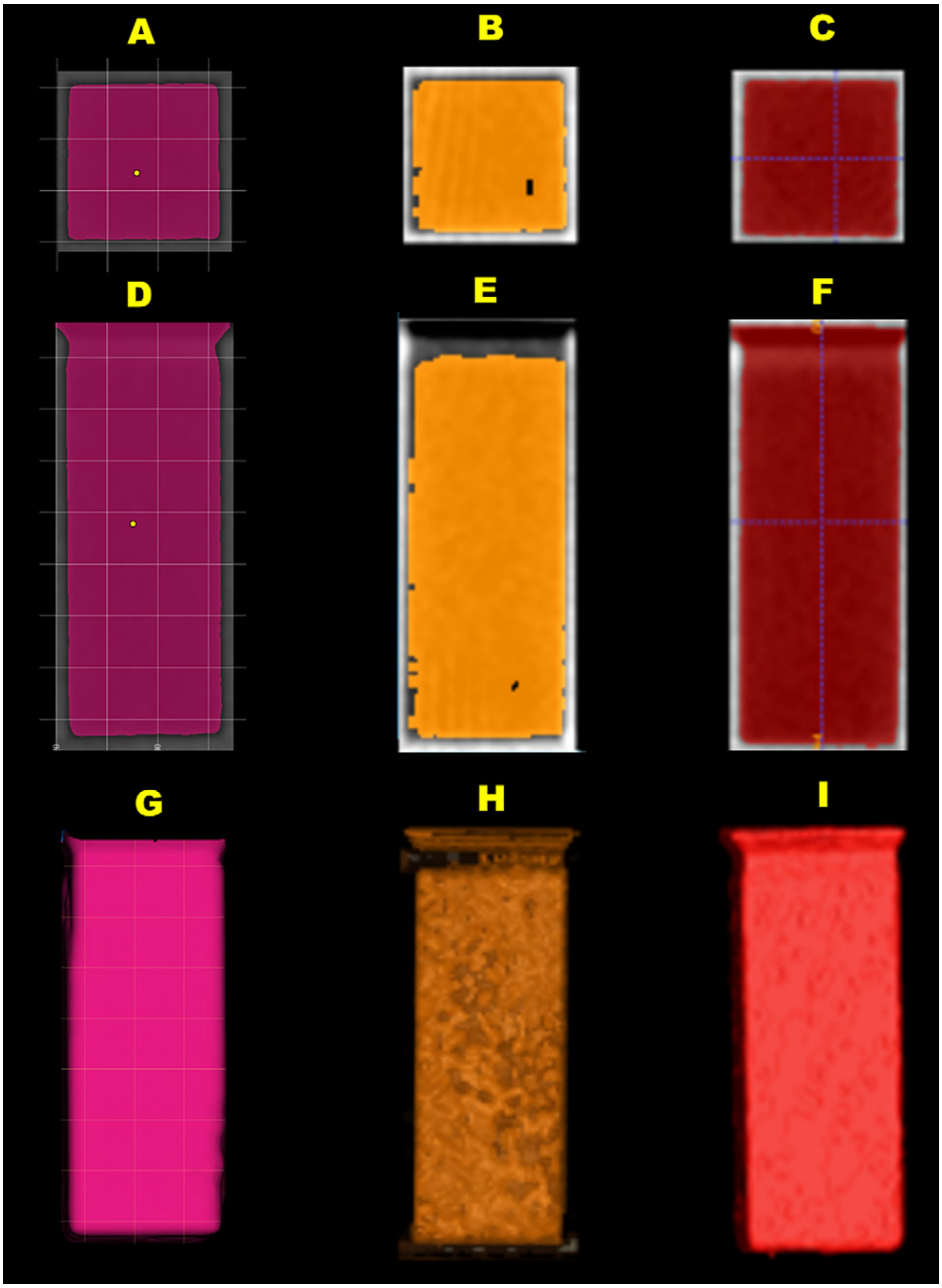

Fig 4. Examples of the oropharynx acrylic phantom segmentation process with 3 imaging software programs: **A**-**C,** axial views and **D-F,** sagittal views of the oropharynx acrylic phantom segmentation processes in CBCT images; **G-I,** 3D models of the oropharynx acrylic phantom obtained from segmentations. **A, D,** and **G,** Segmentation with Dolphin3D (*pink*). **B, E,** and **H,** Segmentation with Ondemand3D (*orange*) showing, in the axial and sagittal views, empty spaces (*black*) without segmentation in some areas and overflowed in other regions. **C, F,** and **I,** Segmentation with ITK-Snap (*red*). The segmentation matched precisely the oropharynx acrylic phantom in the axial and sagittal views.

Table I. CBCT imaging software compared in this study

Software	*Description*	*Operational system*	
Dolphin3D	Version 11.7; Dolphin Imaging & Management Solutions, Chatsworth, Calif	Windows	Not free
InVivo Dental	Version 5.0; Anatomage, San Jose, Calif	Windows	Not free
Ondemand3D	Version 1.0.9.1451; CyberMed, Seoul, Korea	Windows	Not free
Mimics	Version 14.12; Materialise, Leuven, Belgium	Windows	Not free
OsiriX	Version 4.0; Pixmeo, Geneva, Switzerland	Mac OS X	Free open source
ITK-Snap	Version 2.2.0; www.itksnap.org	Windows Mac OS X Linux	Free open source

Table II. Descriptive statistics of oropharynx acrylic phantom volume segmentation with the 6 imaging software programs

Software program	*Segmentation volume mean (mm^3)*	*Gold standard volume (mm^3)*	*Difference (mm^3)*	*Difference (%)*
Mimics	9392	9407	−15	−0.2
Dolphin3D	9315	9407	−92	−1.0
OsiriX	9289	9407	−118	−1.3
ITK-Snap	9236	9407	−172	−1.8
OnDemand3D	8809	9407	−598	−6.4
InVivo Dental	8366	9407	−1041	−11.1
Mimics FT	8340	9407	−1067	−11.3
OnDemand3D FT	8344	9407	−1063	−11.3
ITK-Snap FT	8328	9407	−1079	−11.5
OsiriX FT	8310	9407	−1097	−11.7

Programs designated *FT* (fixed threshold protocol) were used with a different segmentation protocol of −1000 to −587 grey levels.

The method repeatability for oropharynx acrylic phantom segmentations was excellent. The intraclass correlation coefficient between the first and second measurements (mean volume) was 0.999. The Wilcoxon signed rank test showed statistically significant differences ($P = 0.005$) between the medians of all 10 oropharynx acrylic phantom segmentation volumes and the gold standard. All 6 imaging software programs, with both interactive and fixed threshold protocols, underestimated the gold standard volume (Table II). When the interactive thresholding technique was used, Mimics, Dolphin3D, OsiriX, and ITK-Snap showed less than 2% error differences (underestimated) in volume calculation compared with the gold standard, whereas Ondemand3D and InVivo Dental showed more than a 5% errors compared with the gold standard (Table II). When the fixed threshold interval technique was used (−1000 to −587 grey levels), the software programs (InVivo Dental, Mimics, Ondemand3D, OsiriX, and ITK-Snap) had similar errors and underestimated the phantom volume by 11% compared with the gold standard (Table II).

The method repeatability for the patients' oropharynx measurements was high (ICC >0.94) for all imaging software programs (Table III). When oropharynx segmentations were performed with interactive thresholding, the volume measurements with the 6 imaging softwares were statistically different (repeated measures ANOVA with the Greenhouse-Geisser correction: F = 8.3; $P = 0.006$). The descriptive statistical analysis showed higher oropharynx mean volumes for ITK-Snap (7174 mm^3), Mimics (7163 mm^3), OsiriX (7086 mm^3), and Dolphin3D (7071 mm^3), and lower mean volumes for InVivo Dental (6661 mm^3) and Ondemand3D (6061 mm^3) (Table III). According to the Bonferroni post-hoc paired comparisons, there were no statistically significant differences (P >0.05) among ITK-Snap, Mimics, OsiriX, Dolphin3D, and Ondemand3D. However, the oropharynx volume segmentations with ITK-Snap, Mimics, OsiriX, and Dolphin3D were statistically significantly different from InVivo Dental. The comparison between InVivo Dental and OnDemand3D showed no significant difference (Table IV). When the fixed threshold technique was used, the ANOVA analysis with the Greenhouse-Geisser correction showed overall statistically significant differences (F = 30.7; P <0.001) in the oropharynx volume segmentation among all imaging software programs tested (InVivo Dental, Mimics, Ondemand3D, OsiriX, and ITK-Snap). The post-hoc paired comparisons also showed significant differences

Table III. Descriptive analysis of 33 patients' oropharynx volumes with the 6 imaging software programs

Software	Volume measurement 1 (mm^3)	Volume measurement 2 (mm^3)	ICC	Mean volume (mm^3)	SD (mm^3)	95% CI (lower and upper bounds)
ITK-Snap	7271	7077	0.99	7174	607	5938, 8409
Mimics	7206	7119	0.99	7163	594	5952, 8373
OsiriX	7093	7079	0.99	7086	593	5878, 7086
Dolphin3D	7121	7021	0.99	7071	585	5879, 8263
InVivo Dental	6679	6644	0.99	6661	587	5465, 7858
OnDemand3D	6198	5924	0.94	6061	349	5350, 6772
Mimics FT	6753	6753	1.00	6753	583	5566, 7940
ITK-Snap FT	6616	6627	1.00	6622	577	5445, 7797
OsiriX FT	6646	6645	1.00	6645	575	5475, 7817
OnDemand3D FT	6881	6881	1.00	6881	585	5688, 8075

Programs designated *FT* (fixed threshold protocol) were used with a different segmentation protocol of −1000 to −587 grey levels.
CI, Confidence interval; *ICC*, intraclass correlation coefficient.

for most software pairs, with the exception of InVivo Dental, which had no significant difference compared with ITK-Snap, Mimics, and OsiriX with the fixed threshold technique. Also, there was no significant difference between ITK-Snap and OsiriX (Table V).

DISCUSSION

Currently, several imaging software programs are available for upper airway assessment. A previous systematic review reported 18 imaging software programs for viewing, measuring, and analyzing the upper airway in CBCT images.[1] Only 1 study compared the accuracy and reliability of the imaging software programs for upper airway assessment in CBCT.[14] However, only Dolphin3D, InVivo Dental, and Ondemand3D software were compared; these programs were compatible only with the Windows operating system (Microsoft).[14] In our study, 6 imaging softwares were compared to test the accuracy of oropharynx segmentation (Table I). These programs are compatible not only with the Windows operating system, but also with Macintosh operating system X (Apple, Cupertino, Calif) and Linux operating system. A custom-made oropharynx acrylic phantom with a known volume was used as the gold standard. Dolphin3D, Ondemand3D, and InVivo Dental software were selected in this study because of their popularity among orthodontists and maxillofacial surgeons in the United States. Mimics software was chosen because of its wide use in biomedical engineering. OsiriX and ITK-Snap software were chosen because they are free, open-source softwares commonly used in medicine. Thus, the results of this study can be used as the basis for future airway studies with CBCT in both dentistry and medicine.

In this study, the method's design addressed some important aspects regarding the region of the upper airway analyzed. According to El and Palomo,[14] nasopharynx morphology is complex, and its volume measurement has less reliability than does oropharynx volume measurement. Because of this, to reduce the variability in measurements because of the structural complexity during software comparisons, only the oropharynx was evaluated. After head reorientation, the oropharynx was defined and exported as a new DICOM file maintaining the original 0.3-mm-voxel resolution. This image-processing step allows for standardized upper and lower oropharynx limits, eliminating variability introduced by using different imaging software programs to define the oropharyngeal airway. Furthermore, removing areas of noninterest created smaller files, reducing computing time.

The upper airway volume assessment depends on segmentation accuracy, image quality, and threshold interval selection. The CBCT image quality is impacted by several factors, such as the CBCT device's settings, patient positioning and management, volume reconstruction, and DICOM export.[22] When scanning is performed with high settings (small voxel size, longer scan time), the CBCT images are obtained with better spatial resolution.[23] In this study, both the oropharynx acrylic phantom and the patient's oropharynx were scanned with an i-CAT scanner, with high settings (0.3-mm voxel, 40-second scan time).[14] However, even with these settings, the patient's movement during the scan (40 seconds) might have produced motion-related artifacts, which could have some influence on the segmentation accuracy. Additionally, noise and artifacts also might have affected the contrast of the images. This could affect the discrimination between densities during the thresholding filter process and, consequently, the segmentation accuracy.

The upper airway segmentation and volume measurement are also influenced by the threshold interval selection. Interactive thresholding is a technique where

Table IV. Results of the post-hoc pairwise comparisons (with Bonferroni adjustment for multiple testing) for oropharynx segmentation with interactive thresholding

					95% CI for difference†	
Software (I)	*Software (J)*	*Mean difference (I – J) (mm^3)*	*SE (mm^3)*	*Significance†*	*Lower bound*	*Upper bound*
InVivo	ITK-Snap	−512*	53	.000	−682	−343
	Dolphin3D	−410*	51	.000	−573	−247
	OsiriX	−425*	46	.000	−570	−279
	Ondemand3D	600	364	1.000	−555	1755
	Mimics	−501*	43	.000	−638	−364
ITK-Snap	InVivo	512*	53	.000	343	682
	Dolphin3D	102	57	1.000	−78	283
	OsiriX	88	53	1.000	−82	258
	Ondemand3D	1113	378	.090	−86	2311
	Mimics	11	44	1.000	−128	151
Dolphin3D	InVivo	410*	51	.000	247	573
	ITK-Snap	−102	57	1.000	−283	78
	OsiriX	−15	46	1.000	−159	130
	Ondemand3D	1010	365	.139	−146	2167
	Mimics	−91	41	.517	−222	40
OsiriX	InVivo	425*	46	.000	279	570
	ITK-Snap	−88	53	1.000	−258	82
	Dolphin3D	15	46	1.000	−130	159
	Ondemand3D	1025	363	.122	−126	2176
	Mimics	−76	29	.183	−167	15
Ondemand3D	InVivo	−600	364	1.000	−1755	555
	ITK-Snap	−1113	378	.090	−2311	86
	Dolphin3D	−1010	365	.139	−2167	146
	OsiriX	−1025	363	.122	−2176	126
	Mimics	−1101	365	.075	−2259	56
Mimics	InVivo	501*	43	.000	364	638
	ITK-Snap	−11	44	1.000	−151	128
	Dolphin3D	91	41	.517	−40	222
	OsiriX	76	29	.183	−15	167
	Ondemand3D	1101	365	.075	−56	2259

Based on estimated marginal means. Because the mean difference could have positive or negative values, the letters *I* and *J* show that the mean difference (*I* – *J*) was obtained by subtracting the first-column program (*I*) from the second-column program (*J*).
CI, Confidence interval.
*Statistically significant at $P < 0.05$; †Bonferroni adjustment for multiple comparisons.

the threshold interval is determined by the operator's visual discrimination. When an interactive thresholding technique was used, although the reliability of the measurements was considered high, all imaging softwares produced different segmentations and volumes for both the oropharynx acrylic phantom and the patient's oropharynx. All programs underestimated the oropharynx acrylic phantom segmentation compared with the gold standard. These findings indicate that giving the operator freedom to determine the threshold interval based on visual analysis influences the upper airway segmentation accuracy. El and Palomo[14] demonstrated similar results, with Dolphin3D, InVivo Dental, and Ondemand3D software used to segment the upper airway. Also, they concluded that the airway segmentation using interactive thresholding showed high reliability but poor accuracy. This is because interactive thresholding is based on the operator's visual discrimination of the airway boundaries, and human vision is subject to a host of factors: eg, lighting conditions, fatigue, gray-scale ability, and visual acuity.[24] Fixed thresholding eliminates operator subjectivity in boundary selection. When fixed thresholding was used for oropharynx acrylic phantom segmentation, all imaging softwares produced similar segmentation and volume results. This further confirms that the operator's influence in threshold selection affects the final segmentation, and that the thresholding scale is compatible between the programs when a noncomplex structure, such as our oropharynx acrylic phantom, is segmented. However, when the oropharynx was segmented by using the same patient and the same threshold interval via the fixed threshold technique, most imaging software programs produced different segmentation and volume measurements. This could

Table V. Results of the post-hoc pairwise comparisons (with Bonferroni adjustment for multiple testing) for oropharynx segmentation with the fixed threshold interval, −1000 to −587 grey levels

Software (I)	Software (J)	Mean difference (I – J) (mm^3)	SE (mm^3)	Significance†	95% CI for difference†	
					Lower bound	Upper bound
Ondemand3D FT	ITK-Snap FT	260*	28	.000	175	345
	OsiriX FT	236*	28	.000	152	319
	InVivo	220*	37	.000	107	332
	Mimics FT	128*	23	.000	59	197
ITK-Snap FT	Ondemand3D FT	−260*	28	.000	−345	−175
	OsiriX FT	−24	11	.388	−58	10
	InVivo	−40	36	1.000	−149	68
	Mimics FT	−132*	14	.000	−173	−90
OsiriX FT	Ondemand3D FT	−236*	28	.000	−319	−152
	ITK-Snap FT	24	11	.388	−10	58
	InVivo	−16	35	1.000	−122	90
	Mimics FT	−108*	11	.000	−141	−75
InVivo	Ondemand3D FT	−220*	37	.000	−332	−107
	ITK-Snap FT	40	36	1.000	−68	149
	OsiriX FT	16	35	1.000	−90	122
	Mimics FT	−92	31	.060	−186	2
Mimics FT	Ondemand3D FT	−128*	23	.000	−197	−59
	ITK-Snap FT	132*	14	.000	90	173
	OsiriX FT	108*	11	.000	75	141
	InVivo	92	31	.060	−2	186

Based on estimated marginal means. Because the mean difference could have positive or negative values, the letters *I* and *J* show that the mean difference (*I* – *J*) was obtained by subtracting the first-column program (*I*) from the second-column program (*J*). Programs designated *FT* (fixed threshold protocol) were used with a different segmentation protocol of −1000 to −587 grey levels.
*The mean difference is significant at the .05 level; †Bonferroni adjustment for multiple comparisons.

be explained by the fact that, even with the oropharynx's less complex morphology compared with the nasopharynx, there are still small areas in the oropharynx where the software's segmentation algorithms differ in identifying, matching, and filling.

Although all programs underestimated the oropharynx acrylic phantom's physical volume, Mimics (0.2% error), Dolphin3D (1% error), OsiriX (1.3% error), and ITK-Snap (1.8% error) showed better accuracy than did Ondemand3D (6.4% error) and InVivo Dental (11.1% error) (Table II). Similar results were obtained for patients' oropharynx segmentations, in which the mean volumes with Mimics, Dolphin3D, OsiriX, and ITK-Snap were similar (Table III). Even though the Bonferroni post-hoc correction did not show significant differences between Ondemand3D and the other imaging software programs, both Ondemand3D and InVivo Dental underestimated the phantom's volume the most and produced the greatest mean differences between pairs of software measurements of the oropharynx (Tables III and IV). A possible explanation could be related to differences in the softwares' segmentation algorithms to identify, match, and fill the upper airway. With Ondemand3D software, both segmentation control (capability to allow a fine threshold interval adjustment) and sensitivity (capability to identify and segment all voxels with the grey levels compatible with air) were deficient, especially when small spaces of the oropharynx had to be filled and segmented. After segmentation, empty spaces were frequently observed in the axial, coronal, and sagittal slices, explaining the underestimated volumes (Fig 3, *D* and *E*). A similar problem with Ondemand3D was demonstrated in a previous study.[14] With InVivo Dental software, the user can only adjust the threshold interval in the 3D mode of view, not in the axial, coronal, or sagittal slices as traditionally found in all other software programs, so the user cannot accurately check the segmentation of the upper airway. Because of this program limitation, we used the standard threshold interval given by InVivo (−1000 to −587 grey levels), which seems to underestimate the oropharynx acrylic phantom's actual physical volume. Based on our findings, new updates and improvements in the airway modules for InVivo Dental and Ondemand3D are recommended.

Because of the limitation of the InVivo Dental software program, the standard threshold interval given by InVivo Dental (−1000 to −587 grey levels) was chosen as a control threshold interval. This control threshold interval was then applied to the fixed threshold protocol

Table VI. Main advantages and disadvantages of the 6 imaging softwares used in this study

Software	*Advantages*	*Disadvantages*
Dolphin3D	• User friendly • Quick upper airway segmentation • Good segmentation sensitivity (Figs 3, *A-C*; and 4, *A*, *D*, and *G*) • Segmentation can be checked in 2D slices (axial, coronal, and sagittal) • Minimal cross-sectional area analysis	• Not free • Lacks tools to adjust or correct segmentation in 2D slices (in spite of good segmentation sensitivity, in small regions with complex morphology, the segmentation algorithm fills empty space in some areas and overflows in other areas; this requires correction before volume rendering) • Threshold interval units (grey levels) not compatible with other imaging software
InVivo Dental	• User friendly • Quick upper airway segmentation • Threshold interval units (grey levels) compatible with other imaging software	• Not free • Threshold selection performed only in 3D mode of view (lack of parameter to check whether the segmentation matches the airway anatomy correctly in axial, coronal, and sagittal slices)
Ondemand3D	• Quick airway segmentation • Segmentation can be checked in 2D slices (axial, coronal, and sagittal) • Threshold interval units (grey levels) compatible with other imaging software	• Not free • Deficient segmentation control and sensitivity • Failed to render large areas of oropharynx, especially those with complex morphology (Figs 3, *D-F*; and 4, *B*, *E*, and *H*) • Lacks tools to adjust or correct segmentation in 2D slices
Mimics	• User friendly • Quick and easy airway segmentation • Best segmentation control and sensitivity • Segmentation can be checked in 2D slices • Tools to correct segmentation in 2D slices • Threshold interval units (grey levels) compatible with other imaging software	• Not free • Designed for biomedical engineering • Not as user friendly as Dolphin3D and InVivo
OsiriX	• Free and open source • Quick airway segmentation • Good segmentation sensitivity • Segmentation can be checked in 2D slices • Tools to correct segmentation in 2D slices • Threshold interval units (grey levels) compatible with other imaging software	• Designed for use in medicine • Not as user friendly as Dolphin3D and InVivo
ITK-Snap	• Free and open source • Good segmentation control and sensitivity (Figs 3, *G-I*; and 4, *C*, *F*, and *I*) • Segmentation can be checked in 2D slices • Tools to correct segmentation in 2D slices • Threshold interval units (grey levels) compatible with other imaging software	• Designed for use in medicine • Not as user friendly as Dolphin3D and InVivo

for the other programs. However, using this method resulted in underestimation of the oropharynx acrylic phantom for all softwares in our study. An improvement on our method would have been to adjust the threshold interval in InVivo Dental until the oropharynx acrylic phantom segmentation volume reached the actual physical volume of the oropharynx acrylic phantom, and then use this adjusted threshold interval as the control threshold interval for the other programs.

Although all 6 programs perform semiautomatic segmentations, they have different tools and ways to segment the upper airway. The main advantages and disadvantages of each imaging software program are described in Table VI.

In this study, we compared the precision and accuracy of 6 imaging software programs to segment and compute the volume of a phantom representing the oropharynx and the oropharynx airways of 33 patients. The influence of the threshold selection was addressed, and the advantages and disadvantages of each program were described, including some suggestions for future updates. Currently, there are only a few studies about the reliability and accuracy of imaging software programs for the upper airway.[1,10,14] Further research regarding CBCT imaging of the airway is needed to ensure the validity of the upper airway diagnosis.

CONCLUSIONS

1. All 6 imaging software programs were reliable in the volume segmentation of the oropharynx acrylic phantom but underestimated the values compared with the gold standard.
2. Mimics, Dolphin3D, ITK-Snap, and OsiriX values were similar and considerably more accurate than InVivo Dental and Ondemand3D for upper airway assessment.
3. The final segmentation volume depends on the threshold interval selection, the imaging software's segmentation algorithms, and the complexity of airway morphology.
4. Upper airway volume assessment depends on segmentation accuracy; the 6 imaging software programs used different segmentation engines. Contrary to popular belief, there is no established protocol or algorithm for processing DICOM images to produce an assessment of airway volume, and there are differences among the methods that are currently available commercially.

REFERENCES

1. Guijarro-Martinez R, Swennen GRJ. Cone-beam computerized tomography imaging and analysis of the upper airway: a systematic review of the literature. Int J Oral Maxillofac Surg 2011;40:1227-37.
2. Aboudara C, Nielsen I, Huang JC, Maki K, Miller AJ, Hatcher D. Comparison of airway space with conventional lateral headfilms and 3-dimensional reconstruction from cone-beam computed tomography. Am J Orthod Dentofacial Orthop 2009;135:468-79.
3. Aboudara CA, Hatcher D, Nielsen IL, Miller A. A three-dimensional evaluation of the upper airway in adolescents. Orthod Craniofac Res 2003;6:173-5.
4. Grauer D, Cevidanes LSH, Styner MA, Ackerman JL, Proffit WR. Pharyngeal airway volume and shape from cone-beam computed tomography: relationship to facial morphology. Am J Orthod Dentofacial Orthop 2009;136:805-14.
5. Iwasaki T, Hayasaki H, Takemoto Y, Kanomi R, Yamasaki Y. Oropharyngeal airway in children with Class III malocclusion evaluated by cone-beam computed tomography. Am J Orthod Dentofacial Orthop 2009;136:318.e1-9.
6. Kim YJ, Hong JS, Hwang YI, Park YH. Three-dimensional analysis of pharyngeal airway in preadolescent children with different anteroposterior skeletal patterns. Am J Orthod Dentofacial Orthop 2010;137:306.e1-11.
7. Lenza MG, Lenza MM, Dalstra M, Melsen B, Cattaneo PM. An analysis of different approaches to the assessment of upper airway morphology: a CBCT study. Orthod Craniofac Res 2010;13:96-105.
8. Ogawa T, Enciso R, Shintaku WH, Clark GT. Evaluation of cross-section airway configuration of obstructive sleep apnea. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2007;103:102-8.
9. Zhao Y, Nguyen M, Gohl E, Mah JK, Sameshima G, Enciso R. Oropharyngeal airway changes after rapid palatal expansion evaluated with cone-beam computed tomography. Am J Orthod Dentofacial Orthop 2010;137(Suppl):S71-8.
10. Schendel SA, Hatcher D. Automated 3-dimensional airway analysis from cone-beam computed tomography data. J Oral Maxillofac Surg 2010;68:696-701.
11. Tso HH, Lee JS, Huang JC, Maki K, Hatcher D, Miller AJ. Evaluation of the human airway using cone-beam computerized tomography. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2009;108:768-76.
12. Yamashina A, Tanimoto K, Sutthiprapaporn P, Hayakawa Y. The reliability of computed tomography (CT) values and dimensional measurements of the oropharyngeal region using cone beam CT: comparison with multidetector CT. Dentomaxillofac Radiol 2008;37:245-51.
13. Grauer D, Cevidanes LSH, Proffit WR. Working with DICOM craniofacial images. Am J Orthod Dentofacial Orthop 2009;136:460-70.
14. El H, Palomo JM. Measuring the airway in 3 dimensions: a reliability and accuracy study. Am J Orthod Dentofacial Orthop 2010;137:S50.e1-9.
15. Weissheimer A, Menezes LM, Mezomo M, Dias DM, Lima EMS, Rizzatto SMD. Immediate effects of rapid maxillary expansion with Haas-type and hyrax-type expanders: a randomized clinical trial. Am J Orthod Dentofacial Orthop 2011;140:366-76.
16. InVivo Dental 5.0 reference manual. San Jose, Calif: Anatomage; 2010.
17. Ondemand3D video manual: 3D segmentation. Avaiable at: http://www.ondemand3d.com/pages/resource/gallery/videos. Accessed November 10, 2011.
18. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116-28.
19. Mimics 13.0 reference guide. Ann Arbor, Mich: Materialise.

20. OsiriX User manual: the complete reference. Geneva, Switzerland: Pixeo.
21. Dolphin3D user's manual. Chatsworth, Calif: Dolphin Imaging & Management Solutions; 2006.
22. Liu Y, Olszewski R, Alexandroni ES, Enciso R, Xu T, Mah JK. The validity of in vivo tooth volume determinations from cone-beam computed tomography. Angle Orthod 2009;80:160-6.
23. Ballrick JW, Palomo JM, Ruch E, Amberman BD, Hans MG. Image distortion and spatial resolution of a commercially available cone-beam computed tomography machine. Am J Orthod Dentofacial Orthop 2008;134:573-82.
24. Mah P, Reeves TE, McDavid WD. Deriving Hounsfield units using grey levels in cone beam computed tomography. Dentomaxillofac Radiol 2010;39:323-35.