BJR

https://doi.org/10.1259/bjr.20170663

Received: 03 September 2017 | Revised: 30 January 2018 | Accepted: 01 February 2018

Cite this article as:
Andersen TN, Darvann TA, Murakami S, Larsen P, Senda Y, Bilde A, et al. Accuracy and precision of manual segmentation of the maxillary sinus in MR images—a method study. *Br J Radiol* 2018; **91**: 20170663.

FULL PAPER

Accuracy and precision of manual segmentation of the maxillary sinus in MR images—a method study

[1,2]TOBIAS N ANDERSEN, MS, [2,3]TRON A DARVANN, [2,4]SHUMEI MURAKAMI, [2,4]PER LARSEN, YURIE SENDA, [1]ANDERS BILDE, [1]CHRISTIAN V BUCHWALD and [2,4,5]SVEN KREIBORG

[1]Department of Otorhinolaryngology, Head and Neck Surgery and Audiology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark
[2]3D Craniofacial Image Research Laboratory (School of Dentistry, University of Copenhagen, Centre of Head and Orthopaedics, Copenhagen University Hospital Rigshospitalet; and DTU Compute, Technical University of Copenhagen), Copenhagen, Denmark
[3]Department of Oral and Maxillofacial Surgery, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark
[4]Department of Oral and Maxillofacial Radiology, Osaka University Graduate School of Dentistry, Osaka, Japan
[5]Department of Pediatric Dentistry and Clinical Genetics, School of Dentistry, University of Copenhagen, Copenhagen, Denmark

Address correspondence to: Mr Tobias N Andersen
E-mail: *tobiasnicolaiandersen@gmail.com*

Objective: To assess the accuracy and precision of segmentation of the maxillary sinus in MR images to evaluate the potential usefulness of this modality in longitudinal studies of sinus development.
Methods: A total of 15 healthy subjects who had been both craniofacial CT and MR scanned were included and the 30 maxillary sinus volumes were evaluated using segmentation. Two of the authors did segmentation of MRI and one of these authors did double segmentation. Agreement in results between CT and MRI as well as inter- and intraexaminer errors were evaluated by statistical and three-dimensional analysis.
Results: The intraclass correlation coefficient for volume measurements for both method error, inter- and intraexaminer agreement were > 0.9 [maximal 95% confidence interval of 0.989–0.997, $p < 0.001$] and the limit of agreement for all parameters were < 5.1%. Segmentation errors were quantified in terms of overlap [Dice Coefficient (*DICE*) > 0.9 = excellent agreement] and border distance [95% percentile Hausdorff Distance (*HD*) < 2 mm = acceptable agreement]. The results were replicable and not influenced by systematic errors.
Conclusion: We found a high accuracy and precision of manual segmentation of the maxillary sinus in MR images. The largest mean errors were found close to the orbit and the teeth.
Advances in knowledge: MRI can be used for 3D models of the paranasal sinuses with equally good results as CT and allows longitudinal follow-up of sinus development.

INTRODUCTION

The use of three-dimensional (3D) software to construct models for representation of body compartments has gradually gained ground and has become useful for diagnostics, surgical planning and simulation.[1–3] Furthermore, 3D models could become useful in following development of sinuses, which is of particular interest in patients with chronic rhinosinusitis (CRS). CRS may slow and alter pneumatisation[4] and is a major concern in individuals with Cystic Fibrosis[5,6] and Primary Ciliary Dyskinesia.[7,8] In general, the maxillary sinus is one of the most common sinuses for disease[9] and surgery, which is why it is of particular interest.

In understanding pathology, a complete knowledge of the normal development is necessary and is currently based on cross-sectional studies using CT,[10–13] cone beam CT (CBCT),[14–16] and a few using MRI.[17–19] Hence, longitudinal studies are clearly called for to elucidate the growth and development of the paranasal sinuses further. In principal, CT scans would be the best modality;[20] however, the increased cancer risk associated with consecutive CT scans during childhood[21] makes such a study unacceptable. Even though lower doses are seen with CBCT,[22] an extensive follow-up would imply a significant amount of radiation, and the Food and Drug Administration still recommend their use only when clinically necessary. An alternative solution would be to employ longitudinal MRI without using ionising radiation.

The primary purpose of the present study was to assess the accuracy and secondary the precision of segmentation of the maxillary sinus in MR images in order to evaluate the potential usefulness of this modality in longitudinal studies of normal individuals and patients with sinus disease.

METHODS AND MATERIALS

Subjects

15 subjects (10 females and 5 males; median age: 18 years, range: 14–20 years) without sinonasal disease or history of sinus surgery were included in the study. While one patient underwent MR 28 days after the CT scan, all the others had both scans done within the same day. The majorities of patients were diagnosed with osteoarthrosis in the temporomandibular joint and were referred to CT to evaluate bony changes and MR for articular disk condition. Other indications were trigeminal neuralgia, severe trismus, phlegmone (cellulitis) and mandibular protrusion. Since this is a retrospective study, subjects did not receive any additional radiation exposure.

Scanning procedure

All 15 subjects were scanned with a 1.5T MR scanner (neurovascular head and neck coil; GE Healthcare, Milwaukee, WI). We applied a 3D T_1 weighted image sequence. The imaging protocol consisted of sagittal fast gradient echo sequences (TR/TE/FA/NEX, 10.5 ms/4.8 ms/15 degree/1) 280 × 280 × 224 mm FOV; 512 × 512 × 320 matrix size (voxel size was 0.55 × 0.55 × 0.7 mm).

CT examination was performed by a 64-row multidetector CT scanner (Light Speed VCT; GE Healthcare, Milwaukee, WI) and applied helical (spiral) mode. A dose reduction algorithm [3D mA Modulation (Smart mA), GE] was applied. CT images were taken at 120 kVp and 200 mA. Axial images were obtained with the X-ray beam parallel to the occlusal plane or mandibular plane and displayed with bone window [WW/WL: 4,000/800 Hounsfield units (HU)]. The field of view was 25 × 25 cm. The matrix size was 512 × 512 (voxel size was 0.49 × 0.49 × 0.625 mm). No enhancement methods were used in either CT or MR.

Image Registration

For each image pair, the CT scan was oriented according to the Frankfort horizontal plane and the mid-sagittal plane by manually placing four anatomical landmarks (the two *orbits*, the deepest point on the infraorbital rim, and the two *porions*, the highest point on the upper margin of the external acoustic meatus bilaterally). Subsequently, the MR image was registered to the CT using a voxel-based rigid image registration algorithm (Image Registration ToolKit, IRTK, IXICO PLC, London, UK) applying normalised mutual information as a similarity measure.[23] Thereafter, fast flickering between the CT and MR images was done for quality control.

Segmentation

The aligned scans were imported into Analyze™ (v. 12.0, Minneapolis, MN). Segmentation of the CT images was carried out with "Auto Trace", a two-dimensional seeded region growing and level-set method, using a seed pixel and predefined thresholds of −1024 to −318 HU. Every slice in the transversal plane was reviewed and any errors were manually corrected. Segmentation of MRI was manual using the "Spline Edit" tool, a curve interpolated between points manually set by the observer for each slice. In the context of this work, the segmentation performed on CT is considered as the gold standard (GS). While only one author did segmentation of CT, two authors did segmentation of MRI and one of these did double segmentation with a minimum of 1-month interval between the two. Observer 1 (O1) and observer 2 (O2) were calibrated in terms of segmentation procedure and any uncertainty of the general delineation was discussed with an ENT doctor and an oral maxillofacial radiologist before the study was initiated. Thereafter, both observers carried out a pilot segmentation on three subjects each, which were the last ones to be segmented with a minimum of 1.5-month interval from the training period.

Analysis

An error analysis was carried out based on segmentation of left and right sinuses from 15 individuals, resulting in a total sample size of 30.

Volume error

The volume of each sinus was calculated by multiplying the number of voxels with the voxel dimensions. The intraclass correlation coefficient (ICC) was the measure of agreement for volume comparisons. A modified Bland-Altman approach was used to test for random and systematic errors by plotting the relative difference in volume of the sinuses segmented in CT and MRI.[24]

Magnitude and spatial distribution of error

An investigation of the magnitude of local errors was carried out by comparing pairs of segmented sinus surfaces from the same individual.

Six types of comparisons were carried out:

(1) O1A*vs*GS: comparison of observer 1's first segmentation (A) to the gold standard
(2) O1B*vs*GS: comparison of observer 1's second segmentation (B) to the gold standard
(3) O2*vs*GS: comparison of observer 2's segmentation to the gold standard
(4) O1A*vs*O1B: intraexaminer comparison between observer 1's two segmentations
(5) O2*vs*O1A: interexaminer comparison of observer 2's segmentation to observer 1's first segmentation
(6) O2*vs*O1B: comparison of observer 2's segmentation to observer 1's second segmentation

Dice Coefficient (*DICE*) and Hausdorff Distance (*HD*)[25] were used as comparison metrics. *DICE* provides a measure of the amount of overlap between two segmentations and ranges from 0 (no overlap) to 1 (full overlap). *HD* represents the maximum distance between corresponding segmentation boundary points. A 95% Hausdorff percentile (*HD95*) was also calculated, representing the maximum distance when the largest 5% of distances were discarded (this is the default setting in the Insight Toolkit).[26] Segmentation boundaries were formed as polygonal surfaces using Marching Cubes.[27] Corresponding points on two surfaces were determined as the distance d_1 from a point in one (source) surface to the nearest location in the other (target) surface. Another set of distances, d_2, were also calculated by switching the role of source and target surfaces, providing, in general, different values than d_1. *HD* is therefore, in practice, calculated as the

maximum of all distances d_1 *and* d_2. Histogram distributions of distances were also plotted, as well as mean histogram distributions (one for each of the six types of comparison).

An investigation of the spatial distribution of the local differences ("errors") (d_1 or d_2) was carried out by color coding the source surface by the distance value, providing a distance or "error" map *E*. In the plotted distributions and color coded surfaces the distances were given a sign depending on whether the source surface was inside or outside of the target surface. A mean signed distance ($\bar{D}$) different from zero thus indicates a systematic difference between a pair of segmentations. Mean error maps $\bar{E}$, averaging the error maps from all individuals were also constructed, but it required establishment of detailed point correspondence between surfaces from different individuals; this was achieved using a method similar to the method by Szeliski et al[28].

Statistics

Calculation of ICC was done in SPSS (v. 23) based on inter- and intraobservations with a fixed number of two examiners. The two-way mixed procedure was used and results for average measures were reported. Bland Altman plots were calculated and illustrated using RStudio (v. 0.99.879). Power analysis was based on a paired t-test and found to be 0.93 in detecting a 2% difference between CT and MRI with a two-sided significance level of 0.05. Normality was checked with a Shapiro-Wilk test and verified by plotting the residuals in a QQ-plot.

RESULTS

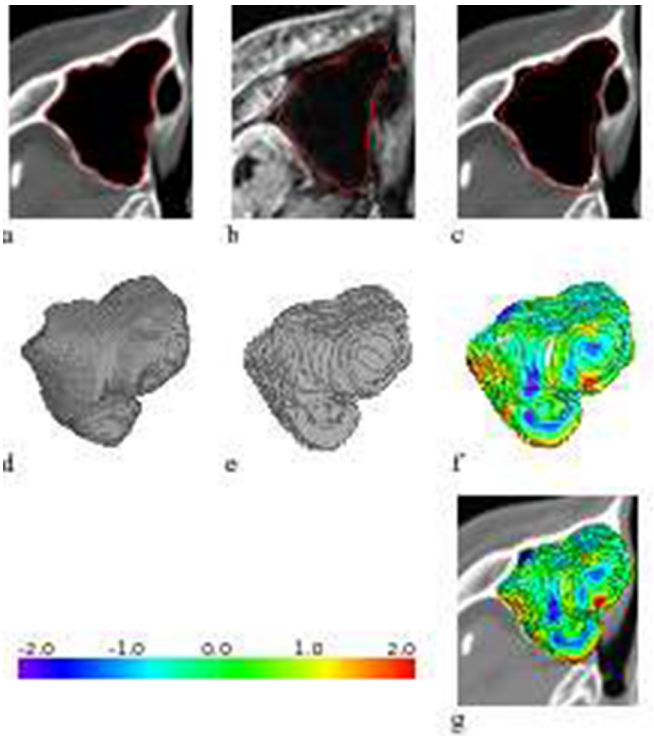
Registration

The visual quality control of registration results with fast flickering between CT and MRI showed good agreement of the maxillary borders and related anatomical structures (Figure 1).

Segmentation

Figure 2 shows an example of segmentation of a maxillary sinus. It may be seen how tracing of the sinus' border differs between MR and CT in some locations. Figure 2 also shows corresponding 3D surface reconstructions, indicating that similar 3D shapes are obtained in the two different modalities. An example error map *E* for comparison type O1A*vs*GS is also shown, indicating local errors up to about ± 2 mm. Segmentation of a single sinus took between 1 and 3 h depending on the shape and size of the sinus. Overall, O2 used considerably less time on segmentation than O1.

Figure 2. Example of segmentation. Axial slices. (a) Semi-automatic delineation (red) on CT. (b) Manual delineation (red) on MR. (c) Manual delineation (red) on MR shown on corresponding CT slice. (d) CT surface reconstruction from semi-automatic segmentation. (e) MR surface reconstruction from manual segmentation. (f) The MR surface reconstruction colour coded according to an error map *E* providing an accuracy of MR surface reconstruction in terms of calculated closest point distance in mm between the surfaces in d) and e). (g) Same as f) but shown with CT image slice for spatial context.



Volume reliability

The mean volume of assessed sinuses was 21.5 cm^3 (range 12 to 28 cm^3). In accordance with the classification by Landis and Koch[29], the intraexaminer and interexaminer ICC (Table 1) indicate an excellent accuracy and precision (ICC range 0.995–0.998, $p < 0.001$). The limits of agreement were less than 5%. The Bland-Altman plots (Figure 3) did not

Figure 1. Example of images included in the study. (a) Axial slice of CT scan. (b) Axial slice of MR scan shown after voxel-based rigid registration to the CT scan. (c) Fusion with 50% CT and 50% MR after registration.

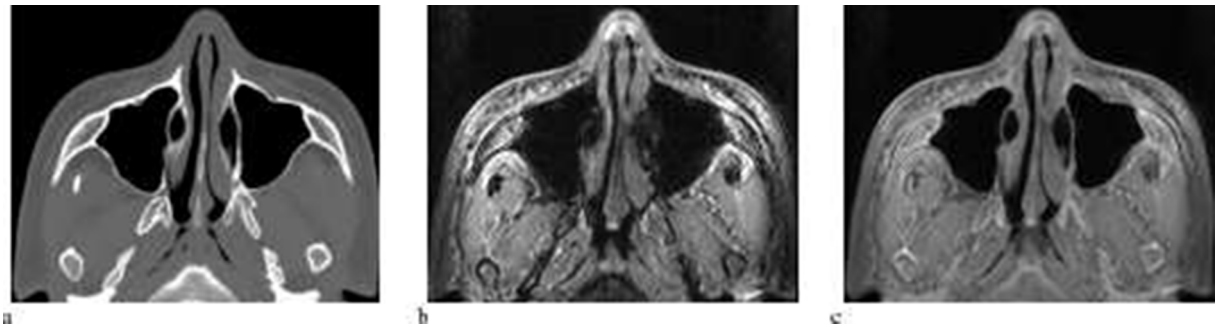


Table 1. Method, intraexaminer and interexaminer reliability for volume measures of the maxillary sinus using CT and MR segmentation (n = 30).

Comparison	Intraclass correlation coefficient	95% CI	Mean Diff [%]	SD [%]	Limit of agreement (Mean Diff. ± 1.96 SD) [%]
O1AvsGS	0.998	0.996–0.999	0.20	1.6	−2.9 to 3.3
O1BvsGS	0.996	0.992–0.998	0.72	2.0	−3.2 to 4.7
O2vsGS	0.997	0.994–0.999	−0.06	1.9	−3.9 to 3.8
O1vsGSa	0.998	0.995–0.999	0.46	1.6	−2.6 to 3.5
O1AvsO1B	0.998	0.995–0.999	0.52	1.8	−3.0 to 4.0
O1AvsO2A	0.995	0.989–0.997	- 0.26	2.46	−5.1 to 4.6

CI, confidence interval; GS, gold standard; Mean Diff, mean difference; O1, observer 1; O1A, first segmentation by observer 1; O1B, second segmentation by observer 1; O2, observer 2; O2A, first segmentation by observer 2; SD, standard deviation.
[a]For O1, a mean of volumes for O1A and O1B was used to compare with GS.

show any evidence of systematic errors for any of the observers.

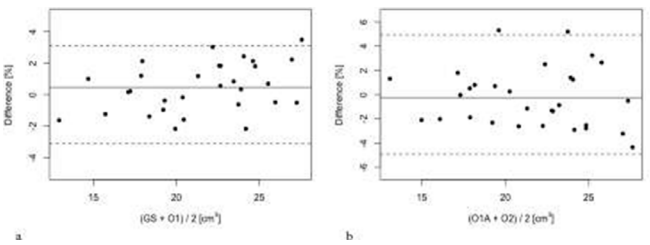
Comparisons

Table 2 presents results of the six comparisons in terms of Dice Coefficient (*DICE*), Hausdorff Distance (*HD*), 95% percentile Hausdorff Distance (*HD95*) and mean signed distance ($\bar{D}$). Minimum *DICE* values are >0.9 indicating excellent overlap between segmentations for all six comparisons. The maximum HD was large (about 8 mm), while the maximum *HD95* was much smaller (<1.7 mm) and smaller for observer 1 than for observer 2. The mean signed distance was very close to zero, indicating no systematic differences.

Figure 4 presents mean of the distance distributions for the four comparisons O1A *vs* O1B, O2 *vs* O1A, O1A *vs* GS and O2 *vs* GS (see the section on "Magnitude and spatial distribution of error" for explanation). It may be seen that the values are approximately normally distributed and with mean values close to 0; thus no systematic errors seem to be present. It is also evident that the narrowest distribution is the one describing O1A *vs* O1B, followed by O2 *vs* O1A, O1A *vs* GS and finally O2 *vs* GS.

Mean error maps $\bar{E}$ for the comparisons O1A *vs* GS are shown in Figure 5. In addition, the comparisons O2 *vs* GS, O1A *vs* O1B and O2 *vs* O1A can be seen in the supplemental material (Supplementary Figure 1). Regions with the largest mean error were seen especially at the superior and inferior parts of the maxillary sinus, and also at the medial part to some extent.

Figure 3. Bland-Altman plots of precision of volume estimate. GS, gold standard (CT); O1A, first segmentation by observer 1; O1, mean of two segmentations (O1A and O1B, see the text) by observer 1; O1B, second segmentation by observer 1; O2, segmentation by observer 2. Solid line: mean difference between GS- and MR-volume. Dotted lines: two standard deviations of difference between GS- and MR-volume.



DISCUSSION

Although 3D models based on MRI would seem to be potentially useful in examining the development of the paranasal sinuses in children and adolescents because of no ionising radiation exposure, nearly all 3D studies of the sinuses to date are based on CT or CBCT data and accuracy or precision of MR segmentation have not been analysed so far.

The present study showed high accuracy and precision of segmentation of the maxillary sinus in MRI compared with CT. The ICC for volume-measurements indicated excellent approximation of volume. In 95% of the cases, it was possible to predict the sinus volume with a maximal error of 5% using MRI. The large *DICE* values found (>0.9) confirmed this, indicating that accuracy as well as inter- and intraobserver error were small. The high HD values found reflected the presence of noise in the segmentation border. HD is a measure of agreement that is known to be sensitive to outliers. Using *HD95* (95th quantile of distances) revealed that 95% of the distances were below 2.5 mm. In comparison, the average maxillary sinus develops approximately 1 cm in all dimensions and triples in volume from 3 to 18 years of age.[17] In general, O1 had higher accuracy compared to O2, which may be explained by considerably more time spent on segmentation by O1. In addition, we also presented analysis of spatial localisation of deviations and found the largest mean errors to be located in the superior and inferior part of the sinus; areas that are in close relation to the orbit and the teeth respectively. This is most likely because of the difficulty to distinguish the paper-thin bone of the orbital floor from air in the sinus, while the teeth constitute an uneven floor of the maxillary sinus and their placement relative to the maxilla varies between individuals.

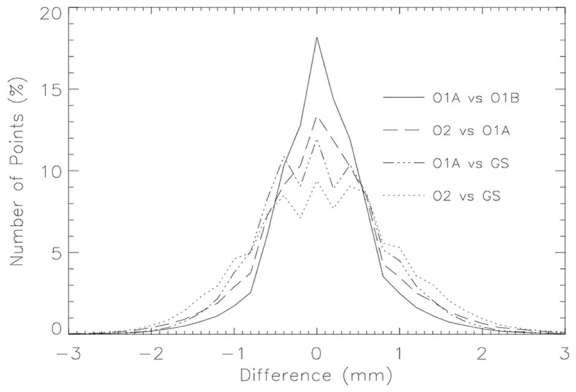
Both left and right sinuses were included for each of the 15 subjects. Therefore, the method error has contributions from both intersubject and left-right intrasubject variability. This approach is sensible since the segmentation procedure should be carried out on both left and right sinuses.

Table 2. Comparison metrics

Comparison	$\bar{D}$ (mm)		*DICE*			*HD* (mm)			*HD95* (mm)		
		Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
O1AvsGS	0.06	0.93	0.92	0.94	3.64	2.37	5.49	1.43	1.25	1.91
O1BvsGS	0.04	0.93	0.91	0.94	3.37	2.50	5.76	1.44	1.24	1.77
O2vsGS	0.07	0.92	0.90	0.93	4.37	2.61	7.94	1.72	1.25	2.45
O1AvsO1B	0.10	0.94	0.93	0.96	3.33	2.31	4.83	1.25	0.92	1.71
O2vsO1A	0.09	0.92	0.91	0.94	4.33	2.63	8.09	1.62	1.25	2.09
O2vsO1B	0.11	0.92	0.91	0.94	4.39	2.72	8.00	1.67	1.31	2.27

GS, gold standard; O1A, first segmentation by observer 1; O1B, second segmentation by observer 1; O2, observer 2.
Values of $\bar{D}$ (mean signed distance), mean and range (Min and Max) of Dice Coefficient (*DICE*), Hausdorff Distance (*HD*) and 95% Hausdorff percentile (*HD95*) for six comparisons as indicated.
Three uppermost rows pertain to measurement of accuracy; three lowermost to measurement of precision

Figure 4. Distribution of error magnitude. Mean histograms $\bar{H}$ shown for four types of comparisons as indicated. GS, gold standard; O1A, first segmentation by observer 1; O1B, second segmentation by observer 1; O2, observer 2.

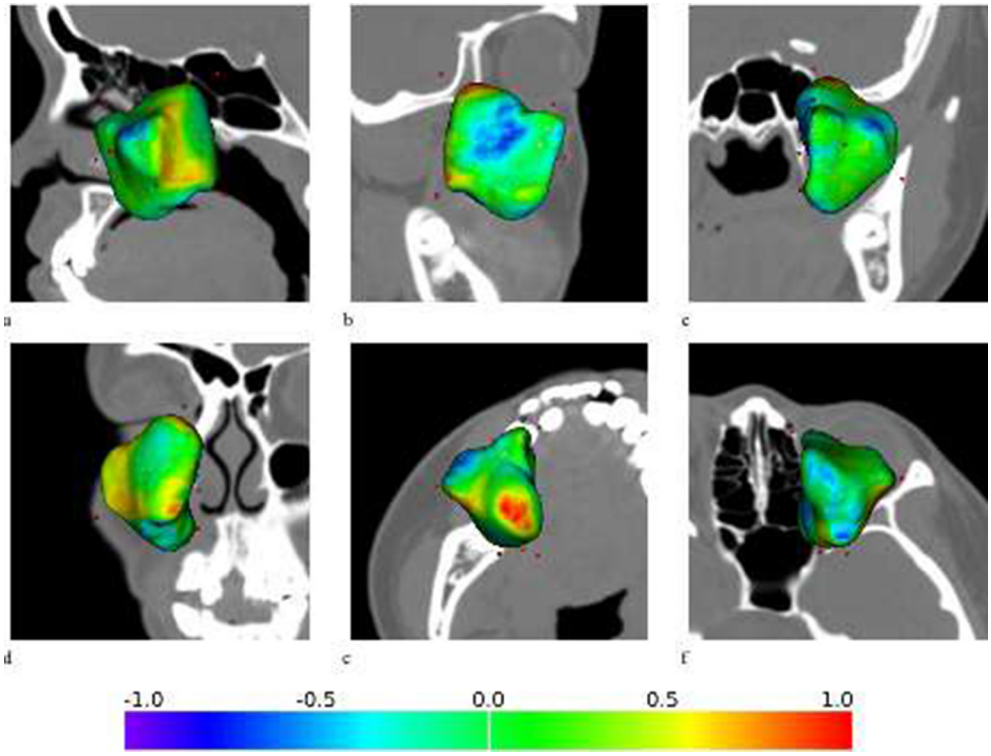


Some previous method studies in the field have used manual segmentation by a single observant as GS,[30,31] but this has shown to have greater than expected intervariability.[32] Furthermore, the precision of manual and semi-automatic segmentation on CT images of the maxillary sinus is previously shown to not be significantly different.[33] The choice of semi-automatic segmentation as GS in this study was based on these considerations as well as the authors' experience with segmentation based on predefined and fixed HU as more reproducible than manual segmentation, since very few or in some cases even no corrections are made by the observer. Nevertheless, this has not been investigated in a study and should therefore be mentioned as a limitation. Another limitation is the limited number of two observers with only one of these doing double segmentation. Therefore, the reader should interpret the results of repeatability in light of this.

Due to ethical considerations, CT data on healthy children is sparse which is why the minimum age in our study was 14 years of age. This may leave a limitation since it can be argued that this study cannot be generalised to children. On the other hand, we included patients with various sinus volumes (range 12–28 cm^3) and shapes. Furthermore, our study finds that the job of distinguishing paper-thin bone from tissue and air can be overcome with excellent results. This issue does not differ in a paediatric population and the authors argue that the method can be applied in younger individuals with reasonable results. After all, manual segmentation is probably the best approximation of sinus volume based on MRI compared to far simpler stereological methods used in previous MRI studies.[18] Because of its time requirements, however, this method is only useful in research contexts and CT substitutes derived from MRI data could be an interesting and more time efficient alternative.[34]

Previous MRI studies on paranasal sinuses have refused using segmentation for volume assessments[17–19] since the delineation of bony structures on MRI at first seems difficult and imprecise. These results have a particular interest in patients with CRS, but may also be useful in other fields such as obstructive sleep apnoea, where volumetric assessments from MRI have shown to be useful in diagnostics.[35] Finally, this allows us to use manual segmentation for a future project with longitudinal MRI data and

Figure 5. Mean error map $\bar{E}$ for the comparison O1A (a) *vs* GS (b). The error maps are shown on a mean maxillary sinus surface (mean of 15 surfaces) and with CT slices shown behind the surface for indication of orientation. Red dots represent landmarks used for alignment. Lateral (a), medial (b), anterior (c), posterior (d), superior (e) and inferior (f) view. Difference between the compared error maps is shown in mm (range −1 to 1). Negative (bluish) values indicate that the compared segmentation (a) is smaller than the other (b) (a is inside b surface), while positive (red dish) colours indicate that the compared segmentation (a) is larger than the other (b) (a is outside b surface). GS, gold standard; O1A, first segmentation by observer 1.



has a great interest for otolaryngologists in understanding pathological sinus development.

CONCLUSION

Our results show high accuracy and precision of segmentation of the maxillary sinus in MR images. While maximal errors were found to be relatively large, there were only a few of them. The mean errors were small and not of clinical relevance. The largest mean errors were seen in relation to the orbit and the teeth.

ACKNOWLEDGEMENT

We thank Professor Bjarne Kjær Ersbøll, Department of Applied Mathematics and Computer Science, Technical University of Denmark, for statistical advice.

CONSENT

Written and informed consent was achieved from each subject, and the "Osaka University Ethical Committee" approved the study (application no. "H21-E16", "Improvement of accuracy of image diagnosis using CT and MRI").

REFERENCES

1. Olson G, Citardi MJ. Image-guided functional endoscopic sinus surgery. *Otolaryngol Head Neck Surg* 2000; **123**: 188–94. doi: https://doi.org/10.1067/mhn.2000.107453
2. Sorensen MS, Mosegaard J, Trier P. The visible ear simulator: a public PC application for GPU-accelerated haptic 3D simulation of ear surgery based on the visible ear data. *Otol Neurotol* 2009; **30**: 484–7. doi: https://doi.org/10.1097/MAO.0b013e3181a5299b
3. Luka B, Brechtelsbauer D, Gellrich NC, König M. 2D and 3D CT reconstructions of the facial skeleton: an unnecessary option or a diagnostic pearl? *Int J Oral Maxillofac Surg* 1995; **24**: 76–83. doi: https://doi.org/10.1016/S0901-5027(05)80866-5
4. Milczuk HA, Dalley RW, Wessbacher FW, Richardson MA. Nasal and paranasal sinus anomalies in children with chronic sinusitis. *Laryngoscope* 1993; **103**: 247–52. doi: https://doi.org/10.1288/00005537-199303000-00002
5. Eggesbø HB, Eken T, Eiklid K, Kolmannskog F. Hypoplasia of the sphenoid sinuses as a diagnostic tool in cystic fibrosis. *Acta Radiol* 1999; **40**: 479–85. doi: https://doi.org/10.3109/02841859909175571
6. Kim HJ, Friedman EM, Sulek M, Duncan NO, McCluggage C. Paranasal sinus development in chronic sinusitis, cystic fibrosis, and normal comparison population: a computerized tomography

correlation study. *Am J Rhinol* 1997; **11**: 275–81. doi: https://doi.org/10.2500/105065897781446676
7. Pifferi M, Bush A, Caramella D, Di Cicco M, Zangani M, Chinellato I, et al. Agenesis of paranasal sinuses and nasal nitric oxide in primary ciliary dyskinesia. *Eur Respir J* 2011; **37**: 566–71. doi: https://doi.org/10.1183/09031936.00068810
8. Sommer JU, Schäfer K, Omran H, Olbrich H, Wallmeier J, Blum A, et al. ENT manifestations in patients with primary ciliary dyskinesia: prevalence and significance of otorhinolaryngologic co-morbidities. *Eur Arch Otorhinolaryngol* 2011; **268**: 383–8. doi: https://doi.org/10.1007/s00405-010-1341-9
9. Diament MJ, Senac MO, Gilsanz V, Baker S, Gillespie T, Larsson S. Prevalence of incidental paranasal sinuses opacification in pediatric patients: a CT study. *J Comput Assist Tomogr* 1987; **11**: 426–31. doi: https://doi.org/10.1097/00004728-198705000-00011
10. Apuhan T, Yıldırım YS, Özaslan H. The developmental relation between adenoid tissue and paranasal sinus volumes in 3-dimensional computed tomography assessment. *Otolaryngol Head Neck Surg* 2011; **144**: 964–71. doi: https://doi.org/10.1177/0194599811399712
11. Kim J, Song SW, Cho JH, Chang KH, Jun BC. Comparative study of the pneumatization of the mastoid air cells and paranasal sinuses using three-dimensional reconstruction of computed tomography scans. *Surg Radiol Anat* 2010; **32**: 593–9. doi: https://doi.org/10.1007/s00276-009-0618-4
12. Spaeth J, Krügelstein U, Schlöndorff G. The paranasal sinuses in CT-imaging: development from birth to age 25. *Int J Pediatr Otorhinolaryngol* 1997; **39**: 25–40. doi: https://doi.org/10.1016/S0165-5876(96)01458-9
13. Park IH, Song JS, Choi H, Kim TH, Hoon S, Lee SH, et al. Volumetric study in the development of paranasal sinuses by CT imaging in Asian: a pilot study. *Int J Pediatr Otorhinolaryngol* 2010; **74**: 1347–50. doi: https://doi.org/10.1016/j.ijporl.2010.08.018
14. Asaumi R, Sato I, Miwa Y, Imura K, Sunohara M, Kawai T, et al. Understanding the formation of maxillary sinus in Japanese human foetuses using cone beam CT. *Surg Radiol Anat* 2010; **32**: 745–51. doi: https://doi.org/10.1007/s00276-010-0678-5
15. Tolstunov L, Thai D, Arellano L. Implant-guided volumetric analysis of edentulous maxillary bone with cone-beam computerized tomography scan. Maxillary sinus pneumatization classification. *J Oral Implantol* 2012; **38**: 377–90. doi: https://doi.org/10.1563/AAID-JOI-D-11-00212
16. Alsufyani NA, Flores-Mir C, Major PW. Three-dimensional segmentation of the upper airway using cone beam CT: a systematic review. *Dentomaxillofac Radiol* 2012; **41**: 276–84. doi: https://doi.org/10.1259/dmfr/79433138
17. Adibelli ZH, Songu M, Adibelli H. Paranasal sinus development in children: a magnetic resonance imaging analysis. *Am J Rhinol Allergy* 2011; **25**: 30–5. doi: https://doi.org/10.2500/ajra.2011.25.3552
18. Barghouth G, Prior JO, Lepori D, Duvoisin B, Schnyder P, Gudinchet F. Paranasal sinuses in children: size evaluation of maxillary, sphenoid, and frontal sinuses by magnetic resonance imaging and proposal of volume index percentile curves. *Eur Radiol* 2002; **12**: 1451–8. doi: https://doi.org/10.1007/s00330-001-1218-9
19. Yan YC, Wu SC, Yuan XY, Gu QL, Bai ZH, Guo HW. Preliminary study on normal aerification of paranasal sinuses in children. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 2011; **46**: 650–3.
20. Kirmeier R, Arnetzl C, Robl T, Payer M, Lorenzoni M, Jakse N. Reproducibility of volumetric measurements on maxillary sinuses. *Int J Oral Maxillofac Surg* 2011; **40**: 195–9. doi: https://doi.org/10.1016/j.ijom.2010.10.008
21. Brenner D, Elliston C, Hall E, Berdon W. Estimated risks of radiation-induced fatal cancer from pediatric CT. *AJR Am J Roentgenol* 2001; **176**: 289–96. doi: https://doi.org/10.2214/ajr.176.2.1760289
22. Ludlow JB, Ivanovic M. Comparative dosimetry of dental CBCT devices and 64-slice CT for oral and maxillofacial radiology. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2008; **106**: 106–14. doi: https://doi.org/10.1016/j.tripleo.2008.03.018
23. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit* 1999; **32**: 71–86. doi: https://doi.org/10.1016/S0031-3203(98)00091-0
24. Pollock MA, Jefferson SG, Kane JW, Lomax K, MacKinnon G, Winnard CB. Method comparison-a different approach. *Ann Clin Biochem* 1992; **29 (Pt 5**: 556–60. doi: https://doi.org/10.1177/000456329202900512
25. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015; **15**: 29. doi: https://doi.org/10.1186/s12880-015-0068-x
26. Johnson HJ, McCormick MM, Ibáñez L. The ITK software guide. In: *Book 2: Design and Functionality*. **Fourth ed**: Kitware inc.; 2015.
27. William E, Lorensen HEC. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics* 1987; **21**: 163–9.
28. Szeliski R, Lavallée Stéphane, Lavallée S. Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines. *Int J Comput Vis* 1996; **18**: 171–86. doi: https://doi.org/10.1007/BF00055001
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74. doi: https://doi.org/10.2307/2529310
30. El HPalomo JM, Palomo JM. Measuring the airway in 3 dimensions: a reliability and accuracy study. *Am J Orthod Dentofacial Orthop* 2010; **137**(4 Suppl): S50 e1-9–discussion S50-2. doi: https://doi.org/10.1016/j.ajodo.2010.01.014
31. Bui NL, Ong SH, Foong KW. Automatic segmentation of the nasal cavity and paranasal sinuses from cone-beam CT images. *Int J Comput Assist Radiol Surg* 2015; **10**: 1269–77. doi: https://doi.org/10.1007/s11548-014-1134-5
32. Tingelhoff K, Eichhorn KW, Wagner I, Kunkel ME, Moral AI, Rilk ME, et al. Analysis of manual segmentation in paranasal CT images. *Eur Arch Otorhinolaryngol* 2008; **265**: 1061–70. doi: https://doi.org/10.1007/s00405-008-0594-z
33. Tingelhoff K, Moral AI, Kunkel ME, Rilk M, Wagner I, Eichhorn KG, et al. Comparison between manual and semi-automatic segmentation of nasal cavity and paranasal sinuses from CT images. *Conf Proc IEEE Eng Med Biol Soc* 2007; **2007**: 5505–8. doi: https://doi.org/10.1109/IEMBS.2007.4353592
34. Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. *Med Phys* 2011; **38**: 2708–14. doi: https://doi.org/10.1118/1.3578928
35. Abbott MB, Donnelly LF, Dardzinski BJ, Poe SA, Chini BA, Amin RS. Obstructive sleep apnea: MR imaging volume segmentation analysis. *Radiology* 2004; **232**: 889–95. doi: https://doi.org/10.1148/radiol.2323031581