In vitro assessment of a three-dimensional segmentation algorithm based on the belief functions theory in calculating renal volumes by MR imaging

P.-H. Vivier, M. Dolores, Isabelle Gardin, P. Zhang, Caroline Petitjean, J.-N. Dacher

Editorial Manager(tm) for American Journal of Roentgenology
Manuscript Draft

Manuscript Number:

Title: In vitro assessment of a three-dimensional segmentation algorithm based on the belief functions theory in calculating renal volumes by MR imaging

Article Type: Original Research

Section/Category: Genitourinary Imaging

Keywords: kidney; magnetic resonance imaging; organ size; sensitivity and specificity; image processing, computer-assisted

Corresponding Author: Prof. Jean-Nicolas Dacher, M.D., Ph.D.

Corresponding Author's Institution: University Hospital of Rouen, LITIS EA 4108

First Author: Pierre-Hugues Vivier, M.D., M.Sc

Order of Authors: Pierre-Hugues Vivier, M.D., M.Sc; Michael Dolores, MSc; Isabelle Gardin, PhD; Peng Zhang, PhD; Caroline Petitjean, PhD; Jean-Nicolas Dacher

Abstract: OBJECTIVE: Renal volume measurement is an essential part of split renal function assessment in MR urography. The aim of this study was to assess the accuracy and repeatability of a three-dimensional segmentation algorithm based on the belief functions theory for calculating renal volumes from MR images.

MATERIALS AND METHODS: The true volume of 20 various sized animal kidneys was obtained by fluid displacement. Each kidney was examined using two different MR units. 3D proton-density weighted-acquisitions with incremental slice thickness were obtained. The MR volume was then measured with a segmentation algorithm based on the belief functions theory. Two independent observers performed all segmentations twice. Accuracy, intra- and inter-observer variability were evaluated by the Bland-Altman method. The number and type of manual corrections were recorded, as well as the entire processing time.

RESULTS: The mean renal volume estimated by fluid displacement was 114 mL (range, 38-224 mL). Regarding renal volume assessment obtained from adjacent axial MR images, maximal standard deviations of the difference were: 2.2 mL (accuracy), 0.6 mL and 1.8 mL (respectively intra- and inter-observer variability). Segmentation of axial slices provided better accuracy and reproducibility as compared with coronal ones. Overlapped coronal slices yielded poor results due to the partial volume effect. The mean processing time, including optional manual modifications, was less than 75 seconds.

CONCLUSION: The belief functions theory is an accurate and reproducible mathematic method to assess renal volume from MR adjacent images.

Suggested Reviewers:

Opposed Reviewers:

ABSTRACT:

OBJECTIVE: Renal volume measurement is an essential part of split renal function assessment in MR urography. The aim of this study was to assess the accuracy and repeatability of a three-dimensional segmentation algorithm based on the belief functions theory for calculating renal volumes from MR images.

MATERIALS AND METHODS: The true volume of 20 various sized animal kidneys was obtained by fluid displacement. Each kidney was examined using two different MR units. 3D proton-density weighted-acquisitions with incremental slice thickness were obtained. The MR volume was then measured with a segmentation algorithm based on the belief functions theory. Two independent observers performed all segmentations twice. Accuracy, intra- and inter-observer variability were evaluated by the Bland-Altman method. The number and type of manual corrections were recorded, as well as the entire processing time.

RESULTS: The mean renal volume estimated by fluid displacement was 114 mL (range, 38-224 mL). As regards renal volume assessment obtained from adjacent axial MR images, maximal standard deviations of the difference were: 2.2 mL (accuracy), 0.6 mL and 1.8 mL (respectively intra- and inter-observer variability). Segmentation of axial slices provided better accuracy and reproducibility as compared with coronal slices. Overlapped coronal slices yielded poor results due to the partial volume effect. The mean processing time, including optional manual modifications, was less than 75 seconds.

CONCLUSION: The belief functions theory could be considered an accurate and reproducible mathematic method to assess renal volume from MR adjacent images.

INTRODUCTION

Dynamic MR Urography (MRU) is increasingly performed in routine practice as it can provide both morphologic and functional information without radiation exposure. Dynamic MRU can be used to assess renal excretion as well as split renal function. Rohrschneider et al. [1-4] have reported an excellent correlation between dynamic MRU and diuresis renal scintigraphy, which is considered the reference examination [5, 6]. An accurate renal volume evaluation is essential to calculate split renal function [7-9].

Furthermore, previous studies have shown that MRU could determine other functional parameters such as renal blood flow, and single kidney glomerular filtration rate which both require normalization to renal volume [10-12].

In previously reported studies, renal segmentation has been manually performed following a threshold [13, 14] or non threshold stage [15-19]. Although new segmentation techniques have been recently applied to the kidney, few studies have evaluated their accuracy and repeatability [13-16].

The three-dimensional segmentation algorithm evaluated in the present study was based on the belief functions theory which permits managing imprecise and uncertain information, such as partial volume effect and noise [20-22] (Vauclin S et al., presented at the 2005 International Symposium on Signals, Circuits and Systems). This theory represents a connection between fuzzy reasoning and probability. An imaged organ is generally composed of connected voxels sharing similar characteristics such as the grey level which was analyzed in this study. The aim of our research was to assess the accuracy and repeatability of this algorithm in calculating renal volumes from ex vivo MR images.

MATERIAL AND METHODS

Measurements were performed on isolated animal kidneys with known volumes in order to assess the algorithm accuracy and reproducibility.

Animal kidneys and standard volume measurements

The study was performed in 9 lamb and 11 pig kidneys obtained from an abattoir. The hilar structures were extensively removed. Intra-sinusal cavities were flat. The kidneys were first soaked in a basin filled with 0.9% saline for 2 hours in order to expand to their fullest volume capacity. Volumes were then immediately measured by a technician not involved in the MR measurements. The kidney volume was considered equal to the volume of displaced fluid [15]. The final result, used as the reference kidney volume, was the average of 4 successive measurements.

Then kidneys were prepared for MR imaging. They were placed in suspension in fat within a plastic container to simulate normal perirenal tissue. Fat was composed of a mixture of sunflower oil and hydrogenated refined copra oil so that the mixture was liquid when heated and solid at room temperature. A first layer of heated fat was placed in the container, and then refrigerated. Once solid, 4 kidneys were placed on the fat layer. The container was then filled with liquid fat, and then refrigerated again.

MR acquisitions

MR acquisitions were performed within six hours following reference measurements. Fat-suppressed gradient-echo proton-density weighted acquisitions were carried out. 3D acquisitions of each kidney were performed, using incremental section thickness on the two MR units available at our institution: a 1-T unit (Philips, Gyroscan NT, The Netherlands) and a 1.5-T unit (Symphony, Siemens, Germany). Acquisition parameters are summarized in

Tables 1 and 2. Coronal and axial acquisitions were performed and the signal was obtained via an abdominal phased-array coil. SENSitivity Encoding (SENSE®, 1.6 factor) was used for all acquisitions performed with the 1-T unit. A 4 mm overlapped sequence with an actual 8 mm slice thickness, available on the 1-T unit, was also tested. Neither parallel imaging nor overlapped sequences were available on the 1.5-T equipment at the time of the experiment.

Segmentation method

In order to label a given voxel, its grey level was evaluated as well as that of adjacent voxels within the same slice, and also within the 2 adjacent slices above and below (3 D segmentation). Data fusion permitted to decide whether or not the assessed voxel was within or beyond the limits of the organ. The segmentation process is fully explained in the appendix.

Processing

All images were stored on CD-Roms obtained from regular CD burners. DICOM images were transferred to personal computers database (Pentium® 4.3 GHz, 512 Mo). The MR renal volume measurements were performed using the belief functions theory based MR-Urography plug-in incorporated to the ImageJ software [23, 24]. The method for renal volume measurement is illustrated in Figure 1. After a stack of images was accessed, the first step was to limit the volume of interest. The observer indicated both first and last images showing a visible part of kidney. A stack of images was determined in between. Then, a rectangle was drawn on a mid-stack image showing the kidney in its largest dimensions in order to encompass the kidney on all the selected images. This rectangle was automatically copied on all images for automatic segmentation.

After segmenting, a visual quality control was performed. Manual correction of the segmented region of interest (ROI) on each image was feasible at this stage. Lastly, the renal volume was automatically calculated by a voxel-count method.

The MR kidney volumes were measured independently by two independent observers: a radiologist with 3 years of experience in MR Urography and a computer analyst who participated in the research. Each stack of images was evaluated on four separate occasions by each observer (Fig. 2). Manual corrections were only allowed for two of the four measurements. The obtained kidney volume, the number of manual corrections performed (when allowed) and the processing time were routinely recorded in the computer database. The observers were blinded to the reference volumes, to their own results, and to all results obtained by the other observer.

Quantitative and statistical analyses

a. Water displacement: Reproducibility of measurements was assessed by the Bland-Altman method [25]. For each kidney, the mean of four results was calculated. One of these four results was randomly chosen, then subtracted from the mean. Intra-observer variability was calculated as follows: for each kidney, two of the four measurements were randomly chosen, and then one was subtracted from each other. The mean difference and standard deviation of the difference (SDD) were calculated. Results were plotted on graphs showing the 95% limits of agreement (mean difference ± 2 SDD).

b. MR renal volume: The accuracy, intra- and inter-observer variability of MR measurements were also assessed by the Bland-Altman method. Accuracy was obtained from the errors in measurements due to the algorithm. For each kidney, the error was defined as the difference between the reference volume and the first MR measurement of each observer. This calculation was performed for each MR unit, for all types of acquisition (coupling scan

plane and thickness), and for manually modified vs. non modified segmentation. For all kidneys (n=20), the mean difference and SDD were calculated as previously described.

In order to assess intra-observer variability, the difference between the first and the second measurements of each observer was calculated for each kidney. In order to assess inter-observer variability, the difference between the first measurements of both observers was calculated for each kidney.

A matched-pairs Wilcoxon rank sum test (Statview 5.0, SAS Institute, Cary, NC) was used to evaluate the required manual modifications. For each section thickness, comparison was made between axial and coronal acquisitions.

In order to compare the present study with previously reported series [13, 14], the volumes obtained from 3 mm thick MR coronal slices (same thickness as in comparators) were also plotted against reference volumes using a linear regression to obtain the coefficient of determination R^2.

RESULTS

Reference standard volumes

The mean renal volume estimated by fluid displacement was 114 mL (range; 38-224 mL). Mean renal volume of lamb kidneys was 63 mL (range; 38-99 mL) and mean renal volume of pig kidneys was 155 mL (range; 74-224 mL). Water displacement accuracy SDD was 1.25 mL. Intra-observer SDD was 2 mL.

MR-calculated volumes

An experiment was used to explain how the limits of agreement were determined by the Bland-Altman technique (Fig. 3). No influence of kidney volumes on the SDD values could be found in any of the experiments.

Accuracy, intra- and inter-observer variability graphs are respectively shown in Figures 4, 5 and 6. The widest intervals were reported for each section thickness.

For example, the accuracy related SDDs obtained from 4 mm thick adjacent axial images without manual modifications were 1.07 mL for Observer 1 (Fig. 3) and 1.05 mL for Observer 2 on the 1-T unit, and 1.75 mL for Observer 1 and 2.1 mL for Observer 2 on the 1.5-T unit. Therefore, the 2.1 mL SDD defined the limits of agreement (4xSDD = 8.4 mL; interval corresponding to [-3 mL; 5.4 mL], Fig. 4). All 95% limits of agreement spanned the axis of zero.

In the following section, all results are reported for axial and coronal slices respectively. Maximal SDD values of adjacent slices were 2.2 and 7.5 mL for accuracy; 0.6 and 3.3 mL for intra-observer variability; 1.8 and 5.4 mL for inter-observer variability. In segmenting overlapped 4 mm slices, maximal SDD values obtained for accuracy were 1.2 and 10.5 mL. They were 0.6 and 8.1 mL for intra-observer variability. SDD values were 0.6 and 5.7 mL for inter-observer variability.

The mean number of manually corrected images per stack is shown in Figure 7. Image stacks were composed of 17 to 75 images depending on the scan plane and section thickness. Processing time was always less than 75 seconds including manual corrections. Axial images required less modification than coronal images ($p < 0.05$) for each section thickness. For axial acquisitions, approximately one image per stack was modified. In the coronal plane, the mean number of modified images per stack did not differ significantly whatever the 2, 3 or 4 mm section thickness that was chosen. Conversely, the number of modifications was significantly higher for 5 mm thick slices and for 4 mm thick overlapped slices ($p < 0.02$).

R^2 was equal to 0.99 for both observers and for each trial when assessing the accuracy of measurements in 3 mm thick coronal MR images.

DISCUSSION

Assessing the renal volume is an essential part of MRU in order to determine the split renal function [7, 14, 26]. Although feasible, manual segmentation of the kidneys is excessively time consuming in order to be used in routine evaluation. Rohrschneider et al. [2, 3] suggested inferring the renal volume from the renal parenchyma area measured on the single mid-renal coronal scan used for the dynamic study. This approximation did not take into account any cases of atrophy or hypertrophy located outside the selected plane. Considering these limitations, we have evaluated newly developed software based on the belief functions theory in terms of accuracy, reproducibility and processing time.

Fresh animal kidneys were used rather than physical objects. This choice was made due to the specific shape and surface to volume ratio [27] of kidneys which are prone to induce partial volume effect, itself impacting on the quality of segmentation. Lamb and pig kidneys were selected for their size in order to simulate respectively children and adult kidneys [16, 28]. Water displacement, was chosen as the reference method in spite of a 5 mL (4 x 1.25 mL) range of uncertainty.

The segmentation algorithm was tested on two different MR units. Geometric parameters (matrix, field of view) were defined based on clinical practice. Basically, axial and coronal scans were assessed with incremental slice thickness. The results obtained on one unit could not be compared to those obtained with the other one due to different reconstruction algorithms and voxel sizes (Table 1). A 350 mm asymmetric field of view commonly used in routine adult practice was chosen for all experiments. As smaller voxels induce a less partial

volume effect, results could be improved in cases of a reduced field of view more adapted to pediatric imaging.

The overall evaluation of the segmentation method appeared satisfactory when processing adjacent slices. The quality of results (accuracy and reproducibility) appeared to decrease with the thickness of slices in the coronal plane (Figs. 4-6). In contrast, slice thickness did not seem to influence the quality of segmentation in the axial plane.

Results obtained from axial images were better than those from coronal images. This finding is in agreement with previously published data [27]. In fact, the axial plane provides a less partial volume effect since it is grossly perpendicular to the long axis of the kidneys. Otherwise, more slices were available for measurement (Table 2). Furthermore, in this ex vivo study, coronal slices could have been more affected by a partial volume effect than in vivo due to the flattening of the kidneys. It is noteworthy that whatever the slice thickness, the intra-observer variability of segmentation of axial MR slices remained lower than that of water displacement.

Inter-observer SDD of adjacent axial scans segmentation remained equal or below 1.8 mL in the absence of any manual modification. The impact of manual modifications was positive on inter-observer variability when segmenting adjacent axial slices (inter-observer SDD < 0.5 mL). In contrast, this impact was negative when segmenting adjacent coronal slices; then, inter-observer SDD reached 3.3 mL with 4 mm thick slices, and as high as 5.4 mL with 5 mm thick slices. Again, this apparent contradiction could be explained by the marked partial volume effect observed in the coronal plane. Results appear to indicate that the semi-automatic segmentation algorithm was more effective as compared to human operators in cases of significant partial volume effect. Comparable impact of manual modifications was obtained in terms of intra-observer variability.

Overall results obtained from the segmentation of coronal overlapped 4 mm slices were less satisfying. SDD reached 10.5 mL for accuracy, 8.1 mL for intra-observer variability, and 5.7 mL for inter-observer variability. In our experience, this type of acquisition should be avoided to assess renal volume. In this situation, the algorithm in fact proved more accurate when used with no manual modification.

Several studies have been published in the last decade regarding kidney segmentation. Some of them (Boykov Y et al., presented at the 2001 Medical Image Computing and Computer-Assisted Intervention; Song T et al., presented at the 2005 International Society for Magnetic Resonance in Medicine Scientific Meeting; Yuksel SE et al., presented at the 2005 Computer Assisted Radiology and Surgery Conference) have focused on the time-intensity curves of specific regions such as the renal cortex or medulla. In these studies, the accuracy and reproducibility of the segmentation procedures were not assessed on multiple models. Furthermore, the kidney volume was not evaluated.

In contrast, the authors of some clinical studies have focused their attention in the assessment of kidney volume. Manual segmentation with [13, 14] or without a preliminary threshold stage [15-19] has been used. Manual segmentation has been shown to be time consuming and produces a relatively poor reproducibility. Semi-automatic segmentation (median filter, user-defined intensity thresholds, morphologic erosions and dilatations, and region growing steps) has also been used to determine renal volumes [7-9].

Few authors have evaluated the reproducibility of their segmentation technique. Among them, Bakker et al. carried out two experiments, one in pig kidneys [15], and also in humans [16]. As in our study, intra- and inter-observer variability was obtained using the Bland-Altman method. Five millimeters thick images were acquired in the coronal plane. The SDD for intra-observer variability was 8.2 mL in [15] and 7.3 mL in [16], as compared with 3.3 mL with manual modifications (0.6 mL without manual modification) in the part of our

study testing the same slice thickness with an inferior spatial resolution. The SDD for inter-observer variability were 6.4 mL in [15] and 9.9 mL in [16] as compared with 5.4 mL with manual modifications (0.6 mL without manual modification) in our study.

Similarly, few authors have assessed the accuracy of their segmentation method. The comparison of our accuracy results with other published data is thus limited. Coulam et al. [13] and van den Dool et al. [14] used linear regression to assess the accuracy of their volume assessment of pig kidneys (manual segmentation followed by a threshold stage). However, these results are subject to discussion since linear regression is known to measure the strength of a relation between two variables rather than the agreement between them [25]. R^2 were respectively 0.86 and 0.98, as compared with 0.99 in our study. Van den Dool et al. [14] also reported the intra- (3%) and inter-observer (1.9%) variability as relative mean differences in healthy adult humans. Analysis of Bland-Altman graphs in the present study shows that errors did not depend on kidney volumes. Therefore, we were able to obtain absolute results in mL in the present study.

The processing time was reported in the in-vitro study by Bakker et al. [15]. It ranged from 5 to 8 minutes as compared with less than 75 seconds in the present series.

The main advantage of the belief functions theory is its ability to merge 3D information originating from adjacent voxels to delineate the kidney boundary. The algorithm is only semi-automatic since the first part of the process (definition of the volume of interest) is manual. Belief functions theory has also been used for other purposes such as segmentation of brain MR images [29] (Capelle AS et al., presented at the 2004 International Conference on Information Fusion). The algorithm used in that research has been previously used to segment intra-thoracic organs from CT images in the context of conformational external radiotherapy (Vauclin S et al., 2005 ISSCS symposium).

Despite the overall optimal results obtained with this segmentation method, some manual corrections may be necessary (Fig. 7) and software was designed to account for this adjustment. Different situations occurred. For example, a common situation requiring manual modification was the segmentation of intra-sinusal fat (Fig. 8). The segmentation was basically appropriate in the experiment but it had to be modified due to the mode of comparison (water displacement) which included the remaining intra-sinusal fat.

Our study has limitations and cannot be immediately extrapolated to MR urography in humans. First, the study did not test the influence of respiratory motion. Moreover, MR urography is mainly used in children where motion artifacts can deteriorate the image quality. Secondly, the investigated kidneys were not contrast enhanced and their positioning was artificial. Thirdly, a 2 class algorithm (see appendix) adapted to ex vivo kidneys was used, while in humans a 3 class algorithm is obligatory to segment the kidney from perirenal fat, the urinary excretory tract and adjacent organs. At least, more than two observers would have been preferable to strengthen the results of the study.

CLINICAL APPLICATION

For clinical use, a 3 class algorithm was made available as a plug-in in the public domain software ImageJ [23, 24, 30, 31]. This tool can be freely accessed at the National Institutes of Health website (http://rsb.info.nih.gov/ij/). The program runs as a downloadable application, on any computer with a Java 1.1 or later virtual machine. The algorithm enables functional renal parenchyma volume assessment from adjacent contrast enhanced T1-weighted DICOM images covering the entire kidneys. We suggest performing the volume assessment sequence at the end of the dynamic phase of MR urography. At this stage, the renal excretory system is usually contrast enhanced. We re-inject intravenously the same low dose of gadolinium chelate as that used for the dynamic study (0.025 to 0.05 mmol/kg) [32-

34]. The volume assessment sequence is then acquired at the tubular phase 60 seconds after this second injection. Renal parenchyma signal is then homogeneous and superior to that of adjacent organs, while inferior to the signal of the enhanced urinary tract.

Radiologists involved in MR urography commonly use this tool at our institution with no specific help from the computer analyst. Segmentation results in daily clinical practice are satisfactory except in much dilated or atrophic kidneys where manual segmentation remains useful.

APPENDIX

The employed segmentation method was based on the belief functions theory. The basic principle of the algorithm is to define which class each voxel belongs to. A class is a set of voxels with closed grey level values. A volume of interest is composed of a cluster of adjacent voxels belonging to the same class. The belief functions algorithm includes 5 consecutive stages.

Stage 1 is a definition of the information source including a central voxel V, as well as its neighbors V_n. A 3 x 3 x 3 filter was used so the classification of any voxel relies on its own grey value as well as on the values of its 26 adjacent voxels.

Stage 2 is qualification of the classes: number of classes, center of gravity Ck of each class. In this experiment, two classes were defined corresponding to fat and kidney tissue. The two Ck values were routinely extracted using the K-means clustering algorithm.

Stage 3 is the calculation of each voxel (V) basic belief assignment (BBA) with respect to each class (CL). BBA m[V], also called mass of belief, associated to each studied voxel was defined as follows:

$$\mathbf{m[V](\{CL_k\}) = \delta.e^{-(\gamma.d(V,Ck))}}$$

where $d(V,C_k)$ is the Euclidean distance between the voxel V intensity and the characteristic value C_k of CL_k. BBA $m[V](\{CL_k\})$ quantifies the probability that voxel V belongs to class k (CL_k), based on its own characteristics. This equation means that the highest the mass of belief, the closest the voxel V grey level and C_k. The constant γ modifies the influence of $d(V,C_k)$ in the computation of BBA. Preliminary experiments have shown that a 0.05 γ value provided satisfactory results. δ is a constant coefficient belonging to the interval]0,1[used to avoid assigning the totality of the mass to CL_k when the distance $d(V,C_k)$ is null. In other words, if the voxel V intensity is equal to the characteristic value C_k, V will

not be automatically considered to belong to the CL_k. From our experience, a value of 0.9 yielded the best results.

The BBA $m[V_n]$ corresponding to the neighbor voxels V_n are computed the same way, following the equation:

$$\mathbf{m[V_n](\{CL_k\}) = \alpha.\delta.e^{-(\gamma.d(Vn,Ck))}}$$

The goal of the α coefficient is to dampen the contribution of the mass of belief of neighbor voxels V_n. To classify voxel V, the information contained in voxels V_n is less reliable than that contained in voxel V itself. Coefficient α decreases exponentially with the distance between V and V_n to take into account the spatial nature of the relations between voxels,.

Stage 4 is the combination of information sources provided by the central voxel V and its neighbors. For each voxel V, BBA corresponding to V and its 26 neighbor voxels are aggregated to obtain a unique BBA m[V], using the Dempster's conjunctive rule of combination [35].

Stage 5 is decision making. The decision to assign the voxel V to a class k (CL_k) is taken by analyzing a probability $P(V_n \in CL_k)$. If this probability is higher than a specific threshold, the voxel is assigned to CL_k. If not, the decision will be taken at one of the following iterations. The first threshold value was 0.8 and progressively decreased at each iteration.

REFERENCES

1. Rohrschneider WK, Becker K, Hoffend J, et al. Combined static-dynamic MR urography for the simultaneous evaluation of morphology and function in urinary tract obstruction. II. Findings in experimentally induced ureteric stenosis. *Pediatr Radiol* **2000**;30:523-532
2. Rohrschneider WK, Haufe S, Clorius JH, Troger J. MR to assess renal function in children. *Eur Radiol* **2003**;13:1033-1045
3. Rohrschneider WK, Haufe S, Wiesel M, et al. Functional and morphologic evaluation of congenital urinary tract dilatation by using combined static-dynamic MR urography: findings in kidneys with a single collecting system. *Radiology* **2002**;224:683-694
4. Rohrschneider WK, Hoffend J, Becker K, et al. Combined static-dynamic MR urography for the simultaneous evaluation of morphology and function in urinary tract obstruction. I. Evaluation of the normal status in an animal model. *Pediatr Radiol* **2000**;30:511-522
5. O'Reilly PH. Standardization of the renogram technique for investigating the dilated upper urinary tract and assessing the results of surgery. *BJU Int* **2003**;91:239-243
6. Piepsz A. Radionuclide studies in paediatric nephro-urology. *Eur J Radiol* **2002**;43:146-153
7. Grattan-Smith JD, Perez-Bayfield MR, Jones RA, et al. MR imaging of kidneys: functional evaluation using F-15 perfusion imaging. *Pediatr Radiol* **2003**;33:293-304
8. Jones RA, Easley K, Little SB, Scherz H, Kirsch AJ, Grattan-Smith JD. Dynamic contrast-enhanced MR urography in the evaluation of pediatric hydronephrosis: Part 1, functional assessment. *AJR Am J Roentgenol* **2005**;185:1598-1607
9. McDaniel BB, Jones RA, Scherz H, Kirsch AJ, Little SB, Grattan-Smith JD. Dynamic contrast-enhanced MR urography in the evaluation of pediatric hydronephrosis: Part 2, anatomic and functional assessment of uteropelvic junction obstruction. *AJR Am J Roentgenol* **2005**;185:1608-1614
10. Debatin JF, Ting RH, Wegmuller H, et al. Renal artery blood flow: quantitation with phase-contrast MR imaging with and without breath holding. *Radiology* **1994**;190:371-378
11. Hackstein N, Bauer J, Hauck EW, Ludwig M, Kramer HJ, Rau WS. Measuring single-kidney glomerular filtration rate on single-detector helical CT using a two-point Patlak plot technique in patients with increased interstitial space. *AJR Am J Roentgenol* **2003**;181:147-156
12. Niendorf ER, Grist TM, Lee FT, Jr., Brazy PC, Santyr GE. Rapid in vivo measurement of single-kidney extraction fraction and glomerular filtration rate with MR imaging. *Radiology* **1998**;206:791-798
13. Coulam CH, Bouley DM, Sommer FG. Measurement of renal volumes with contrast-enhanced MRI. *J Magn Reson Imaging* **2002**;15:174-179
14. van den Dool SW, Wasser MN, de Fijter JW, Hoekstra J, van der Geest RJ. Functional renal volume: quantitative analysis at gadolinium-enhanced MR angiography--feasibility study in healthy potential kidney donors. *Radiology* **2005**;236:189-195
15. Bakker J, Olree M, Kaatee R, de Lange EE, Beek FJ. In vitro measurement of kidney size: comparison of ultrasonography and MRI. *Ultrasound Med Biol* **1998**;24:683-688
16. Bakker J, Olree M, Kaatee R, et al. Renal volume measurements: accuracy and repeatability of US compared with that of MR imaging. *Radiology* **1999**;211:623-628

17. Cheong B, Muthupillai R, Rubin MF, Flamm SD. Normal Values for Renal Length and Volume as Measured by Magnetic Resonance Imaging. *Clin J Am Soc Nephrol* **2007**:38 - 45
18. Cheung CM, Shurrab AE, Buckley DL, et al. MR-derived renal morphology and renal function in patients with atherosclerotic renovascular disease. *Kidney Int* **2006**;69:715-722
19. Cibulskyte D, Engberg A, Pedersen M, et al. Kidney volume increases during long-term ciclosporin a treatment. *Transplant Proc* **2006**;38:2714-2718
20. Dempster AP. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics* **1968**;39:957-966
21. Shafer G. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, **1976**
22. Smets P, Kennes R. The transferable belief model. *Artificial Intelligence* **1994**;66:191-234
23. Lefort C, Marouteau-Pasquier N, Pesquet AS, Pfister C, Vera P, Dacher JN. Dynamic MR urography in urinary tract obstruction: implementation and preliminary results. *Abdom Imaging* **2006**;31:232-240
24. Rasband WS. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA. *http://rsbinfonihgov/ij/* **1997-2005.**
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**;1:307-310
26. Riccabona M, Fritz GA, Schollnast H, Schwarz T, Deutschmann MJ, Mache CJ. Hydronephrotic kidney: pediatric three-dimensional US for relative renal size assessment-initial experience. *Radiology* **2005**;236:276-283
27. Luft AR, Skalej M, Welte D, Kolb R, Klose U. Reliability and exactness of MRI-based volumetry: a phantom study. *J Magn Reson Imaging* **1996**;6:700-704
28. Oswald J, Schwentner C, Lunacek A, Deibl M, Bartsch G, Radmayr C. Age and lean body weight related growth curves of kidneys using real-time 3-dimensional ultrasound in pediatric urology. *J Urol* **2004**;172:1991-1994
29. Capelle AS, Colot O, Fernandez-Maloigne C. Evidential segmentation scheme of multi-echo MR images for the detection of brain tumors using neighborhood information. *Information Fusion* **2004**;5:203-216
30. Abramoff MD, Magelhaes, P.J., Ram, S.J. Image Processing with ImageJ. *Biophotonics International* **2004**;11:36-42
31. Escott EJ, Rubinstein D. Free DICOM image viewing and processing software for your desktop computer: what's available and what it can do for you. *Radiographics* **2003**;23:1341-1357
32. Grenier N, Pedersen M, Hauger O. Contrast agents for functional and cellular MRI of the kidney. *Eur J Radiol* **2006**;60:341-352
33. Taylor J, Summers PE, Keevil SF, et al. Magnetic resonance renography: optimisation of pulse sequence parameters and Gd-DTPA dose, and comparison with radionuclide renography. *Magn Reson Imaging* **1997**;15:637-649
34. Teh HS, Ang ES, Wong WC, et al. MR renography using a dynamic gradient-echo sequence and low-dose gadopentetate dimeglumine as an alternative to radionuclide renography. *AJR Am J Roentgenol* **2003**;181:441-450
35. Dempster AP. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics* **1967**;38:325-339

TABLES

TABLE 1: Sequence parameters on both MR units

MR unit	1-T (Philips)	1.5-T (Siemens)
Sequence type	THRIVE	VIBE
TR / TE (ms)	10 / 2.5	11 / 1.93
Flip angle (°)	10	15
Number of Signals Acquired	1	
Field of view (mm)	350 x 278	350 x 284
Acquisition matrix (pixels)	272 x 140	256 x 168
Reconstruction matrix (pixels)	512 x 512	512 x 336
Acquisition pixel size (mm)	1.3 x 2	1.4 x 1.7
Reconstruction pixel size (mm)	0.7 x 0.7	

Note - THRIVE = T1 High Resolution Isotropic Volume Examination, VIBE = Volumetric Interpolated Breath-hold Examination.

TABLE 2: Section thickness versus number of images per stack and acquisition time.

	1-T (Philips)		1.5-T (Siemens)	
	Axial	Coronal	Axial	Coronal
Section thickness (mm)	Numbers of images per stack / *Acquisition time (sec)*			
2	75 / *75*	35 / *46*	75 / *111*	35 / *42*
3	55 /*55*	25 / *31*	55 / *78*	25 / *37*
4	50 / *51*	20 / *22*	50 /*59*	20 / *28*
5	45 / *46*	17 / *19*	45 / *51*	17 / *21*
4 overlapped	50 / *26*	20 / *9*	Not applicable	

FIGURES:

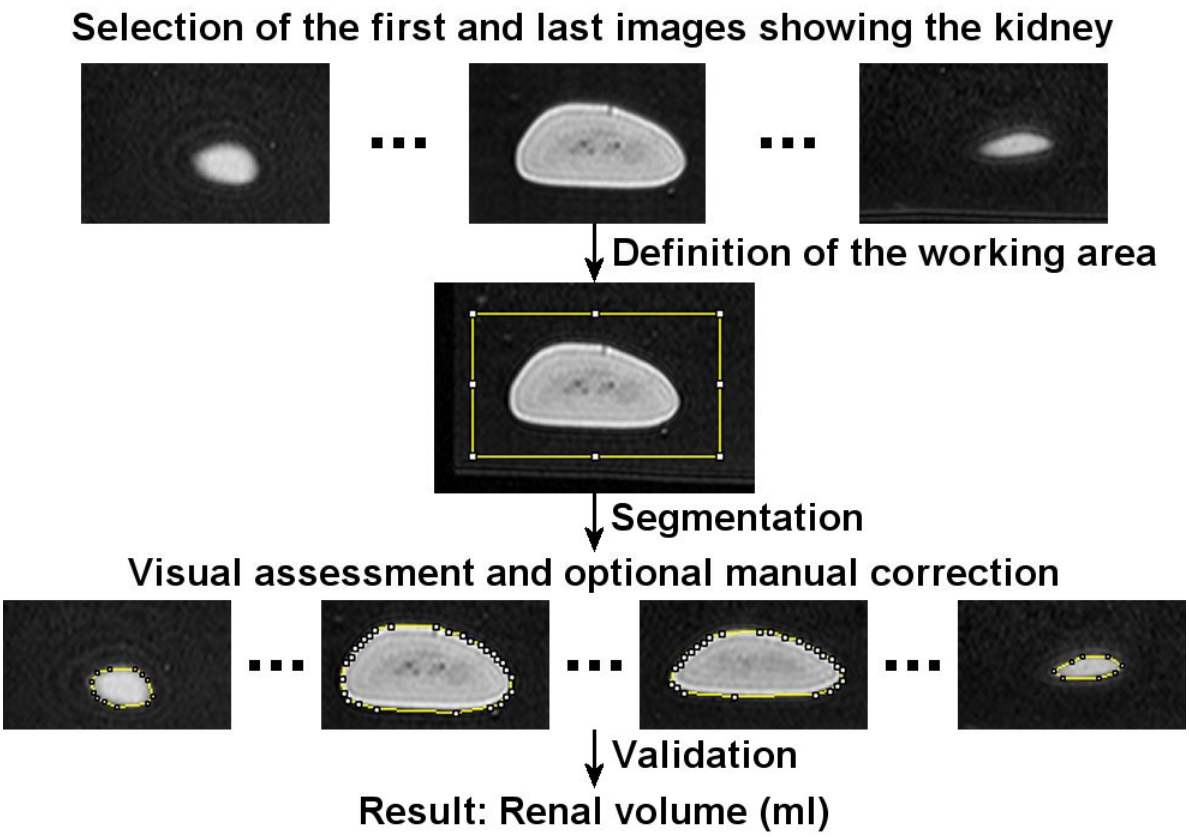


Fig. 1— Processing of a stack of axial images.

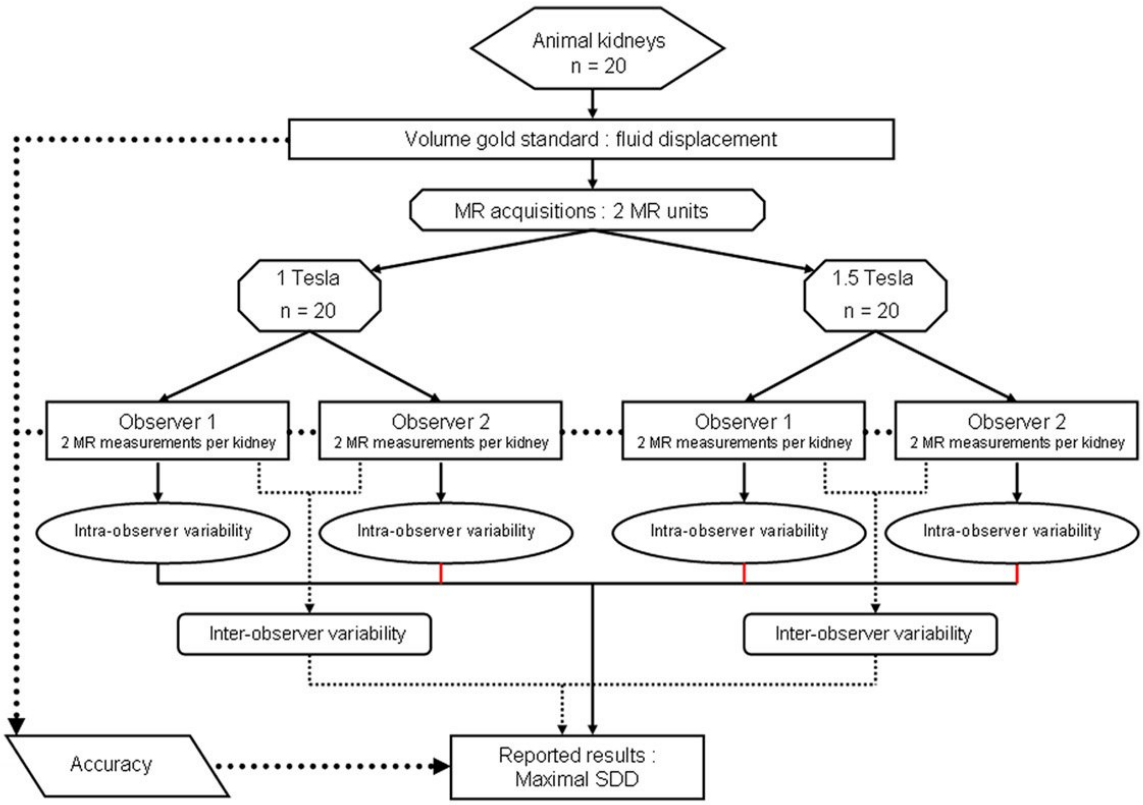


Fig. 2—Flowchart of the study. For each orientation (axial and coronal) and for each slice thickness, the process was performed twice: with and without manual modifications allowed.

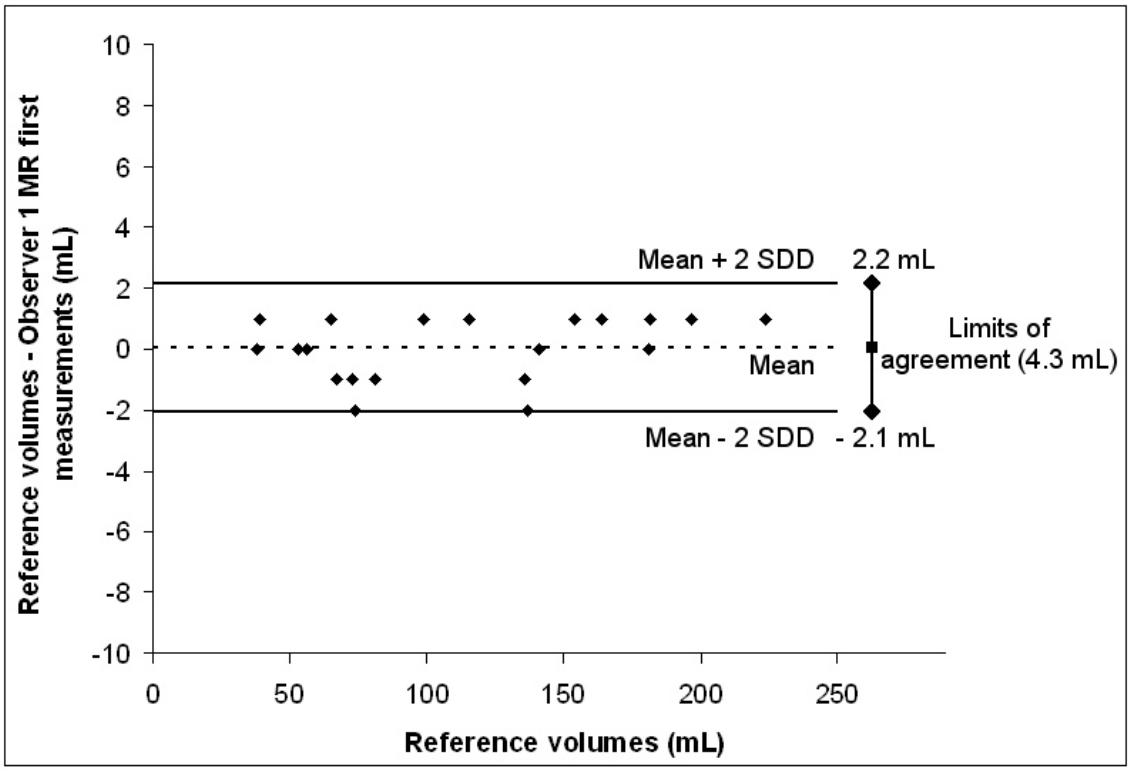


Fig. 3—Example of a Bland-Altman graph used for the accuracy assessment of the MR segmentation method drawn from the following experiment: 1-T unit, axial acquisition, 4 mm adjacent slices, Observer 1 first measurements, without manual modifications. Mean = 0.1 mL; SDD = 1.07 mL; 95% limits of agreement = 4 x 1.07 = 4.3 mL corresponding to [-2.1 mL; +2.2 mL].

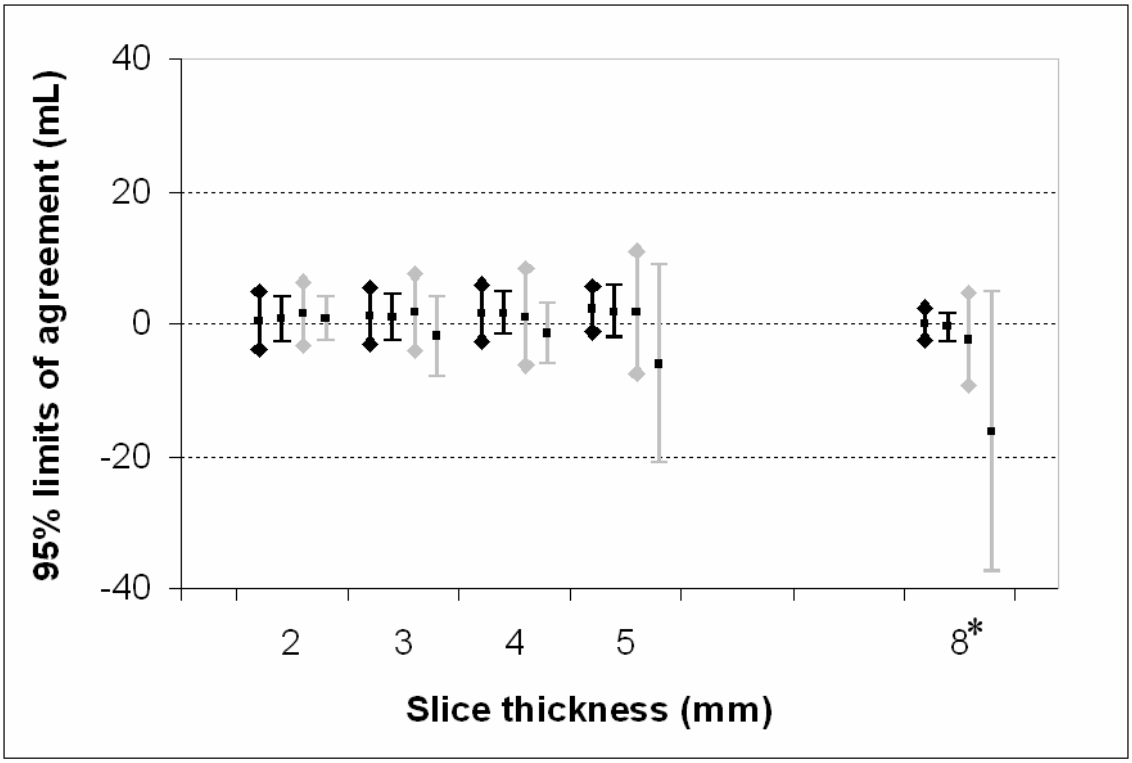


Fig. 4—Accuracy of the MR segmentation method. Plot showing maximal limits of agreement of MR calculated renal volumes vs. slice thickness.

Data from axial acquisitions are in black. Data from coronal acquisitions are in grey.

♦ no manual modification

- manual modification allowed

8*: 4 mm overlapped slices.

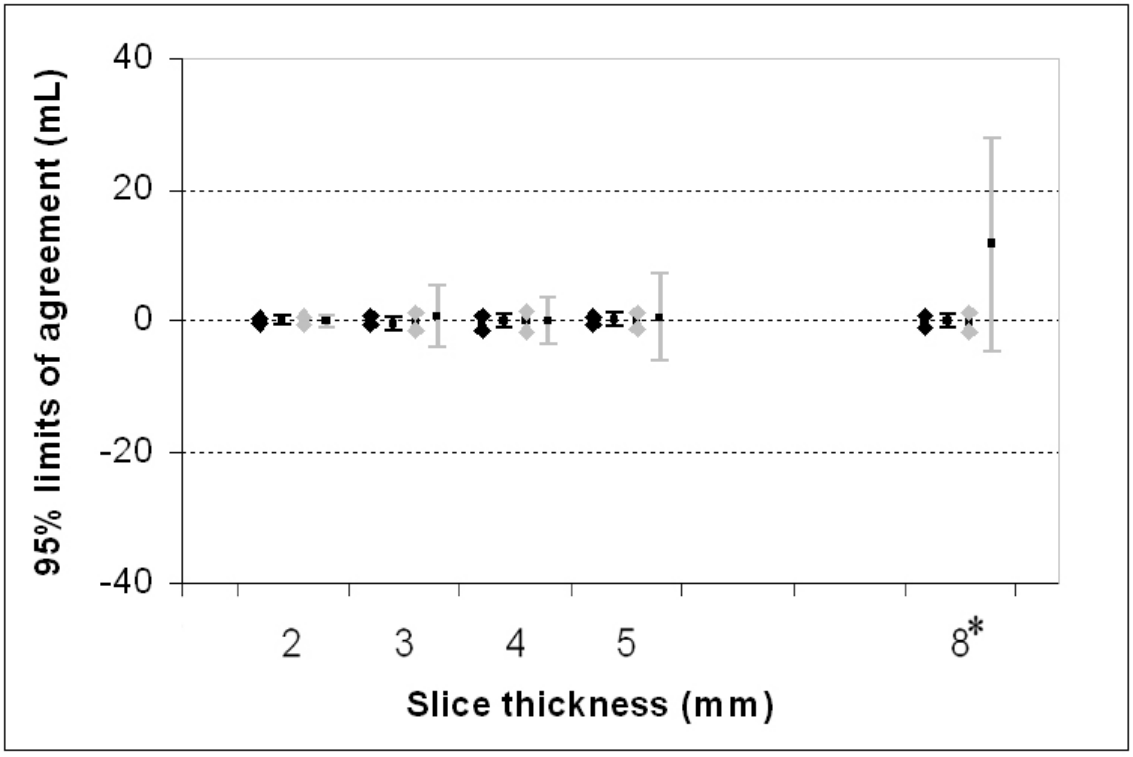


Fig. 5—Intra-observer variability of the MR segmentation method. Plot showing maximal limits of agreement of variability vs. slice thickness.

Data from axial acquisitions are in black. Data from coronal acquisitions are in grey.

♦ no manual modification

- manual modification allowed

8*: 4 mm overlapped slices.

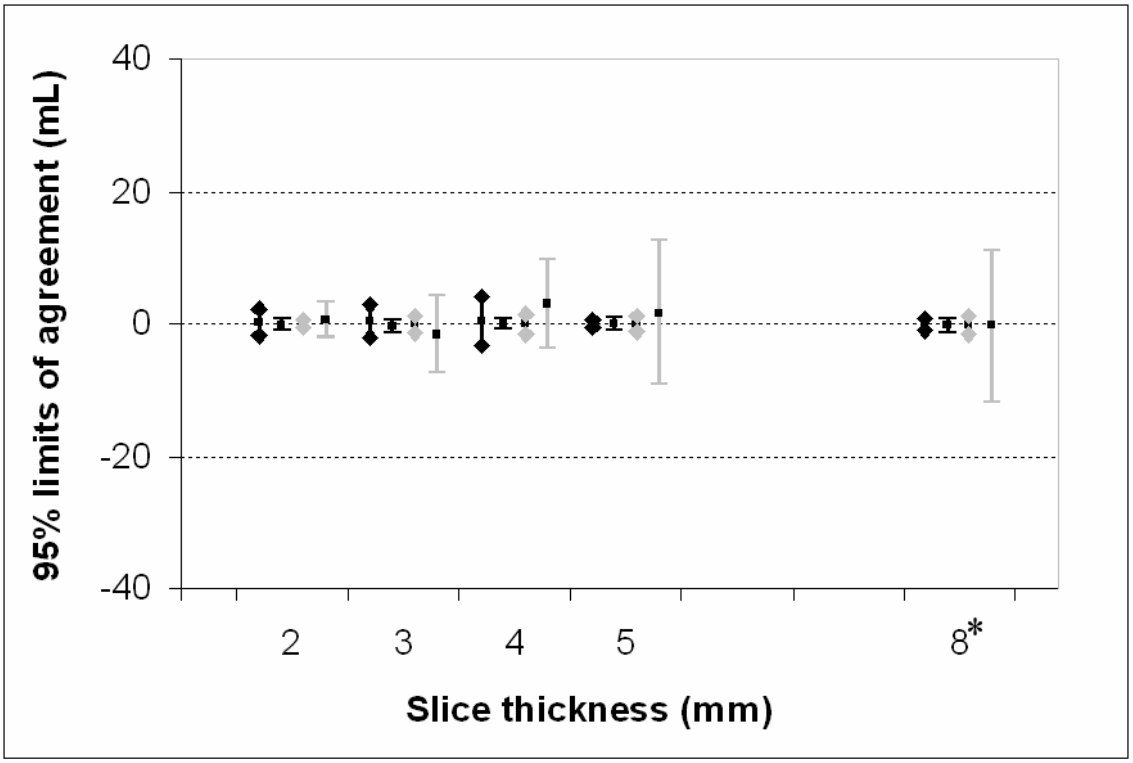


Fig. 6—Inter-observer variability of the MR segmentation method. Plot showing maximal limits of agreement of variability vs. slice thickness.

Data from axial acquisitions are in black. Data from coronal acquisitions are in grey.

♦ no manual modification

- manual modification allowed

8*: 4 mm overlapped slices.

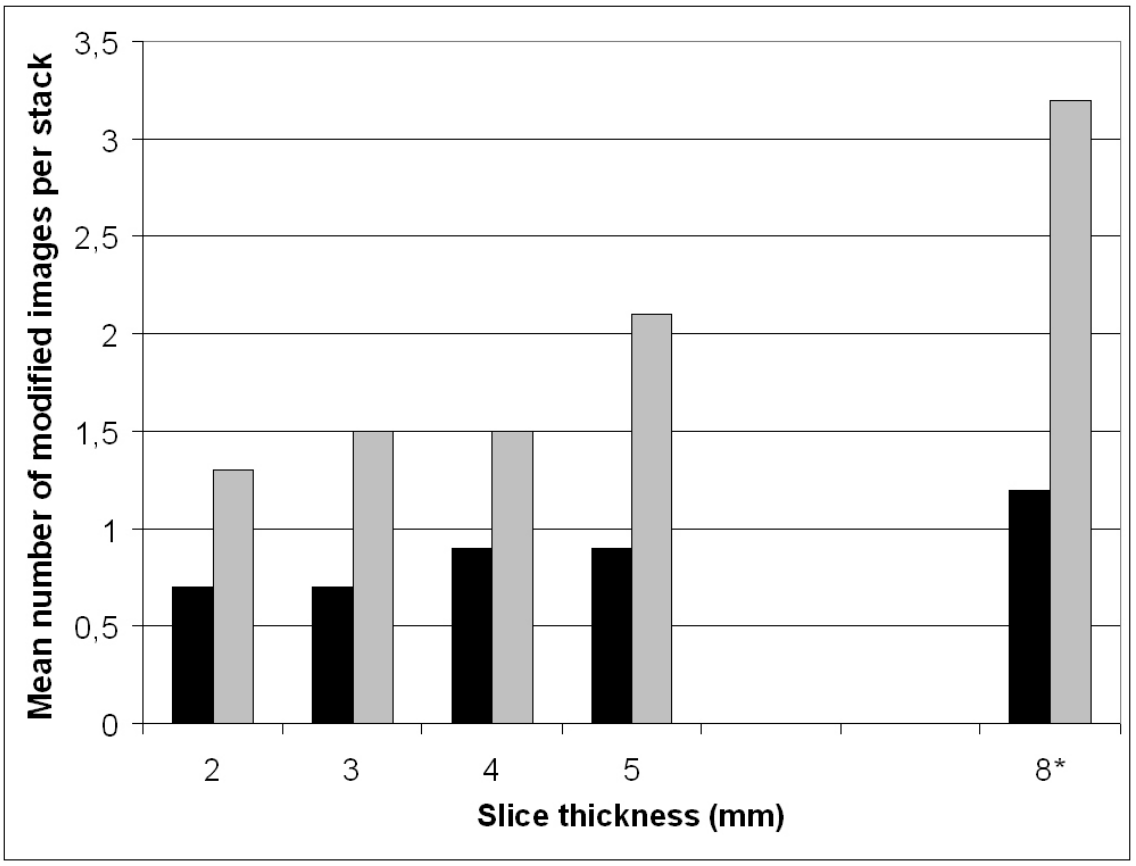
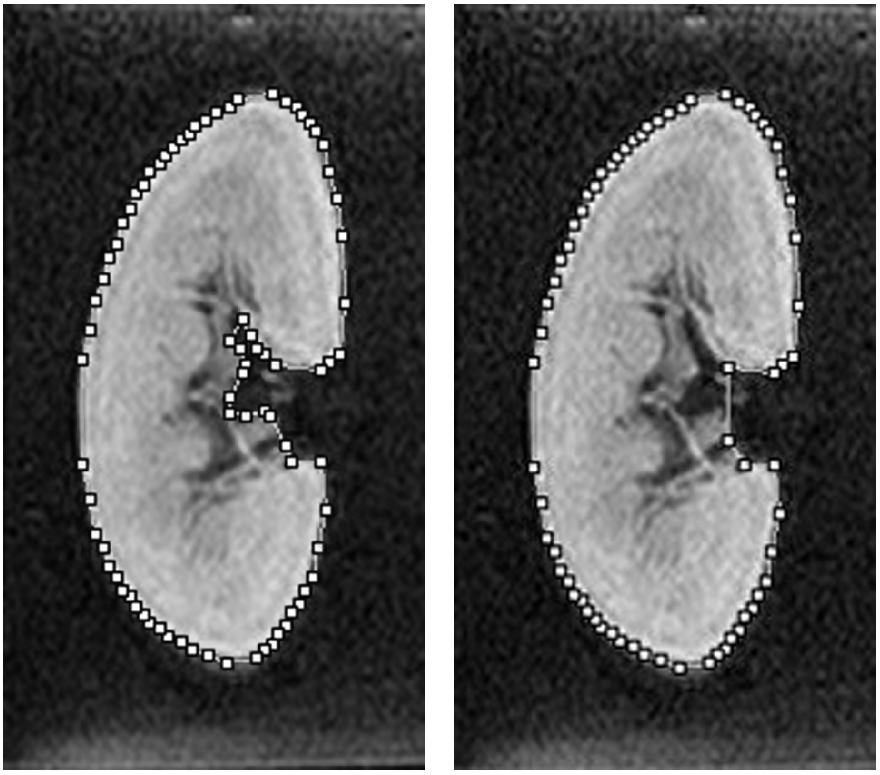


Fig. 7—Mean number of modified images per stack. Black: axial slices. Grey: coronal slices. 8*: 4 mm overlapped slices.



A **B**

Fig. 8—Segmentation of intra-sinusal fat on a coronal slice. (A) Appropriate semi-automatic segmentation of remaining intra-sinusal fat. (B) Manual modification applied to match fluid displacement measurement.

IN VITRO ASSESSMENT OF A THREE-DIMENSIONAL SEGMENTATION ALGORITHM BASED ON THE BELIEF FUNCTIONS THEORY IN CALCULATING RENAL VOLUMES BY MR IMAGING

Pierre-Hugues Vivier, MD, MSc, Michael Dolores, MSc, Isabelle Gardin, PhD, Peng Zhang, PhD, Caroline Petitjean, PhD, Jean-Nicolas Dacher, MD, PhD.

From the Department of Radiology, University Hospital of Rouen, 1 rue de Germont, F-76031 Rouen CEDEX, France (P.H.V., J.N.D); and LITIS Laboratory EA4108, School of Medicine and Pharmacy, University of Rouen, 22 boulevard Gambetta, F-76183 Rouen, France (P.H.V, M.D., P.Z., I.G., C.P., J.N.D).

Corresponding author:

Jean-Nicolas Dacher

Service de radiologie centrale

CHU Charles Nicolle, 1 rue de Germont

76031 Rouen Cedex

Tel (33) 2.32.88.64.96, Fax : (33) 2.42.88.82.35

Email: jean-nicolas.dacher@univ-rouen.fr

Supported in part by grants : from the « Société Française de Radiologie », France, STIC 2005 « Uro IRM », Nancy, France, and « Conseil Régional de Haute Normandie », France.

Acknowledgments: The authors thank:

- R. Medeiros, Rouen University Hospital Medical Editor, for his valuable advice in editing the manuscript.
- P. Vannoorenberghe, who initiated the belief functions theory in the Laboratory
- J.F. Menard who supervised the statistical analysis

Copyright & Disclosure Form

Click here to download Copyright & Disclosure Form: authors agreement.tif

In vitro assessment of a three-dimensional segmentation algorithm based on the belief functions theory in calculating renal volumes by MR imaging

Corresponding author:

Jean-Nicolas Dacher

Service de radiologie centrale

CHU Charles Nicolle, 1 rue de Germont

76031 Rouen Cedex

Tel (33) 2.32.88.64.96, Fax : (33) 2.42.88.82.35

Email: jean-nicolas.dacher@univ-rouen.fr

Type of article: original research

Disclosures of possible conflict of interest and/or commercial involvement: none

---

Dear Madam, Dear Sir,

Would you please consider for publication in the American Journal of Roentgenology, the enclosed article entitled:

"In vitro assessment of a 3D segmentation algorithm based on the belief functions theory in calculating renal volumes by MR Imaging"

The evaluated software is now routinely used in our Department (as well as in several other departments in Europe). However, we have deliberately chosen to limit the clinical application section as much as possible in order to remain within the scientific objectives of the study. Clinical examples are available and can be provided, if ever the Editorial Committee of the American Journal of Roentgenology would require some of them.

We are looking forward to reading the comments of the Reviewers of the Journal.

Yours Sincerely,

PH Vivier JN Dacher

Table

STARD checklist for reporting of studies of diagnostic accuracy

(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	Keywords (online submission)
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	2
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	NA
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	NA
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	NA
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	3-4
Test methods	7	The reference standard and its rationale.	3
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	3
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	NA
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	3 & 5
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	5
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	5-6
	13	Methods for calculating test reproducibility, if done.	5-6
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	NA
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	NA
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	NA
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	3
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	NA
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	NA
	20	Any adverse events from performing the index tests or the reference standard.	NA
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	6-7
	22	How indeterminate results, missing data and outliers of the index tests were handled.	NA
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	NA
	24	Estimates of test reproducibility, if done.	6-7
DISCUSSION	25	Discuss the clinical applicability of the study findings.	12-13

NA : non applicable