Renal volumetry with magnetic resonance imaging

Rolf H Christensen¹, Torbjörn Lundgren², Peter Stenvinkel³ and Torkel B Brismar⁴

Abstract
Background: No gold standard exists for renal volumetry in vivo.
Purpose: To devise and evaluate segmentation methods on magnetic resonance imaging (MRI) datasets.
Material and Methods: Five combinations of MRI pulse sequences and measuring methods were used to measure the renal volumes of five men aged 54–72 years scanned before autologous renal stem cell transplantation and three, six, and 12 months post transplantation.
Results: Renal volume did not change after stem cell transplantation. The results varied considerably: the reproducibility (coefficient of variation) was 4.0–6.0% and measurements took 1–13 min per kidney. Manual segmentation of images from the volumetric interpolated breath-hold examination (VIBE) without fat saturation sequence provided best reproducibility but was time-consuming. Use of the ellipsoid formula from half Fourier acquisition single shot turbo spin echo (HASTE) provided the fastest measurement, but resulted in lower reproducibility.
Conclusion: Renal volumetry based on images from the pulse sequence VIBE without fat saturation acquired using an out-of-phase TE may be investigated further, possibly in combination with the quick ellipsoid formula.

Keywords
Magnetic resonance imaging (MRI), volume, calculation, measurement, kidney

Date received: 20 December 2016; accepted: 9 August 2017

Introduction
Renal volumetry may be useful when evaluating potential kidney donors, prior to nephron-sparing surgery due to renal cancer and polycystic renal disease, and when evaluating the effect of stem cell transplantation in patients with chronic kidney disease. There is no gold standard for renal volumetry in vivo, but several techniques may be used.

The most reliable method to measure kidney volume would be to remove the kidney, immerse it in 0.9% saline and to measure the increased saline level. This method, the fluid displacement method, has been referred to as the gold standard (1). It has been used in research involving kidneys from animals and from living donors (2).

Another classical method is to measure the kidney with ultrasonographic (US) equipment and insert the values in the ellipsoid formula (height x width x thickness x π/6 = volume). This method, however, has been reported to underestimate renal volume (3). Furthermore, regional defects, such as scars and partial resections, will hamper the results.

The increased availability of computed tomography (CT) and magnetic resonance imaging (MRI) has entailed new methods of organ volumetry in vivo. Advantages of MRI to CT are that MRI is not as dependent on contrast media and yields no radiation.

¹Division of Medical Imaging and Technology, CLINTEC, Karolinska Institutet, Stockholm, Sweden
²Division of Transplantation Surgery, CLINTEC, Karolinska Institutet, Stockholm, Sweden
³Division of Renal Medicine, Karolinska Institutet, Stockholm, Sweden
⁴Department of Radiology, Karolinska University Hospital, Stockholm, Sweden

Corresponding author:
Rolf H Christensen, Department of Radiology C1-43, Karolinska University Hospital, 141 86 Stockholm, Sweden.
Email: rolf.christensen@stud.ki.se; rolf.christensen@hotmail.com

Creative Commons CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
The exclusion of simple renal cysts is also much easier due to higher conspicuity. In recent research, renal volumetry has usually been performed with CT or MRI (4,5). Methods and new software algorithms are often devised by mathematicians and software programmers (6), and may seem confusing to the clinical radiologist. In this void of established methods, work towards a gold standard is called for.

The aims of the present study were therefore to develop and evaluate methods for measuring renal volume with MRI, to calculate the reproducibility of these methods in combination with four MRI pulse sequences, and to estimate the time required for segmentation.

Material and Methods

Both kidneys of five male patients (age range = 54–72 years) were mapped with five combinations of MR sequences and measurement methods at four time points: before stem cell transplantation; and after three, six, and 12 months. This transplantation had been carried out within the framework of a larger study on the potentially beneficial effects of autologous stem cell transplantation in kidneys in patients with type 2 diabetic chronic kidney disease stage III-IV, where renal volume change was one of the outcomes. The overall outcome of the autologous transplantation study has been published elsewhere (7). A total of 200 kidney image stacks were thus processed. Permission was obtained from the Swedish Ethical Review Board (EPN), reference number: dnr 2013/288-31/3. The kidneys were mapped with MRI without contrast media, with a clinical 1.5T unit (Magnetom Aera, Siemens, Erlangen, Germany). The image processing software ImageJ was used on a standard computer. Two freeware plug-ins for ImageJ were used: Versatile Wand (VW) and Image Edge, both of which are freeware available on the Internet (8,9). All segmentation work was carried out by a single operator (RCH), trained by a radiologist with 13 years of experience (TBB). Four MRI sequences were used: diffusion-weighted imaging (DWI), three-dimensional (3D) T1-weighted volumetric interpolated breath-hold examination (VIBE) with fat saturation, VIBE without fat saturation, and T2-weighted (T2W) half Fourier acquisition single shot turbo spin echo (HASTE) (Table 1, Fig. 1). HASTE images were obtained in the axial and coronal imaging planes, all other series in the axial imaging plane. Parenchymal cysts less than 15 mm in diameter were not excluded from the measurements.

With MRI, a kidney is typically rendered with 25–40 slices, making up one stack of slices/images. In a 8-bit grayscale, pure black is given the value 0 and pure white the value 255, with the values in between representing shades of gray. With the voxel count method, the number of pixels representing the kidney in one tomography slice is counted. The area of these pixels is then multiplied by the slice thickness and the volume of this kidney slice is obtained. This is repeated for all the slices in one stack and the volume of the kidney is obtained. The voxels that are to be counted are the ones that have been selected, or segmented. The segmentation of the kidney constitutes the separation of it from the surrounding tissues and substances surrounding it (Figs. 2 and 3). Most methods in recent literature appear to be based on voxel counting and, with the exception of the ellipsoid formula method, this approach was applied throughout this study.

Various software applications were tried out and evaluated in order to find a simple, reproducible, time-efficient, and precise method. Some of the combinations of pulse sequence and segmentation method were clearly not compatible. Therefore, only the most optimal segmentation method for each pulse sequence was applied on all datasets. The most time-consuming part is the segmentation of regions of interest (ROI) in each slice of each image stack. For this reason, an automation method was called for. There seem to be two main approaches to segmentation and the voxel count method. (1) A ROI is segmented on the basis of grey values, whereby values are usually set for the brightest and/or darkest pixels of the kidney parenchyma, and all pixels with values between them are considered to

| Table 1. The imaging parameters for the sequences used. |
|---|---|---|---|
| DWI | VIBE with fat saturation | VIBE without fat saturation | T2 HASTE |
| Echo time (TE) | 62 | 1.9–1.93 | 1.9–1.93 | 93 |
| Repetition time (TR) | 7 600 | 4.02–5.04 | 3.96–4.02 | 2000 |
| Flip angle | 90 | 10 | 12 | 180 |
| Matrix | 192 × 115–154 | 288 × 186 | 288 × 186 | 320 × 211 |
| Slice thickness (mm) | 5 | 2.5 | 2.5 | 4 |
| Imaging gap (mm) | 0 | 0 | 0 | 0 |

All images were obtained in the axial plane except for HASTE which was also obtained in the coronal imaging plane.
belong to the kidney. This is the basis of the wand application, and of the threshold method, where the minimum and maximum levels are called threshold values. (2) The second approach, edge detection, is based on gray value gradients. Two semi-automatic segmentation methods and one manual segmentation method were selected. In addition, the simpler and far less time-consuming ellipsoid formula method was included.

**Semi-automatic segmentation based on gray values**

VW is a plug-in for ImageJ based on gray values. A wand is used to segment an area with pixels that lie within a certain range (tolerance) on the gray scale. The connectedness was set to 4-connected, so that the selection should include contiguous pixels at four sides. (With 8-connectedness the selection includes the “cornering” pixels, as well. However, this is likely to produce “leakage” (Fig. 3), right. The option *Include Holes* was chosen, in order to include single pixels that deviated excessively as a result of noise. Such “stray” pixels can be seen in Fig. 3, left and right. However, when there were actual holes in the parenchyma because of the pelvis and/or a calyx, these “real” holes must be excluded from the segmentation, and *Include Holes* was not chosen in such cases. Such a hole can be seen in the VIBE FS-image in Fig. 1. The ideal result of using the VW can be seen in Fig. 3, left, but suboptimal results may also occur, as shown to the right in Fig. 3.

MRI is inconsistent with regard to gray tones and contrast (Fig. 4). To minimize the effects of this

---

**Fig. 1.** The four pulse sequences used. DWI, diffusion-weighted imaging; VIBE, volumetric interpolated breath-hold examination; T2-HASTE, time 2 weighted half Fourier acquisition single shot turbo spin echo.

**Fig. 2.** Slice from the MRI pulse sequence VIBE with fat saturation, rendering the right kidney. The image has been tightly cropped. A ROI has been segmented along the edges of the kidney. MRI, magnetic resonance imaging; VIBE, volumetric interpolated breath-hold examination; ROI, region of interest.
inconsistency between stacks in MR images (Fig. 5), the stacks were made as uniform as possible: the images were cropped as tightly as possible around the kidneys, in order to minimize irregular gray values. Then it was a simple procedure to compare the average overall gray values of the entire stacks (Fig. 5). If one stack had a lower average gray value than another, this average value could easily be altered for the whole stack by using the Gamma Tool in ImageJ. Instead of adapting the settings of the wand, the images themselves were thus adapted to each other to be as uniform as possible. However, this was time-consuming.

Semi-automatic segmentation based on edge detection

With the aim of avoiding the inconsistency problems described above when using the VW, the edge detection method was then tested. The measurement and calculation of the images of the kidneys with the pulse sequences DWI and T2–HASTE were carried out with the edge detection plug-in “Image Edge,” which has the advantage of processing all slices in one stack simultaneously. Image Edge offers several functions. The function Area filter was selected empirically. Its two most prominent variables, Median filter radius and Deriche alpha value, were set to 4 and 0.5, respectively, throughout the DWI and HASTE stacks. These two variables were also selected empirically, with the aim of finding a balance between too many lines, making the image too complex, and too few lines, requiring manual completion (Fig. 6). Then the wand tool was used to count the number of pixels within the white lines generated.

Manual segmentation

Initially, manual segmentation with the computer mouse was tried on DWI-stacks. This segmentation method is time-consuming. Moreover, it was not considered consistent enough, so it was first abandoned. In most MR images, organ edges are somewhat fuzzy, making segmentation more difficult, imprecise, and inconsistent (Fig. 2). An exception may be the pulse sequence VIBE without fat saturation. As may be seen in Fig. 7, the organs, when imaged using this sequence, are surrounded by a consistent black edge,
due to the chemical shift effect, whereby the signals from water and fat cancel each other out where the signals from these entities are exactly equal. Image Edge was first tried with the pulse sequence VIBE without fat saturation, and with different settings, but no results were considered satisfactory (Fig. 8). However, as the black edge surrounding the kidney may serve as a guiding line (Fig. 7), it was then assumed that manual segmentation may be an ideal method for this pulse sequence, so these images were segmented by hand (Fig. 7).

The ellipsoid formula

The ellipsoid formula differs from the three other methods in not being based on segmentation. This formula was applied on the HASTE images (Fig. 9). These values were then entered into the ellipsoid formula: volume = height × width × depth × π/6.

Results

The calculated reproducibility of the rendered kidney volume ranged from 4.0% for VIBE NF/manual to 6.0% for DWI/edge detection (Table 2). The time used to measure one kidney was approximately 1–13 min, with the shortest time for the Haste/Ellipsoid sequence/method and the longest for VIBE FS/Manual. All patients had a small renal cyst each, in the size range of 2–14 mm, except for one patient who also had a second cyst, less than 5 mm in diameter.
Discussion

Methods for measuring and calculating kidney volumes with MR images were devised and evaluated in the absence of a gold standard. These methods varied greatly with respect to reproducibility and time consumption. The quickest alternative was the ellipsoid formula method in combination with the HASTE pulse sequence. This method was at least five times quicker than the other methods, which may compensate for its shortcomings with regard to reproducibility. The manual segmentation method in combination with the pulse sequence VIBE without fat saturation unexpectedly yielded the highest reproducibility. This was possibly due to the simplicity of outlining the black edge that was obtained by acquiring the images with an out-of-phase TE rendering the chemical shift artifact advantages. However, this method required about 7 min of segmentation time, which can be cumbersome.
in clinical practice. The combination of DWI and edge detection yielded the lowest reproducibility. This was unexpected, as the manual involvement was limited to minor corrections (Fig. 6). A plausible explanation might be that these are done under free-breathing, inducing breathing artefacts.

At its best, the VW was quick and simple to use, generating accurate and consistent segmentations. However, if there were adjacent structures of approximately the same gray value, such as the bridge to the psoas muscle in Fig. 3, right, or irregularities, such as the disfiguration caused by a previously performed stem cell transplantation, the situation was more complicated, and extra time was needed for manual corrections. The same can be said about fluctuating average gray values in the stacks that did not succumb to adaptation. In this case, it was either noticed that the segmented part of the kidney became “tighter,” excluding parts of the kidney that were included in other levels in the image stack, producing a segment that was too small, or, conversely, that the segmented part started to “leak” to surrounding tissues and substances, requiring manual revision (Fig. 3, right). The ideal wand may be one where the average of several reference pixels can be sampled in an image (as it the case with Photoshop’s wand). Some kidneys were time-consuming to segment with this method, requiring up to 13 min, and manual revision is probably less exact.

The greatest single limiting factor of this study was probably the fact that MRI results in inconsistent values at all levels: within one kidney, between two kidneys in one slice (Fig. 4), between images in one stack, and between stacks (Fig. 5). An additional reason for gray value inconsistency is the effect of fluid in the intestines, as shown in Fig. 5. A high signal from intestines in the MR image indicates great water content, probably because the patient has drunk fluid shortly before the examination.

The pulse sequence VIBE without fat saturation may be investigated further, in particular in combination with the quick ellipsoid formula, to see if reproducibility would increase. Furthermore, the methodological principles applied should be possible to apply to CT images as well. Finally, in the absence of reference values derived from a gold standard, an animal study could be carried out, with the results from the fluid displacement method serving as reference values.

The present study has several limitations. The number of participants was small and there was no gold standard to compare the results to (as no such gold standard exists). As shown in Table 2, the different segmentation techniques resulted in different volumes, but as there is no gold standard for measurements in vivo, it is not possible to tell which technique is the most accurate. Thus, when evaluating the reproducibility, the patients served as their own reference for each segmentation technique. As no change in kidney volume could be observed after stem cell transplantation (made into the left kidney), both the right and the left kidney were used for the estimation of reproducibility. However, the left kidney was slightly more challenging to delineate due to scars after the surgery appearing between the preoperative imaging and the postoperative scans. From a comparative point of view, it would be interesting to use all segmentation techniques on all pulse sequences. However, some techniques simply resulted in an imbroglio of obviously erroneous segmentations when applied to images from some pulse sequences. The most optimal segmentation technique was therefore used for each pulse sequence. Also, cysts less than 15 mm in diameter were not excluded from the measurements because small cysts only contribute marginally to total volume, while exclusion of those would risk inducing greater operator dependent error and in order to reduce quantification time. Furthermore, the exact time required for segmentation for each stack is only an estimate, as the exact times were not recorded for all sessions.

In conclusion, the results of the various combinations of sequences and methods varied greatly. VIBE without fat saturation in combination with the manual segmentation method gave the best reproducibility results, but manual segmentation was time consuming.

| Table 2. Reproducibility, expressed as coefficients of variation, and estimated average time required to measure one kidney volume. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Sequence / method | DWI/ edge detection | HASTE/ edge detection | VIBE FS/ VW | VIBE NF/ Manual | HASTE/ Ellipsoid |
| Right kidney volume | 203 (58) | 184 (53) | 221 (70) | 202 (52) | 203 (61) |
| Left kidney volume | 175 (47) | 171 (42) | 199 (42) | 180 (40) | 213 (67) |
| CV% | 5.95 % | 4.34% | 4.44% | 4.03% | 4.94 % |
| Estimated time | 6 | 6 | 13 | 7 | 1 |

CV values for the five combinations of pulse sequences/methods, and estimated time required for segmentation, expressed in minutes. CV%, coefficient of variation in percent; DWI, diffusion-weighted imaging; HASTE, half Fourier acquisition single shot turbo spin echo; VIBE FS, volumetric interpolated breath-hold examination with fat saturation; VW, Versatile Wand; VIBE NF, volumetric interpolated breath-hold examination without fat saturation.
Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References
1. Bakker J, Olree M, Kaatee R, et al. In vitro measurement of kidney size: comparison of ultrasonography and MRI. Ultrasound Med Biol 1998;24:683–688.
2. Vasconcelos VT, Katayama RC, Ribeiro MF, et al. Kidney weight and volume among living donors in Brazil. Sao Paulo Med J 2007;125:223–225.
3. Pedersen M, Karstoft K, Lodrup A, et al. Advantages and controversies in the era of intrarenal volumetry. Am J Nephrol 2011;33:40–45.
4. Will S, Martirosian P, Würslin C, et al. Automated segmentation and volumetric analysis of renal cortex, medulla, and pelvis based on non-contrast-enhanced T1-and T2-weighted MR images. MAGMA 2014;27:445–454.
5. Zhou W, Xie Y. Interactive medical image segmentation using snake and multiscale curve editing. Comput Math Methods Med 2013;2013:325903.
6. Zeng J, Zhao Y, Li W, et al. Algorithm for image retrieval based on edge gradient orientation statistical code. ScientificWorldJournal 2014;2014:705763.
7. Stenvinkel P, Wadstrom J, Bertram T, et al. Implantation of autologous selected renal cells in diabetic chronic kidney disease stages 3 and 4—clinical experience of a “first in human” study. Kidney Int Rep 2016;1:105–113.
8. Versatile Wand Tool. Available at: https://imagej.nih.gov/ij/plugins/versatile-wand-tool/index.html (accessed 26 May 2017).
9. Canny Edge Detector: Available at: http://imagej.net/plugins/canny/index.html (accessed 26 May 2017).