Do radiologists agree when reviewing ultrasound examinations performed by a sonographer and a radiologist?

Carina Stenman (1,2), Robert Glavas (1), Ann-Lis Enlund (1), Kjell Jansson (1), Lars Thorelius (3), Örjan Smedby (1,2,4)

1. Department of Radiology and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

2. Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden

3. Telemedicine clinic, 08005 Barcelona, Spain

4. School of Technology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

Corresponding author:

Carina Stenman, Department of Radiology and Department of Medical and Health Sciences, Linköping University, SE-581 85 Linköping, Sweden.

Telephone: +0046101032775

Fax: +46101031799 Email: carina.stenman@liu.se

Abstract

Background: Ultrasound examinations are usually performed by a radiologist or, in suitable cases, by a sonographer or radiographer. Standardized scanning protocols and cine-loop documentation may permit the transfer of tasks from a radiologist to a sonographer or radiographer.

Purpose: To study the diagnostic variability in standardized ultrasound examinations of the kidney by comparing inter-reader agreement between two radiologists who reviewed examinations acquired by a sonographer and a radiologist, as well as inter-operator agreement between the sonographer and the radiologist.

Material and methods: After approval by the local research ethics committee, 98 adult patients, aged from 18 to 92, referred for diagnostic renal sonographic examination and were prospectively enrolled. Both kidneys were imaged using standardized scanning protocols, and the entire examination was documented with cine-loops. Two radiologists reviewed the examinations for different types of pathology, including tumors, cysts, decreased cortical thickness, increased echogenicity and hydronephrosis. Inter-reader and inter-operator agreement was evaluated with kappa coefficient and intra-class correlation.

Results: The most common finding was cysts, which were found in 32 to 40 cases. Tumors were found in three to 10 cases. With one exception, the kappa values for inter-operator agreement (0.65–1.00) were higher than those for inter-reader agreement (0.31–1.00). With two exceptions, no systematic differences between operators or between observers were found.

Conclusion: Using a standardized cine-loop technique, we found slightly better inter-operator agreement than inter-reader agreement. This suggests that it may be easier to exchange an operator than to exchange a reader.

Keywords: cine-loop imaging, renal sonography, agreement

Introduction

Ultrasound imaging is a simple and widely available technique to investigate patients in clinical practice, and is often the first-line imaging modality due to its availability, low cost and lack of ionizing radiation. With improved acquisition techniques and storage capability, the interest in sonographic examinations of the kidneys and other organs has increased substantially (1, 2). It is often suggested that patients should undergo ultrasound imaging as the initial diagnostic imaging test, with further imaging modalities performed on the basis of clinical judgment (3). Ultrasound measurements of kidney volume and size are reliable predictors of renal function in patients with chronic renal disease (4). With modern equipment, it is possible to image an entire organ or a section of an organ, acquiring serial images over time, and store them digitally as cine-loops in a dedicated PACS with the opportunity to replay the entire examination afterwards (5). When played as a cine-loop, the images provide a dynamic impression of the organ or area of interest comparable with real-time (6).

In areas such as obstetric sonographic examinations, the cine-loop technique has several advantages compared with a review of still images (7). According to Pallan et al. (8), the value of reviewing static ultrasound images is very limited.

Cine-loop documentation for ultrasound examinations in different areas has been used in radiology departments, with positive results (1, 7, 9-11). It has also been proposed to transfer, in certain cases, the acquisition of ultrasound examinations from radiologists to radiographers, while the radiologists remain responsible for the interpretation (5).

The aim of this study was to study the diagnostic variability in standardized ultrasound examinations of the kidneys by comparing inter-reader agreement between two radiologists reviewing ultrasound examinations acquired by a sonographer and a radiologist. Also, inter-operator agreement between the sonographer and the radiologist was studied.

Material and Method

The study population consisted of 98 adult patients, aged from 18 to 92 years, who had been referred for diagnostic renal sonographic examination, and who were prospectively enrolled between November 2012 and September 2014. The study was approved by the local research ethics committee. All patients who came for an ultrasound examination of the kidneys with a clinical question, e.g. hydronephrosis, concrements or tumor, were invited to participate in the study. Two weeks prior to the examination, the patients received a letter explaining how the examination was to be performed and that there would be two examiners. All patients gave written informed consent to having two sonographic examinations made by two different examiners using the standardized examination method. Each patient was examined by one sonographer and one radiologist during the same visit and using the same machine. The order in which the sonographer and radiologist performed the examination was alternated randomly. The patients did not fast prior the examination. The sonographer had performed ultrasound examinations for seven years and the radiologist for four years when the study started. The second examiner was always blinded to the results of the first examination.

All sonograms were obtained with GE LOGIC e9 system (GE Healthcare, Medical Systems, Milwaukee, WI) using a convex transducer C1-6 with 3 to 6 MHz. Ultrasound parameters such as gain adjustment, focal zone locations and depth were changed on a case-by-case basis throughout the examination. Each kidney was imaged with the use of the standardized examination method (12). The examination always started with the left kidney in the longitudinal plane, from left to the right, and then continued transversely, from cranial to caudal, with the patient lying in both the supine and decubitus positions, and with four cine-loops of each kidney. After that, the same procedure was performed with the right kidney. When needed, additional cine-loops were recorded, depending on the patient's anatomy. The goal was to cover the kidneys entirely, with some margin. The urinary bladder was also included in the examination. Each of the cine-loops covered 5-10 centimeters in 5-10 seconds.

All examinations were reviewed by two radiologists. The radiologists did not work at the hospital where the examinations were performed, but at two different hospitals in the same county. They had access to the dedicated PACS (Syngo Dynamics, Siemens Medical Systems, Erlangen, Germany) at their own hospital, where all cine-loops were stored. The two radiologists worked independently and without contact with each other during the review. Both radiologists had worked with ultrasound for approximately 15 years and had been working using the standardized method the last 10 years. Thus, they were used to reviewing examinations made by someone else at a workstation. For each reviewed examination, the radiologist filled out a protocol (a separate protocol for each operator) including different types of pathology that might be seen in an ultrasound examination of the kidneys: renal parenchyma (normal or thin), echogenicity (normal or increased), the presence or absence of hydronephrosis, the presence or absence of renal masses, the presence or absence of cysts, and the number and size of the cysts, if present. Also, the size of the kidneys was measured. The protocol included no identifying information other than the medical record number. The protocols were compared for inter-reader and inter-operator agreement.

Statistical analyses

Agreement between readers and operators on findings was assessed as agreement with kappa statistics (13). Inter-operator and inter-reader agreement on measurements was expressed as intra-class correlation (ICC) using a two-way random-effects model and absolute agreement. McNemar's test was used for analyzing systematic differences between readers and between operators. Calculations were made with Bright Stat version 1.2.0 (14).

Results

The 98 patients comprised 45 males and 53 females, range 18 - 92 years old (median 55 years). Out of the 98 patients examined, all were judged to be diagnostic. Examples of the acquired images are shown in Figure 1 (a) and (b), and the frequency of findings is summarized in Table 1.

The most common findings were cysts, which were seen in 32–37 cases in the right kidney and in 34–40 cases in the left kidney, whereas the total number of cysts detected varied from 118 to 137.

Tumors were found in 8–10 cases in the right kidney and in 3–8 cases in the left kidney. Hydronephrosis, the least common finding, was seen on each side in 1-3 cases. As seen in Table 2, the mean measurement results were very similar between operators, readers and left and right kidney.

Inter-reader and inter-operator agreement

In general, the agreement in findings between operators was very high, with kappa values above 0.75 (Table 3), except for hydronephrosis of the right kidney. For this finding, however, only a few cases had a positive result, and the readers agreed in 96-98% of cases. The inter-reader agreement was somewhat lower than the inter-operator agreement, although the confidence intervals overlapped to a great extent (Table 3). Here, the lowest kappa values (below 0.60) were seen for tumor of the left kidney (with operator 2) and decreased cortical thickness. Still, the readers agreed in 95% of cases on tumor of the left kidney (with operator 2), and in 87-90 % on decreased cortical thickness for both operators.

Significant systematic differences between the operators (readers) were found only for the number of cysts, which was greater for operator 1 and reader 1 (Tables 1 and 3).

For the length of kidneys and the sizes of cysts, both the inter-operator and the inter-reader agreement assessed as ICC was high, mostly above 0.80 (Table 4).

Discussion

For most of the diagnostic findings in this study, the agreement between operators was almost perfect (Table 3). On the other hand, between readers the agreement tended to be somewhat lower. The results suggest that when ultrasound examinations are performed with a standardized technique and cine-loop documentation, the operator's profession may be less relevant, as long as the reader's competence is high. With the standardized technique with stored cine-loops, it thus seems feasible to replace a radiologist examiner with a radiographer or sonographer if the interpretation of images is performed by a radiologist.

Most important is agreement about severe pathology such as tumors. For tumors in the left kidney, examined by operator 2, the rather low inter-reader kappa value of 0.52 (Table 3) was combined with a high agreement percentage (95%), which suggests that the kappa value may reflect imbalanced results for a rather rare finding. Still, the fact that the readers disagreed in five cases out of 98 should be borne in mind. The same explanation is likely for inter-reader agreement regarding hydronephrosis in the right kidney, where the kappa value was 0.31 for reader 1 and 0.65 for reader 2, but the number of cases with actual disagreement was low. Other possible explanations include differences in how readers judge hydronephrosis and the fact that multiple cysts can be confused with general or even local hydronephrosis.

Also, for quantitative measurements of the diameter of cysts and the length of the kidneys (Table 4), both the inter-operator and the inter-reader agreement was high, with ICC values of 0.75 – 0.96. In the sonographer's examinations, a few more cysts were seen than in those made by a radiologist. For all other findings, the operators agreed well.

In the study by Gaarder et al. it was shown that standardized cine-loop ultrasound in renal ultrasound is a highly reliable method, and allows for a skill mix between radiographer and radiologist (11). Their study also demonstrated that examinations of the kidneys acquired by a trained radiographer have similar diagnostic value to examinations performed by a radiologist. Dormagen et al. found that ultrasound scans can be evaluated reliably later by a non-performing reader (9).

An advantage of the cine-loops is that they can be shared electronically via a secure network and the interpreting physician can request expert consultation (7). With extensive recent technical improvements, the quality of ultrasound examinations has increased substantially (2). Cine-loops are said to be comparable with real-time examinations (6). Van Holsbeke et al. found that ultrasound examiners tend to be more confident in excluding malignancy during real-time scans than when looking at static ultrasound images of adnexal masses. When looking at static images, the dynamic aspect is lost and some pathology can be misinterpreted (6).

Today, the most common ultrasound recording technique is to store static images instead of cine-loops. When working exclusively with static images, differences between examiners are of great importance, in particular when the examiner is the one who writes the report. Batch reading of static images by a radiologist acquired by a sonographer may lead to the patient being called back for additional imaging to be carried out (15).

It has been seen in other studies that evaluation of static images is associated with lower diagnostic specificity (1, 7). In cases of uncertainty, the patient may have to be rescanned by a supervising sonologist if the first examiner is less experienced. This can cause unnecessary anxiety for patients. In such situations, using such a standardized method with cine-loops could therefore be favorable.

In the current study, the agreement between readers was somewhat lower than that between operators. This suggests that it may be more difficult to replace a reviewer than to replace an examiner.

For the examiner, it is important to produce images and cine-loops of high quality in order to enable the reviewer to make a correct diagnosis of the patient. The high agreement between examiners in this study suggests that examiners might be replaceable. With the right training and education, different professions could perform ultrasound in an acceptable way. The standardized ultrasound method with stored cine-loops may enable the sonographer or radiographer to take over some of the work traditionally performed by radiologists. The purpose is not to replace the radiologists, but rather for the professions to work in close collaboration, in order to provide both improved care and improved patient outcomes in situations where radiologists are not available to interpret ultrasound images in directly after the examination.

It is important to follow a predefined standardized scanning acquisition pattern when the series of scans are performed. This also facilitates parallel comparison of a new and an old examination when reviewing at the workstation several times, regardless of who carried out the previous examination (12). This might offer the reviewer the opportunity to observe findings that were not noted at the time of examination. In this way, ultrasound will be more like other radiological modalities.

In other areas of radiology, where radiographers have taken over some of the radiologists' duties, benefits have been claimed in terms of reduced patient waiting times, freeing up of radiologists for other duties, cost-effectiveness, and greater potential for recruitment and retention of radiographers, with higher levels of job satisfaction (16, 17). Also, with regard to the shortage of radiologists, it has been seen in some hospitals that non-physician personnel have been trained to reduce the radiologist's workload (18). When different professions cooperate, better care is provided, and it also helps to maintain quality assurance for ultrasound (19). The role of the radiographer is usually to perform, not to interpret, the examination. However, it has been shown that radiographers are able to differentiate negative from positive findings in routine abdominal ultrasonography, with an accuracy level similar to experienced radiologists (20).

With increasing demands for medical imaging and a growing lack of radiologists' time, examinations performed by sonographers or radiographers may be a useful alternative.

A strength of this study was that the readers did not work together and therefore could not influence each other's practical skills. Nevertheless, they had almost similar results when reviewing the examinations at a workstation. On the other hand, a possible limitation of the study could be that the radiologist and the sonographer had been working together and had used the same standardized method for some years and thus might have influenced each other's practical skills when examining patients. It remains to be seen whether similar high values of agreement can be found between operators who have not been trained in the same environment.

It would be relevant to compare this standardized ultrasound method with another modality within radiology, such as Computed Tomography or Magnetic Resonance Imaging. A larger number of patients, in particular patients with positive findings, would also be desirable.

In conclusion, this study shows a slightly better inter-operator than inter-reader agreement when using the standardized cine-loops technique. This might indicate that the operator is more exchangeable than the reader.

References

1. Jandzinski D, van Wijngaarden E, Dogra V, et al. Renal sonography with 2-dimensional versus cine organ imaging: preliminary results. *J ultrasound med* 2007;26:635-44.

2. Robbin ML, Lockhart ME, Weber TM, et al. Ultrasound quality and efficiency: how to make your practice flourish. *J ultrasound med* 2011;30:739-43.

3. Smith-Bindman R, Aubin C, Bailitz J, et al. Ultrasonography versus computed tomography for suspected nephrolithiasis. *N Engl J Med* 2014;371:1100-10.

4. Zanoli L, Romano G, Romano M, et al. Renal function and ultrasound imaging in elderly subjects. *ScientificWorldJournal*. 2014;830649.

5. Stenman C, Thorelius L, Knutsson A, et al. Radiographer-acquired and radiologist-reviewed ultrasound examination--agreement with radiologist's bedside evaluation. *Acta Radiol* 2011;52:70-4.

6. Van Holsbeke C, Yazbek J, Holland TK, et al. Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses. *Ultrasound Obstet Gynecol* 2008;32:828-31.

7. Scott TE, Jones J, Rosenberg H, et al. Increasing the detection rate of congenital heart disease during routine obstetric screening using cine loop sweeps. *J ultrasound med* 2013;32:973-9.

8. Pallan M, Linnane J, Ramaiah S. Evaluation of an independent, radiographer-led community diagnostic ultrasound service provided to general practitioners. *J Public Health* (Oxf). 2005;27:176-81.

9. Dormagen JB, Gaarder M, Drolsum A. Standardized cine-loop documentation in abdominal ultrasound facilitates offline image interpretation. *Acta Radiol* 2015;56(1):3-9.

10. Bude RO, Fatchett JP, Lechtanski TA. The use of additionally trained sonographers as ultrasound practitioners: our first-year experience. *J ultrasound med* 2006;25:321-7.

11. Gaarder M, Seierstad T, Soreng R, et al. Standardized cine-loop documentation in renal ultrasound facilitates skill-mix between radiographer and radiologist. *Acta Radiol* 2015;56(3):368-73.

12. Thorelius L. Sonodynamics fundamentals 2014 [cited 2014 27 February]. Available from: www.sonodynamics.com.

13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.

14. Stricker D. BrightStat.com: free statistics online. Comput Methods Programs *Biomed* 2008;92(1):135-43.

15. Rosen MP, Levine D, Carpenter JM, et al. Diagnostic accuracy with US: remote radiologists' versus on-site radiologists' interpretations. *Radiology* 1999;210:733-6.

16. Smith TN, Baird M. Radiographers' role in radiological reporting: a model to support future demand. *Med J Aust* 2007;186:629-31.

17. Field LJ, Snaith BA. Developing radiographer roles in the context of advanced and consultant practice. *J Med Radiat Sci* 2013;60:11-5.

18. Lockhart ME, Robbin ML, Berland LL, et al. The sonographer practitioner: one piece to the radiologist shortage puzzle. *J ultrasound med* 2003;22:861-4.

19. Ramsay PA, Fracchia CA. The role of the sonologist in a sonographer-based practice. *Australas Radiol* 1999;43:20-6.

20. Leslie A, Lockyer H, Virjee JP. Who should be performing routine abdominal ultrasound? A prospective double-blind study comparing the accuracy of radiologist and radiographer. *Clin Radiol* 2000;55:606-9.

Table 1. Frequency of findings for different operators and readers

	Operator 1 (radiographer)		Operator 2 (radiologist)	
Finding	Reader 1	Reader 2	Reader 1	Reader 2
Tumor, right kidney	10	8	10	8
Tumor, left kidney	7	4	8	3
Decreased cortical thickness, right kidney	20	19	20	19
Decreased cortical thickness, left kidney	20	20	19	19
Increased echogenicity, right kidney	8	10	9	10
Increased echogenicity, left kidney	8	10	8	10
Presence of cysts, right kidney	36	32	37	32
Presence of cysts, left kidney	40	34	35	34
Number of cysts, right kidney; median (min, max)	60; 1(1, 7)	63; 1(1, 7)	62; 1(1, 7)	55; 1(1, 7)
Number of cysts, left kidney; median (min, max)	[a,b] 77; 1(1, 7)	[b] 63; 1(1, 7)	[a] 66; 1(1, 7)	63; 1(1, 7)
Hydronephrosis, right kidney	3	3	3	3
Hydronephrosis, left kidney	1	2	1	2

[a] Significantly higher number for operator 1 (p=0.01)

[b] Significantly higher number for reader 1 (p=0.03)

All other values did not have a significant difference.

Table 2. Measurement results for different operators and readers. Mean (SD)

	Operator 1 (radiographer)		Operator 2 (radiologist)	
Measure	Reader 1	Reader 2	Reader 1	Reader 2
Length of right kidney cm	10.6 (1.15)	10.7 (1.17)	10.5 (1.63)	10.5 (1.19)
Length of left kidney cm	10.8 (1.08)	10.8 (1.60)	10.8 (1.68)	10.8 (1.66)
Size of largest cysts, right kidney cm	1.04 (1.60)	0.9 (1.65)	1.06 (1.60)	0.9 (1.49)
Size of largest cysts, left kidney cm	1.12 (1.59)	1.03 (1.65)	1.11 (1.79)	1.03 (1.65)

Table 3. Inter-operator and inter-reader agreement for findings

Finding	Inter-operator agreement kappa (95% confidence limits)		Inter-reader agreement kappa (95% confidence limits)	
	Reader 1	Reader 2	Operator 1 (radiographer)	Operator 2 (radiologist)
Tumor, right kidney	1.00 (0.53-1.00)	1.00 (0.58-1.00)	0.87 (0.43-1.00)	0.87 (0.43-1.00)
Tumor, left kidney	0.78 (0.12-1.00)	1.00 (0.29-1.00)	0.71 (0.14-1.00)	0.52 (-0.04-1.00)
Decreased cortical thickness, right kidney	1.00 (0.70-1.00)	1.00 (0.71-1.00)	0.58 (0.29-0.87)	0.58 (0.29-0.87)
Decreased cortical thickness, left kidney	0.97 (0.68-1.00)	0.97 (0.68-1.00)	0.68 (0.40-0.97)	0.61 (0.68-0.91)
Increased echogenicity, right kidney	0.93 (0.48-1.00)	1.00 (0.58-1.00)	0.75 (0.31-1.00)	0.82 (0.40-1.00)
Increased echogenicity, left kidney	1.00 (0.58-1.00)	1.00 (0.58-1.00)	0.88 (0.44-1.00)	0.88 (0.44-1.00)
Presence of cysts, right kidney	0.89 (0.68-1.00)	1.00 (0.77-1.00)	0.64 (0.42-0.85)	0.75 (0.54-0.97)
Presence of cysts, left kidney	0.89 (0.68-1.00)	1.00 (0.78-1.00)	0.74 (0.53-0.95)	0.70 (0.49-0.92)
Number of cysts, right kidney †	0.92 (0.69-1.00)	0.98 (0.72-1.00)	0.61 (0.35-0.87)	0.68 (0.42-0.94)
Number of cysts, left kidney †	[a] 0.88 (-0.64-1.00)	1.00 (0.75-1.00)	[b] 0.77 (0.53-1.00)	0.76 (0.52-1.00)
Hydronephrosis, right kidney	0.65 (-0.13-1.00)	1.00 (0.21-1.00)	0.31 (-0.47-1.00)	0.65 (-0.13-1.00)
Hydronephrosis, left kidney	1.00 (0.71-1.00)	1.00 (0.30-1.00)	0.66 (-0.46-1.00)	0.66 (-0.46-1.00)

†) Weighted kappa

[a] Significantly higher number for operator 1 (p=0.01)

[b] Significantly higher number for reader 1 (p=0.03)

All other values did not have a significant difference.

Table 4. Inter-operator and inter-reader agreement for measurements, expressed as intra-class correlation (ICC)

	Inter-operator agreement ICC (95% confidence limits)		Inter-reader agreement ICC (95% confidence limits)	
Measure	Reader 1	Reader 2	Operator 1 (radiographer)	Operator 2 (radiologist)
Length of right kidney (cm)	0.86 (0.80-0.90)	0.89 (0.83-0.93)	0.90 (0.85-0.93)	0.91 (0.86-0.94)
Length of left kidney (cm)	0.78 (0.69-0.85)	0.94 (0.91-0.96)	0.81 (0.72-0.87)	0.95 (0.93-0.97)
Size of largest cysts, right kidney (cm)	0.96 (0.94-0.97)	0.94 (0.92-0.96)	0.80 (0.71-0.86)	0.84 (0.87-0.90)
Size of largest cysts, left kidney (cm)	0.82 (0.74-0.87)	0.89 (0.83-0.92)	0.90 (0.84-0.92)	0.75 (0.65-0.82)

All ICC significant at $p < 0.001$ level.

Figure legends

Fig. 1. a, an image from a cine-loop in a longitudinal sweep of the left kidney with a small cyst, examined by a sonographer. b Same patient, view and kidney, examined by a radiologist.

(a) 

(b)