Interobserver Variability in Interpretation of Planar and SPECT Tc-99m-DMSA Renal Scintigraphy in Children

Nermina Beslic¹, Renata Milardovic¹, Amira Sadija¹, Lejla Džananovic², Semra Cavaljuga²

¹Clinic of Nuclear Medicine, University-Clinical Center Sarajevo, Sarajevo, BiH

²Department of Epidemiology and Biostatistics, Faculty of Medicine, University of Sarajevo, Sarajevo, BiH

Corresponding author: prof Nermina Beslic, MD, PhD. Clinic for Nuclear medicine, University Clinical Center, Sarajevo, Bosnia and Herzegovina.

doi: 10.5455/aim.2017.25.28-33

ACTA INFORM MED. 2017 MAR; 25(1): 28-33

Received: Jan 07, 2017 • Accepted: Mar 08, 2017

ABSTRACT

Objective: This study objective was to evaluate interobserver agreement between individual pairs of three nuclear medicine physicians in interpretation of renal cortical scintigraphy in children with respect to the mode of acquisition (planar vs. SPECT), diagnoses and kidney site (left vs. right). **Materials and Methods:** Thirty children were imaged in planar and SPECT mode per protocol upon the injection of Tc-99m DMSA dose adjusted to their body weight. Patients were classified according to diagnoses into four groups. Three nuclear medicine physicians interpreted the findings blindly and independently. Renal defects were interpreted as focal and diffuse, per three renal segments. For the raters we calculated simple percentage agreement, the Cohen kappa statistic with 95% confidence intervals, and the overall kappa defining the levels of reliability as almost perfect or perfect, substantial, moderate, fair and slight agreement. **Results:** Interobserver agreement in planar interpretation was 77,2% (kappa=0.59; 95% confidence interval, 0.41 to 0.75) and SPECT 72,9% (kappa= 0,57; 95% confidence interval, 0,41 to 0,72). In planar interpretation, all individual pairs had moderate agreements except one that had a substantial agreement. In SPECT, all the pairs had moderate agreements except one that had an almost perfect agreement. Overall agreement per kidney site was on planar 73,4% for the left (kappa=0,54, moderate agreement), and 81,1% for the right kidney (kappa 0,63, substantial agreement). On SPECT, there was 72,2% agreement for the left (kappa=0,59, mode rate agreement), and 73,7% for the right kidney (kappa=0,54, moderate agreement). Overall agreement per diagnoses ranged from 70-88,9% on planar (kappa= -0,04 to 0,79), and 50-100% on SPECT (kappa=-0,02-1,000) indicating agreements from slight to substantial. **Discussion:** Our results suggest acceptable levels of interobserver agreement in all individual pairs of raters with respect to the mode of acquisition (planar vs. SPECT), diagnoses and kidney site (left vs. right). For the mode of acquisition, we would recommend hybrid imaging SPECT/ CT method to be used whenever possible in the detection of renal cortical defects on Tc-99m-DMSA scintigraphy. **Keywords:** Interobserver variability, Tc-99m-DMSA, cortical scintigraphy, planar, SPECT.

1. INTRODUCTION

Renal cortical scintigraphy with Tc-99m - dimercaptosuccinic acid (DMSA) is a well established and highly sensitive imaging method for the detection of renal parenchymal lesions. It is used in nuclear medicine pediatric practice to assess for urinary tract infections with or without vesicourethral reflux (VUR), which may lead to pyelonephritis. It helps diagnose pyelonephritis and its sequelae in a timely manner so that proper treatment options could be tailored (1-4).

Specificities of the kidneys relate to rather heterogenous physiologic uptake of radiotracer, which creates higher contrast resolution and better delineation of the lesions (5, 6). Specificities

of pediatric kidneys relate to developmental changes and smaller kidney size in general.

While planar renal cortical scintigraphy was the first to be applied, the introduction of multiple-headed gamma cameras enabled newer modes of acquisition in order to provide larger dataset for reconstruction and analysis. Together with the use of ultra-high energy collimators, it resulted in better spatial resolution and image contrast that was expected to be attributed to for diagnostic difference.

2. MATERIALS AND METHODS

A prospective study included 30 children and 60 renal units, 25 females and 5 males. The patients aged from three

© 2017 Nermina Bešlić, Renata Milardović, Amira Šadija, Lejla Džananović, Semra Cavaljuga

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

PLANAR

		Rater 1						Rater 3					
Rater 2		0	1F	2F	D	D+1F	D+2F	0	1F	2F	D	D+1F	D+2F
	0	36* (18**,18***)	1 (0,1)		1 (0,1)			33 (16,17)	3 (2,1)	1 (1,0)		1 (0,1)	
	1F	5 (2,3)	4 (2,2)			1 (1,0)		3 (1,2)	6 (2,4)				1 (1,0)
	2F		3 (2,1)						3 (2,1)				
	D				2 (0,2)			1 (0,1)			1 (0,1)		
	D+1F					4 (2,2)					2 (1,1)	2 (1,1)	
Rater 3	D+2F			1 (1,0)		1 (1,0)	1 (1,0)			1 (1,0)	1 (1,0)		1 (1,0)
	0	36 (17,19)			1 (0,1)								
	1F	4 (3,1)	8 (3,5)										
	2F	1 (1,0)		1 (1,0)									
	D				1 (0,1)	2 (1,1)	1 (1,0)						
	D+1F				1 (0,1)	2 (1,1)							
	D+2F					1 (1,0)	1 (1,0)						

Table 1. PLANAR readings of 60 kidneys (30 left and 30 right kidneys) from individual pairs of three observers

Obs	Rater	Number of agreement	Total tasks	Percentage of agreement	Kappa agreement	Lower 95% CI [§] Kappa	Upper 95% CI [§] Kappa
1	1 vs. 2	47* (23**, 24***)	60 (30,30)	0.783 (0.767, 0.800)	0.595 (0.586, 0.600)	0.422 (0.370, 0.323)	0.768 (0.802, 0.877)
2	1 vs. 3	49 (23, 26)	60 (30,30)	0.817 (0.767, 0.867)	0.662 (0.593, 0.739)	0.501 (0.365, 0.522)	0.824 (0.821, 0.956)
3	2 vs. 3	43 (20, 23)	60 (30,30)	0.717 (0.667, 0.767)	0.500 (0.445, 0.559)	0.329 (0.227, 0.297)	0.671 (0.664, 0.821)
4	Overall	46.333 (22, 24.333)	60 (30,30)	0.772 (0.734, 0.811)	0.586 (0.541, 0.633)	0.417 (0.321, 0.381)	0.754 (0.762, 0.885)

* data for both units (kidneys)

** data for left kidneys

*** data for right kidneys

§ Confidence Interval

months to 16 years, with a mean of 5.4 years. All patients were referred for evaluation by pediatric nephrologists.

Patients had the history of VUR and/or other morphological changes 9, VUR and urinary tract infection 12, diagnoses other than VUR (such as hydronephrosis, renal abscess or polycystic kidney) 5, and no changes/normal findings 4.

Tc-99m-DMSA was applied intravenously in the dose 40–150 MBq based on the body weight with imaging performed after three hours. If needed, chloral-hydrate for sedation was applied rectally in the dose of 50 mg/kg. For each patient, first planar and in sequence SPECT images were acquired with a dual-head gamma camera (Siemens E.cam, Siemens Healthcare Global, Germany) equipped with a LEHR collimator set at 140 keV with a 20% energy window. Zoom between 1,45–2,25 was used adjusted to the body size. Planar images were acquired in a 256x256 matrix, in the anterior/posterior and right/left lateral views for 5 minutes or 500 000 counts each. For SPECT images, a dual-head gamma camera was rotating clockwise, in a body contour orbit with 180 degrees of rotation, step and shoot mode, 40 views/head with time per view of 40 seconds. Data were acquired in a 128x128 matrix. Upon reconstruction, the transaxial, sagittal and coronal tomographic slices were displayed.

Datasets of planar and SPECT images for each patient were interpreted by three nuclear medicine physicians separately in random order blinded of diagnosis or other findings.

Parenchymal changes were interpreted as (a) focal or (b) diffuse. Focal defects were classified as single or multiple. For localization of defects, each kidney is divided into three segments: (a) upper pole, (b) mid-kidney and (c) lower pole.

Data management and analyses were performed using MS

Office Excel 2010. To assess the inter-rater reliability between pairs of raters (1 and 2, 2 and 3, and 1 and 3) with regard to the findings on planar and SPECT images, we calculated simple percentage agreement, the Cohen kappa statistic with 95% confidence intervals, and overall kappa (as the arithmetic mean of individual pair's coefficients as suggested by Light (1971). Interpretations of the κ statistic were based on the criteria described by Landis and Koch (1977), meaning the level of reliability was defined as follows: κ values of 0.81 to 1.00 indicating almost perfect or perfect agreement; 0.61 to 0.80 – substantial agreement; 0.41 to 0.60 – moderate agreement; 0.21 to 0.40 – fair agreement; and 0.01 to 0.20 – slight agreement.

3. RESULTS

We analysed a total of 60 kidneys/renal units (30 left kidneys and 30 right kidneys) in 30 subjects.

In planar studies, all three observers interpreted the findings for all 60 kidneys. Percentage of agreement between three observers ranged from 0.717 (observers 2 and 3) to 0.817 (observers 1 and 3), with the overall percentage agreement of 0.772. As indicated in Table 1, the Kappa coefficient ranged from 0.500 (observers 2 and 3) to 0.662 (observers 1 and 3), with an overall inter-rater reliability (IRR) between observer 1, 2, and 3 of 0.586 (average of 0.595, 0.662, and 0.500). According to Landis and Koch, all the pairs had moderate agreements except that observers 1 and 3 had a substantial agreement.

When agreement was calculated with respect to the kidney site (left/right kidney), percentage of agreement ranged from 0.667 (observers 2 and 3 on left kidneys) to 0.867 (observers

		Rater 1							Rater 3						
Rater 2		0	1F	2F	D	D+1F	D+2F		0	1F	2F	D	D+1F	D+2F	
	0	3*	9#	1					2	8	2		1		
	1F	1	2	1					9	14	1				
	2F	2	2	1						3	1				
	D			2					1	2	2			1	
	D+1F														
	D+2F														
Rater 3		0	1F	2F	D	D+1F	D+2F		0	1F	2F	D	D+1F	D+2F	
	0	2	8						2	8	2		1		
	1F	10	16						9	14	1				
	2F	2	3	3						3	1				
	D	1	1	2					1	2	2			1	
	D+1F														
	D+2F														

Table 2. PLANAR readings of 60 kidneys (with respect to diagnoses) from individual pairs of three observers

Obs	Rater	Number of agreement				Totaltasks				Percentage of agreement				Kappa agreement				Lower 95% CI [§] Kappa				Upper 95% CI [§] Kappa			
		*	#	°	⊠	*	#	°	⊠	*	#	°	⊠	*	#	°	⊠	*	#	°	⊠	*	#	°	⊠
1	1 vs. 2	7	12	13	15	9	16	17	18	0.778	0.750	0.765	0.833	0.654	0.614	0.611	-0.038	0.224	0.363	0.313	-0.091	1.000	0.865	0.910	0.014
2	1 vs. 3	7	13	13	16	10	15	17	18	0.700	0.867	0.765	0.889	0.595	0.786	0.570	0.308	0.237	0.529	0.263	0.010	0.952	1.000	0.876	0.605
3	2 vs. 3	7	10	12	14	10	14	17	19	0.700	0.714	0.706	0.737	0.595	0.517	0.514	0.078	0.237	0.210	0.230	-0.200	0.952	0.825	0.798	0.356

* VUR and/or other morphological changes
 # VUR and infection
 ° Diagnoses other than VUR
 ⊠ No changes – normal findings

1 and 3 on right kidneys), with the overall percentage agreement for left kidneys of 0.734, and for right kidneys of 0.811. The Kappa coefficient ranged from 0.445 (observers 2 and 3 on left kidneys) to 0.739 (observers 1 and 3 on right kidneys), with an overall IRR between observer 1, 2, and 3 for left kidneys of 0.541 (average of 0.445, 0.586, and 0.593) – indicating moderate agreement between observers, and for right kidneys of 0.633 (average of 0.559, 0.600, and 0.739) – indicating substantial agreement.

After classification in 4 categories regarding the clinical diagnoses (Category *: VUR and/or other morphological changes; Category #: VUR and infection; Category °: Diagnoses other than VUR; Category ⊠: No changes – normal findings), percentage of agreement ranged from 0.700 (observers 1 and 3 on units diagnosed with VUR and/or other morphological changes, and observers 2 and 3 on units with same diagnosis) to 0.889 (observers 1 and 3 on units with normal findings). As indicated in Table 2, the Kappa coefficient ranged from -0.038 (observers 1 and 2 on units with normal findings indicating the presence of the agreement worse than expected, or disagreement) to 0.786 (observers 1 and 3 on units diagnosed with VUR and infection indicating the substantial agreement, according to Landis and Koch). Overall IRR between observer 1, 2, and 3 was smallest on units with normal findings: $\kappa=0.116$ (average of 0.308, 0.078, and -0.038) indicating slight agreement, and largest on units diagnosed with VUR and infection: $\kappa=0.639$ (average of 0.786, 0.614, and 0.517) indicating substantial agreement.

In SPECT studies, observers 1 and 2 interpreted the findings for all 60 kidneys, while observer 3 read 59 units (one right kidney was classified as “uninterpretable”). Percentage of agreement between three observers ranged from 0.644

(observers 2 and 3) to 0.883 (observers 1 and 2), with the overall percentage agreement of 0.729. As indicated in Table 3, the Kappa coefficient ranged from 0.429 (observers 2 and 3) to 0.820 (observers 1 and 2), with an overall IRR between observer 1, 2, and 3 of 0.567 (average of 0.429, 0.452, and 0.820). According to Landis and Koch, all the pairs had moderate agreements except that observers 1 and 2 had an almost perfect agreement.

When agreement was calculated with respect to the kidney site (left/right kidney), percentage of agreement ranged from 0.633 (observers 2 and 3 on left kidneys) to 0.900 (observers 1 and 2 on right kidneys), with the overall percentage agreement for left kidneys of 0.722, and for right kidneys of 0.737.

Figure 1. A 12-year-old girl with urinary tract infection. Ultrasound revealed chronic inflammatory changes in both kidneys without signs of dilatation of the collecting systems. Tc-99m-DMSA planar imaging revealed normal right kidney. Left kidney showed diffusely reduced uptake of radiotracer, particularly in the poles. The findings were consistent with the inflammatory changes.

SPECT

		Rater 1						Rater 3					
Rater 2	0	32* (17**,15***)						27 (13,14)	4 (3,1)	1 (1,0)			
	1F	1 (0,1)	10 (3,7)	1 (1,0)		1 (1,0)		6 (1,5)	6 (3,3)			1 (1,0)	
	2F		1 (1,0)	2 (1,1)					1 (1,0)	1 (1,0)			
	D				2 (0,2)	1 (0,1)		1 (0,1)	1 (0,1)		1 (0,1)		
	D+1F					3 (2,1)	1 (0,1)	1 (0,1)			1 (1,0)	2 (1,1)	
	D+2F					1 (1,0)	4 (3,1)			3 (3,0)	1 (0,1)		1 (1,0)
Rater 3	0	28 (13,15)	5 (1,4)		1 (0,1)		1 (0,1)						
	1F	4 (3,1)	6 (3,3)	1 (1,0)	1 (0,1)								
	2F	1 (1,0)		1 (1,0)		1 (1,0)	2 (2,0)						
	D					2 (1,1)	1 (0,1)						
	D+1F					3 (2,1)							
	D+2F						1 (1,0)						

Table 3. SPECT readings from individual pairs of three observers

Obs	Rater	Number of agreement	Total tasks	Percentage of agreement	Kappa agreement	Lower 95% CI [§] Kappa	Upper 95% CI [§] Kappa
1	1 vs. 2	53* (26**, 27***)	60 (30,30)	0.883 (0.867, 0.900)	0.820 (0.788, 0.848)	0.703 (0.614, 0.691)	0.938 (0.963, 1.000)
2	1 vs. 3	39 (20, 19)	59 (30,29)	0.661 (0.667, 0.655)	0.452 (0.511, 0.375)	0.280 (0.291, 0.119)	0.624 (0.730, 0.631)
3	2 vs. 3	38 (19, 19)	59 (30,29)	0.644 (0.633, 0.655)	0.429 (0.458, 0.393)	0.253 (0.240, 0.123)	0.605 (0.676, 0.663)
4	Overall	43.333 (21.667, 21.667)	59.333 (30,29.333)	0.729 (0.722, 0.737)	0.567 (0.586, 0.539)	0.412 (0.382, 0.311)	0.722 (0.790, 0.765)

* Data for both units (kidneys)

** Data for left kidneys

*** Data for right kidneys

§ Confidence Interval

The Kappa coefficient ranged from 0.375 (observers 1 and 3 on right kidneys) to 0.848 (observers 1 and 2 on right kidneys), with an overall IRR between observer 1, 2, and 3 for left kidneys of 0.586 (average of 0.458, 0.511, and 0.788), and for right kidneys of 0.539 (average of 0.393, 0.375, and 0.848), indicating moderate agreement between observers.

After classification in 4 categories regarding the clinical diagnoses (Category *: VUR and/or other morphological changes; Category #: VUR and infection; Category °: Diagnoses other than VUR; Category □: No changes – normal findings), percentage of agreement ranged from 0.500 (observers 2 and 3 on units diagnosed with VUR and/or other morphological changes) to 1.000 (observers 1 and 2 on units diagnosed with VUR and infection, and units with normal

findings). As indicated in Table 4, the Kappa coefficient ranged from -0.019 (observers 1 and 3 on units with normal findings, as well as observers 2 and 3 on units with normal findings, indicating the presence of the agreement worse than expected, ordiagreement) to 1.000 (observers 1 and 2 on units diagnosed with VUR and infection, and units with normal findings, indicating the perfect agreement, according to Landis and Koch). Overall IRR between observer 1, 2, and 3 was smallest on units with normal findings: $\kappa=0.321$ (average of 1.000, -0.019, and -0.019) indicating fair agreement, and largest on units diagnosed with VUR and infection: $\kappa=0.629$ (average of 1.000, 0.444, and 0.444) indicating substantial agreement.

4. DISCUSSION

Renal cortical scintigraphy with Tc-99m-DMSA has been established as the main diagnostic imaging method for detection of renal defects caused by pyelonephritis, or its sequelae (cortical scarring and shrunken kidney) (7). There has been a number of studies to evaluate for the interobserver agreement in the interpretation of renal cortical studies. It is of a huge importance because the results directly tailor further management of the patients. Renal parenchymal disease evolve through the phases affecting scintigraphic appearances, so it is also important to take into account the timing of imaging. Studies have different designs, and most authors report on satisfying interobserver agreement.

Our study was designed to evaluate for interobserver agreement in relation to the mode of acquisition (planar vs. SPECT), four categories of diagnoses (VUR and/or other morphological changes; VUR and infection; diagnoses other than VUR; no changes – normal findings), and the kidney

Figure 2. Tc-99m-DMSA SPECT images of the same patient performed consequently to planar imaging, depicted in detail the inflammatory changes in the left kidney. Affection of the lower pole with disappearance of the renal contour was prominent.

		Rater 1							Rater 3						
Rater 2		0	1F	2F	D	D+1F	D+2F		0	1F	2F	D	D+1F	D+2F	
	0	*2 7	*8 15						1 7	1 12		1 3			
	1F	1		5 2	2 1			1	2 1	3 2	1		1		
	2F			1	1						1				
	D					1 1	1					1 1			
	D+1F					2	1		1 1				2		
	D+2F					1		4			1	2	1		1

Rater 3		0	1F	2F	D	D+1F	D+2F	
	0	1 8	7 12	2 2	1 1			1
	1F	1		3 2	1			
	2F		1			1		2
	D					1 1		1
	D+1F					1 2		
	D+2F						1	

Table 4. SPECT readings (with respect to diagnoses) from individual pairs of three observers

Obs	Rater	Number of agreement				Total tasks				Percentage of agreement				Kappa agreement				Lower 95% CI [§] Kappa				Upper 95% CI [§] Kappa			
		*	#	°	¤	*	#	°	¤	*	#	°	¤	*	#	°	¤	*	#	°	¤	*	#	°	¤
1	1 vs. 2	9	15	11	18	10	15	17	18	0.900	1.000	0.647	1.000	0.841	1.000	0.517	1.000	0.548	1.000	0.254	1.000	1.000	1.000	0.779	1.000
2	1 vs. 3	6	9	12	12	10	14	17	18	0.600	0.643	0.706	0.667	0.437	0.444	0.560	-0.019	0.008	0.157	0.290	-0.315	0.865	0.731	0.829	0.278
3	2 vs. 3	5	9	12	12	10	14	17	18	0.500	0.643	0.706	0.667	0.265	0.444	0.577	-0.019	-0.172	0.157	0.294	-0.315	0.702	0.731	0.860	0.278

- * VUR and/or other morphological changes
- # VUR and infection
- ° Diagnoses other than VUR
- ¤ No changes – normal findings

side (left vs. right).

Erdogan et al comment that the interobserver variability is one of the important indicators of the reliability of a test (8). For this reason, some kind of standardization of interpretation is also necessary. The authors claim anatomical variations of the kidneys, different experiences of the readers and the severity of renal lesions as the reasons for interobserver variability. Our readers all work in the same nuclear medicine department and at present read the same number of DMSA scans, however, their length of experience differs from ten to twenty years.

Anatomical variations of the kidneys, especially age-related, have been cited by many studies as the source of disagreement. According to Tondeur et al, normal variants such as pear-shaped kidney, hypoactive poles contrasting with important parenchymal mass, triangular kidney and unusual shape of columns of Bertin are amongst the main causes of disagreement (10). Craig et al commented on technically sub-optimal/blurred images in newborns due to poor uptake of radiotracer, and perceived lesions due to normal or exaggerated anatomic structures, such as the pelvicalyceal system (11). Admitting to immature kidneys of the newborns and prominent columns of Bertin in some cases in our study, Observer 3 even rated one neonate kidney as totally uninterpretable due to the diffusely poor uptake of radiotracer.

With respect to the mode of acquisition, our study demonstrated overall agreement on planar imaging 77,2% with a kappa of 0,586 and 72,9% on SPECT with a kappa of 0,567. Knowing that a kappa of 1 indicates perfect agreement, whereas a kappa of 0 indicates agreement equivalent to chance, kappa values for planar and SPECT are in the mod-

erate agreement range between the raters. In our opinion, lower overall agreement for SPECT can be attributed to its lower specificity in comparison to planar imaging. Earlier was postulated that SPECT would create opportunity for reduced specificity due to false-positive findings caused by its enhanced spatial and contrast resolution. Also, it is widely recognized that the kidneys have rather heterogenous uptake of radiotracer due to their anatomic structure. Frequent causes of false positives in children include the columns of Bertin that differ in one individual and between the individuals, and cortical thickness ununified in one kidney (6, 9).

With respect to diagnoses, we classified patients into four groups (1) VUR and/or other morphological changes (2), VUR and infection (3), diagnoses other than VUR and (4) no changes/normal findings). On planar studies, the overall agreement ranged from 70% for VUR and/or other morphological changes to 88,9% for normal findings. On SPECT, the overall agreement ranged from 50% for VUR and/or other morphologic changes to 100% for VUR and infection and normal findings. This may be explained with strikingly abnormal or normal findings, respectively. Unfortunately, in our study, we did not dispose of the data on exact timing of the infection for all patients so we did not take them into account.

As suggested by Craig et al, high agreement for VUR and infection can be attributed to high prevalence of DMSA scan abnormality for the acute infection group (11).

Ladron de Guevara et al evaluated reproducibility for early scans for acute lesions and late scans for residual sequelae six months later, reporting on high reproducibility for both scintigraphies. Slight differences were noted pending the availability of early DMSA scans for the comparison purposes,

however, it was not possible to conclude if the availability of the early scans resulted in overdiagnosis of sequelae or in a higher sensitivity (12). As De Sadeleer et al suggested in their study on planar DMSA scintigraphy, abnormalities seen during an acute phase of infection are often more striking than the residual lesions (13).

When comparing for the kidney side (left vs. right), planar imaging rendered 73,4% agreement for the left kidney ($\kappa = 0,541$) and 81,1% for the right kidney ($\kappa = 0,633$). Lower agreement for the left kidney may be attributed to the physiologic appearance of the left kidney often impressed by the spleen in a way to mimic a photopenic defect. For the same reason, a κ value for the left kidney is in the moderate range, and the right kidney in the substantial range. SPECT rendered 72,2% agreement for the left kidney ($\kappa = 0,586$) and 73,7% for the right kidney ($\kappa = 0,539$). To support this, it is accepted that planar scintigraphy has a spatial resolution similar to that of SPECT, however, SPECT has a higher contrast resolution enabling delineation of the small renal defects, especially if the lesions are deeper within the kidney parenchyma (10). In our study, Observer 1 and Observer 2 demonstrated the highest percentage of agreement amongst the groups in all cases but one. Individual groups of Observer 3 share the lowest level of agreement with another groups in all cases but one. As κ cannot discriminate among different types and sources of disagreement, we hypothesize that despite all three raters come from the same nuclear medicine department, Observer 3 is the least experienced and in total has read fewer DMSA scans than the other two.

In our study, the overall agreement and κ for all individual pairs in relation to the mode of acquisition (planar vs. SPECT), diagnoses and kidney side are concluded to be within the acceptable ranges. From the clinical point of view, this is of paramount importance because the further management of patients should be the same regardless of which nuclear medicine physician interpreted the DMSA scan.

There are authors whose studies reported on significant variations in interpretations of DMSA scans, and low reproducibility of cortical scintigraphy (14,15). They suggested standardized criteria and terminology in interpretations. In our nuclear medicine department we apply standardized protocol and follow standardized criteria for interpretation (position, size, parenchymal contour, focal/diffuse uptake of radiotracer), and normal/abnormal impression of the kidney. However, the way of displaying (color scale, contrast enhancement, etc.) remains an observer's choice.

Since κ is affected by prevalence, we recommend further verification of the results in a larger cohort.

According to our results, a total agreement for planar and SPECT imaging is similar. However, higher variability between the readers is calculated for SPECT imaging, which may be contributed to a higher contrast resolution of SPECT, better delineation of details and thus the increased potential for different readings. That is why for the mode of acquisition, we would recommend hybrid imaging SPECT/CT method to be used whenever possible in detection of renal cortical defects on Tc-99m-DMSA scintigraphy.

REFERENCES

1. Piepsz A, Ham HR. Pediatric applications of renal nuclear medicine. *Semin Nucl Med.* 2006 Jan; 36(1): 16-35.
2. Piepsz A, Blafox MD, Gordon I, Granerus G, Majd M, O'Reilly P, Rosenberg AR, Rossleigh MA, Sixt R. Consensus on Renal Cortical Scintigraphy in Children With Urinary Tract Infection. Scientific Committee of Radionuclides in Nephrourology. *Semin Nucl Med.* 1999 Apr; 29(2): 160-74.
3. Piepsz A. Cortical scintigraphy and urinary tract infection in children. *Nephrol Dial Transplant.* 2002; 17: 560-2.
4. Rossleigh MA. Renal cortical scintigraphy and diuresis renography in infants and children. *J Nucl Med.* 2001; 42: 91-5.
5. De Sadeleer C, Bossuyt A, Goes E, Piepsz A. Renal Technetium-99m-DMSA SPECT in Normal Volunteers. *J Nucl Med.* 1996 Aug; 37(8): 1346-9.
6. Williams ED. Renal single-photon emission computed tomography: should we do it? *Semin Nucl Med.* 1992; 22: 112-21.
7. Mac Kenzie JR. DMSA - The new "gold standard". *Nucl Med Commun.* 1990; 11: 725-6.
8. Erdogan Z, Abdulrezzak U, Silov G, Ozdal A, Turhal O. Evaluation of interobserver variability of parenchymal phase of Tc-99m mercaptoacetyltriglycine and Tc-99m dimercaptosuccinic acid renal scintigraphy. *Indian J Nucl Med.* 2014 Apr; 29(2): 87-91.
9. Itoh K, Yamashita T, Tsukamoto E, Nonomura K, Furudate M, Koyanagi T. Qualitative and quantitative evaluation of renal parenchymal damage by 99mTc-DMSA planar and SPECT scintigraphy. *Ann Nucl Med.* 1995 Feb; 9(1): 23-8.
10. Tondeur MC, De Palma D, Roca I, Piepsz A, Ham HH. Interobserver reproducibility in reporting on renal cortical scintigraphy in children: a large collaborative study. *Nucl Med Commun.* 2009 Apr; 30(4): 258-62.
11. Craig JC, Irwig LM, Howman-Giles RB, Uren RF, Bernard EJ, Knight JF, Sureshkumar P, Roy LP. Variability in the interpretation of dimercaptosuccinic acid scintigraphy after urinary tract infection in children. *J Nucl Med.* 1998 Aug; 39(8): 1428-32.
12. Ladron de Guevara D, Franken F, de Sadeleer C, Ham H, Piepsz A. Interobserver Reproducibility in Reporting on 99mTc-DMSA Scintigraphy for Detection of Late Renal Sequelae. *J Nucl Med.* 2001 Apr; 42(4): 564-6.
13. De Sadeleer C, Tondeur M, Melis K, Van Espen MB, Verelst J, Ham H, Piepsz A. A Multicenter Trial on Interobserver Reproducibility in Reporting on 99mTc-DMSA Planar Scintigraphy: A Belgian Survey. *J Nucl Med.* 2000 Jan; 41(1): 23-6.
14. Gacinovic S, Buscombe J, Costa DC, Hilson A, Bomanji J, Ell PJ. Inter-observer agreement in the reporting of 99Tcm-DMSA renal studies. *Nucl Med Commun.* 1996 Jul; 17(7): 596-602.
15. Jaksic E, Beatovic S, Zagar I, Punkovic N, Stefanovic A, Ajdinovic B, Jansevic S, Han R. Interobserver variability in 99m Tc-DMSA renal scintigraphy reports: multicentric study. *Nucl Med Rev Cent East Eur.* 1999; 2(1): 28-33.
16. Caglar M, Kiratli PO, Karabulut E. Inter- and intraobserver variability of (99m)Tc-DMSA renal scintigraphy: impact of oblique views. *J Nucl Med Technol.* 2007 Jun; 35(2): 96-9.
17. Patel K, Charron M, Hoberman A, Brown ML, Rogers KD. Intra- and interobserver variability in interpretation of DMSA scans using a set of standardized criteria. *Pediatr Radiol.* 1993; 23(7): 506-9.

• Conflict of interest: none declared.