VIA MEDICA

Nuclear Medicine Review 2004
Vol. 7, No. 2, pp. 157–164

ISSN 1506–9680

Original

Cortical scintigraphy in the evaluation of renal defects in children with vesico--ureteral reflux — optimization of the procedure and study interpretation

Mariusz Gadzicki[1], Małgorzata Bieńkiewicz[2], Ewa Młodkowska[2], Małgorzata Knapska[3], Magdalena Kowalewska-Pietrzak[4], Jacek Kuśmierek[2]

[1]Department of Radiology — Diagnostic Imaging, Medical University in Łódź
[2]Department of Nuclear Medicine, Medical University in Łódź
[3]Department of Endocrinology and Radionuclide Therapy, Medical University in Łódź
[4]Department of Pediatric Diseases, Medical University in Łódź

[Received 12 XI 2004; Accepted 16 XI 2004]

Abstract

BACKGROUND: The objective of this study was to analyse the performance of several variants of kidney scintigraphy in children from the standpoint of: scar detection, an assessment of the rating of the pathology and an investigation of interobserver variability involved in the diagnostic procedure.
MATERIAL AND METHODS: The analysis is based on results of a planar kidney scintigraphy and of a tomographic (SPECT) procedure. The latter was performed in two variants: 1) in which slices were obtained with axis of reconstruction identical with longitudinal axis of the body (SPECT I) and 2) in which axes were fitted to the long axis of each kidney separately (SPECT II). The rating of the diagnosed pathology was made using two scales, according to Goldraich and Howard.
Evaluation of the images involved on the one hand, 150 individual kidneys and 75 patients on the other. The assessment was made by three independent observers, differing in experience in nuclear medicine and employed in three independent departments.
In the statistical analysis, as a measure of observer agreement, a proportion of agreeing readings (%) was accepted; in addition, the kappa index of agreement was calculated.
RESULTS: Better agreement among three observers was attained when planar images were read in contrast to SPECT (I and II) results. The reading of SPECT II images yielded a higher frequency of diagnosed pathology (scars) in kidneys and is characterized by better overall agreement in detection by individual observers than a similar evaluation of SPECT I images. The Goldraich scale secures better interobserver agreement of renal scar detection than is seen when the Howard scale was applied to acquire the rating.
CONCLUSIONS: The conclusion may be drawn that kidney scintigraphy is a method still burdened with a substantial subjectivism. Planar scintigraphy should be treated as a basic option for imaging post-inflammatory changes in kidneys.
Key words: DMSA analogue, scintigraphy, kidney, renal scar, children, interobserver variability

Correspondence to: Mariusz Gadzicki
Department of Nuclear Medicine, Medical University in Łódź
ul. Czechosłowacka 8/10, 92–216 Łódź
Tel: (+48 42) 678 36 85

Introduction

The formation of scars in renal parenchyma in children is related to recurrent infections in the urinary system, induced primarily by a vesico-ureteral reflux [1, 2]. In turn, this process leads to the development of arterial hypertension and impairment of renal function and eventually to terminal insufficiency of the kidneys [3].

Medical imaging plays an important role in the diagnosis and monitoring of the scar forming process in kidneys [4]. Static renal scintigraphy with use of dimercaptosuccinic acid (DMSA) labelled with ^{99m}Tc, is at present treated as a reference method for following the course of the developing disease [5]. There are, however, numerous ways of performing the scintigraphy [6]. The differen-

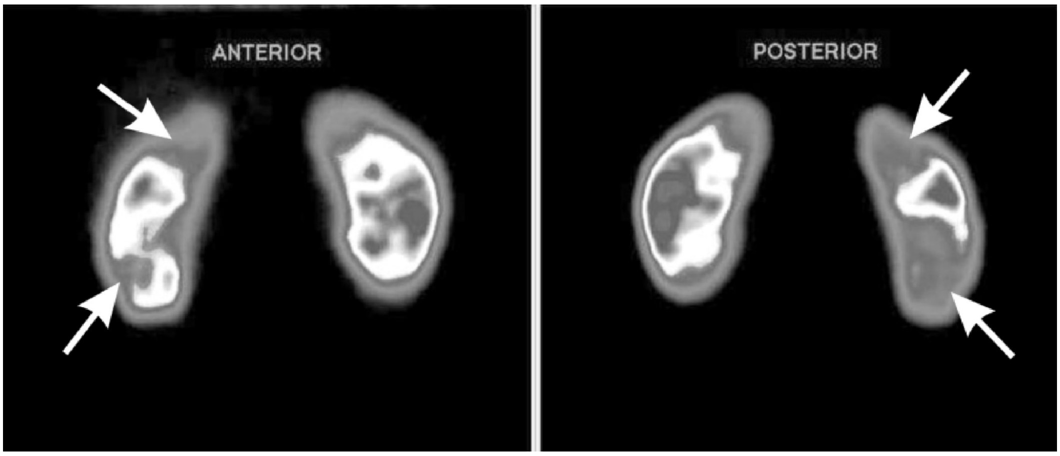

Figure 1. Planar images of kidneys in anterior and posterior views (renal scars are indicated with arrows).

ces encountered apply to: the mode of acquisition and processing of the image — planar or tomographic (SPECT, single photon emission computerized tomography), the selection of the projection — posterior, anterior, posterior oblique, lateral oblique; the collimators used – parallel high resolution or pinhole. The acquired information is also recorded in different matrices, and with varying degrees of magnification. Moreover, the evaluation of the images is performed based on various criteria for the diagnosis of renal scars [7–9], and finally rating the extent of renal scarring is based on various classification systems [10–12].

There are numerous publications pointing to the fact of substantial differences in the interpretation of image assessment between individual observers [13–15]. Investigations aiming specifically at the evaluation of agreement in the interpretation of renal scintigrammes usually confirm the existence of interobserver variability. However, while some authors point to sufficiently good agreement between individual observers [16, 17] others maintain that the agreement is poor [18, 19]. One has to remember, that a statistical analysis of interobserver agreement offers a general and perhaps unique possible estimate of the diagnostic reliability of a method for which there is no independent, valuation of the diagnosis, against which one could calculate such commonly accepted estimators of diagnostic efficacy, as sensitivity, specificity and accuracy.

The objective of the study presented below was to analyse the interobserver agreement of renal scintigrammes evaluation from the standpoint of scar detection in the kidneys, and of scarring ratings. Aiming at the optimization of renal scintigraphy and the interpretation of its results the analysis was performed separately for planar and tomographic (in two versions) investigations. Attempts to optimize the rating of the scarring on the basis of obtained images concentrated on a parallel evaluation of two scales for estimating the intensity of the pathologic process.

Materials and methods

Seventy four children (18 boys, 56 girls) were studied in whom a primary vesico-ureteral reflux had been diagnosed. Seventy eight renal scinigraphic studies were performed with ^{99m}Tc-unithiol as a radiopharmaceutical (this is a strict analogue to DMSA) [20]. The presence or absence of renal scars was based on the evaluation of a scintigraphy performed at least 6 months since the diagnosis of the primary reflux, or since the last diagnosed incident of urinary infection accompanied by elevated body temperature.

^{99m}Tc-unithiol was administered intravenously at an activity calculated individually for the patient's age (acc. to the du Bois formula — age +1/age +7) taking as a reference value activity of 185 MBq for a standard adult human. Acquisition was started 2 hours after the administration of the radiopharmaceutical (RF). A double-head scintillation camera VARICAM (Elscint) with high resolution parallel low energy collimators was used.

Planar images (Figure 1) were recorded in the 128×128 matrix, at internal magnification factor = 2. SPECT acquisition was made by recording 2×60 images over 30 s each; each detector rotated for 180° at an internal magnification = 2 or more, and the data were accumulated in a 128×128 matrix.

Planar images served both for the visualization of scars and for the assessment of the relative uptake of the RF by each of the kidneys. The latter was calculated as a geometrical mean of counts in the region of interest of each kidney in an anterior and posterior projection, taking the sum of counts for both kidneys as 100%.

The tomographic images (SPECT) consisted of 3 series of slices (0.4–0.6 cm thick) in the following planes: transversal, frontal and sagittal. This image processing was carried out in two versions:

1. (SPECT I) — when axis of "cutting" was coincident with the long axis of the body (Figure 2).
2. (SPECT II) — when the cutting axes were those that corresponded to the long axis of each kidney, separately (Figure 3).

In each kidney 3 regions were delineated: upper, middle and lower.

In planar images a scar was diagnosed when there was lack of RF uptake or a reduced uptake, it was visible in more than one projection and/or a deformation was seen in the contour of the kidney.

In the tomographic study a scar was recognized when there was a lack of RF uptake in the renal cortex, a reduced uptake in a cortex of normal thickness, or local thinning of the cortical layer.

The rating of scarring was evaluated according to two scales. Firstly, the scale formulated by Goldraich was employed, which is as follows:

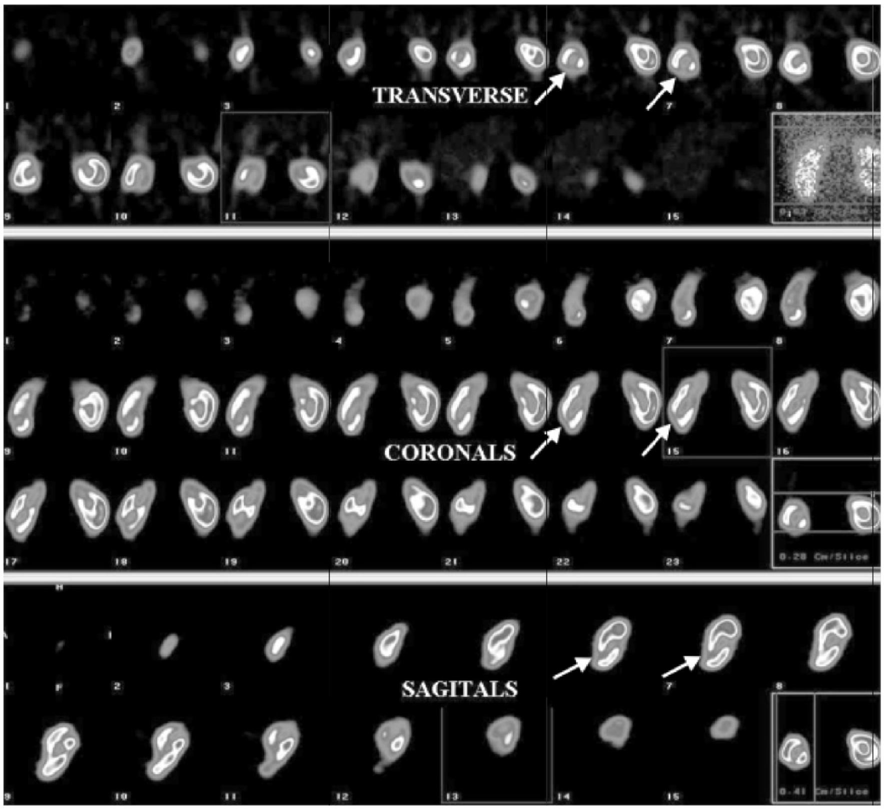

Figure 2. Tomographic images of kidneys (SPECT I) — slices in the transversal, frontal and sagittal planes — axis of "cutting" was coincident with long axis of the body (renal scars are indicated with arrows).

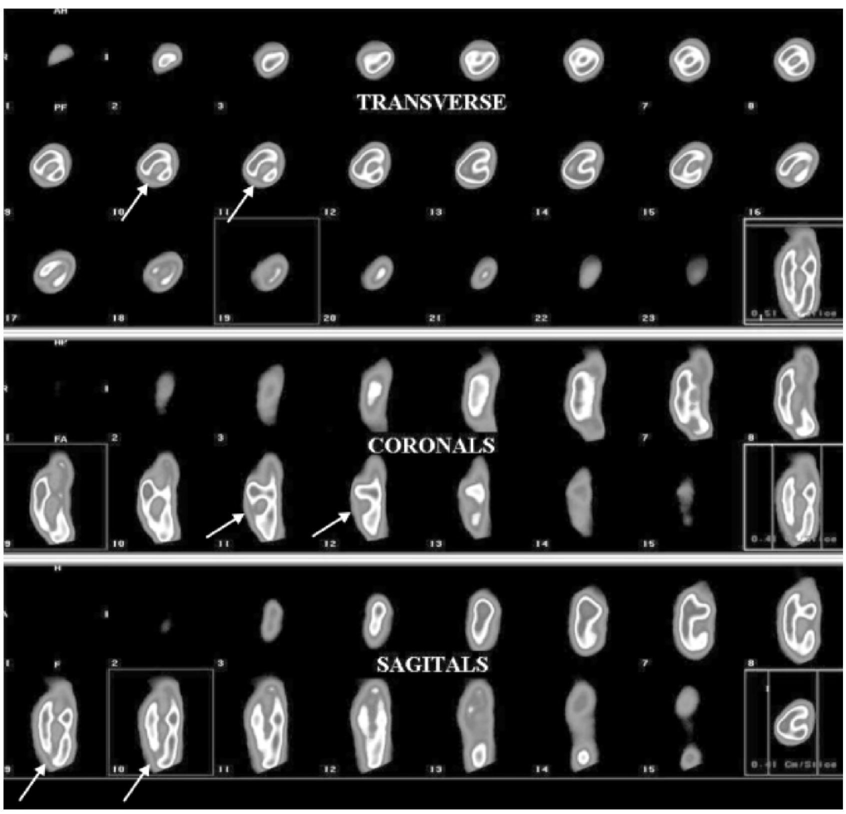

Figure 3. Tomographic images of right kidney (SPECT II) — slices in the transversal, frontal and sagittal planes — axis of "cutting" corresponds to long axis of right kidney (renal scars are indicated with arrows).

— 0 — lack of any focal disturbance of the RF uptake;
— 1 — no more than two cortical defects;
— 2 — more than two cortical defects, the rest of parenchyma not affected;
— 3 — a diffuse reduction of the uptake in the whole kidney with or without focal defects;
— 4 — a small kidney of irregular shape, with the uptake of RF less than 10 per cent of the total for both kidneys [10].

As a second scale that which was proposed by Howard [12] was used. The rating in this scale was based on the following criteria:
— 0 — lack of any focal abnormality;
— 1 — focal changes in 1 region;
— 2 — changes in 2 regions;
— 3 — focal changes in 3 regions;
— 4 — a small kidney of irregular contours with the uptake below 10% of the total renal uptake of RF.

An evaluation of 150 renal images on the one hand, and of 75 patients on the other, was made independently by three observers, coming from 3 nuclear medicine departments. Two of them were experienced nuclear medicine physicians. The third one was rather inexperienced in the interpretation of renal scintigrammes.

A statistical analysis of the image evaluation agreement between 3 pairs of observers was based on the proportion of coincident verdicts (%) and upon the magnitude of the kappa agreement index [21]. The kappa index is a recommended test which provides a measure for the agreement above the expected value of purely incidental character. When value of the kappa index equals 1 the agreement is perfect; when the index assumes a value = 0, the agreement is purely incidental. As a measure of the overall agreement, an arithmetical mean of agreement proportion for 3 pairs of observers was adopted. The reproducibility of the scintigramme evaluation was considered at two levels: as related to patients and separately to individual kidneys. Significance of differences between values expressed in percent was assessed by means of a test for difference between proportions.

Results

Frequency of diagnosed scarring

When two scales were applied for rating the scar (according to those of Howard and Goldraich) in individually evaluated scintigrammes of 150 kidneys the frequency of pathologic images in the planar and SPECT I mode was similar. Using the Howard scale, the scarring of a various extent was found in approx 50% of the kidneys. When the Goldraich scale was used, the proportion increased to approx. 60% ($p < 0.05$).

When tomographic images were reconstructed according to the SPECT II procedure the mean frequency of diagnosed scarring increased, in contrast to that found in a SPECT I reconstruction, to 70% (in individually evaluated kidneys). This applied to both rating scales (71% acc. to the Goldraich's scale — $p < 0.05$; and 68% to the Howard classification — $p < 0.001$. The distribution of the rating by individual observers in 150 kidneys is presented in Table 1.

When images of seventy five patients were considered, the frequency of post inflammatory scarring was similar in the planar and SPECT I mode. The rating scale of Howard yielded 69% and 66% of patients, as seen in images obtained in the two modes, and the Goldraich classification gave slightly higher (but insignificant, $p = 0.1$) results of 77 and 75 per cent for the planar and SPECT I modes, respectively.

The SPECT II mode reconstruction of tomographic images yielded a proportion of patients with renal scarring still slightly higher than that seen in planar mode on the boundary of significance: 85% and 82%, for the Goldraich ($p = 0.1$) and Howard rating ($p < < 0.05$) respectively. The respective data as diagnosed by individual observers are presented in Table 2.

An analysis of ratings by individual observers, as shown in table 1 and 2, led to the conclusion that observer C (with the least experience in nuclear medicine) had an obvious tendency to see scars that were not perceived by his or her remaining two colleagues. The tendency was particularly pronounced when SPECT II images were evaluated.

Table 1. Distribution of the renal scintigraphy ratings (%) for 3 observers (A, B and C) in evaluation of images of 150 kidneys — for different modes of acquisition and classification system

Distribution of renal scintigraphy ratings (%) in kidneys using Howard's classification system															
Observer	Planar					SPECT I					SPECT II				
	0	I	II	III	IV	0	I	II	III	IV	0	I	II	III	IV
A	61	17	13	6	3	73	13	8	4	2	34	27	11	26	2
B	53	30	13	3	1	50	22	18	9	1	45	20	19	15	1
C	41	36	16	6	1	26	40	28	5	1	16	52	27	4	1
mean	51	28	14	5	2	50	25	18	6	1	32	33	19	15	1
mean 0/1	51		49			50		50			32		68		
Distribution of renal scintigraphy ratings (%) in kidneys using Goldraich's classification system															
A	39	27	6	25	3	47	16	2	33	2	27	41	11	18	3
B	51	42	3	3	1	48	29	6	16	1	43	34	5	17	1
C	38	48	7	6	1	25	44	9	21	1	16	54	7	22	1
mean	43	39	5	11	2	40	30	6	23	1	29	43	7	19	2
mean 0/1	43		57			40		60			29		71		

Table 2. Distribution of renal scintigraphy ratings (%) for 3 observers (A, B and C) in evaluation of images for 75 patients (maximum grade of abnormality in either kidney defining overall patient grade) — for different modes of acquisition and classification system

Observer	Planar					SPECT I					SPECT II				
	0	I	II	III	IV	0	I	II	III	IV	0	I	II	III	IV
Distribution of renal scintigraphy ratings (%) in patients using Howard's classification system															
A	43	22	19	9	7	62	18	11	5	4	18	31	15	33	3
B	31	38	24	4	3	33	23	25	16	3	31	24	22	20	3
C	20	42	26	9	3	8	42	39	8	3	4	49	39	5	3
mean	31	34	23	8	4	34	28	25	10	3	18	35	25	19	3
mean 0/1	31	69				34	66				18	82			
Distribution of renal scintigraphy ratings (%) in patients using Goldraich's classification system															
A	20	34	9	30	7	34	23	3	36	4	12	47	16	21	4
B	29	57	4	7	3	33	36	11	17	3	29	41	8	19	3
C	19	59	10	9	3	7	48	12	30	3	4	56	8	29	3
mean	23	50	8	15	4	25	36	8	28	3	15	48	11	23	3
Mean 0/1	23	77				25	75				15	85			

Table 3. Measurement of overall agreement for 3 observers (calculated as mean values for three pairs of the observers) in evaluation of renal scintigraphy in children with post-infectious renal scars — percentage of agreement (%) and kappa statistic [κ]

Category	No.	Planar				SPECT I				SPECT II			
		Howard		Goldraich		Howard		Goldraich		Howard		Goldraich	
		%	κ	%	κ	%	κ	%	κ	%	κ	%	κ
Kidneys 0–1	150	78	0.58	80	0.60	65	0.35	77	0.53	72	0.37	72	0.31
Kidneys 0–4	150	68	0.50	66	0.49	55	0.34	63	0,48	42	0.24	52	0.32
Patients 0–1	75	78	0.51	85	0.57	59	0.20	74	0.31	81	0.36	81	0.30
Patients 0–4	75	60	0.46	61	0.44	43	0.26	48	0.42	37	0.18	54	0.34

0–1 — normal/abnormal, 0–4 — grade of abnormality

Evaluation of the interobserver agreement while rating the renal scarring

The statistical analysis of the interobserver agreement has been based on the proportion of concordant assessments and on the kappa agreement index. The measurements of the overall agreement, calculated as mean values for the three pairs of observers, are presented in Table 3.

In this analysis for individual kidneys the highest agreement of binary diagnoses at 80 per cent, was obtained when the images acquired in the planar mode were evaluated with the applications of the Goldraich rating scale. The tomographic reconstruction was characterized by a somewhat worse concordance of 77% ($p = 0.2$) and 72% ($p = 0.05$) for SPECT I and SPECT II, respectively. The use of the Howard scale yielded similar values of 78 (planar) and 72 (SPECT II) per cent. Only the SPECT I mode yielded a significantly lower value of concordant assessments of 65% ($p < 0.05$).

A similar analysis of interobserver agreement was made when diagnosing 75 patients from the viewpoint of establishing the presence or absence of renal scarring. The highest overall agreement of 85% was attained when planar images were evaluated using a scale of Goldraich. A somewhat lower indicator of agreement (78%, $p = 0.1$) was obtained when the Howard scale was used. Images obtained in SPECT I mode showed less concordance than planar images (59% *vs.* 78% — Howard's classification ($p < 0.01$) and 74% *vs.* 85% according to Goldraich ($p = 0.05$).

The concordance of assessment, which applied to patients, was evidently better in SPECT II mode than in SPECT I, and this applied to both classifications of the rating. The mean proportion of concordance results, in both scales, amounting to 81% for SPECT II was greater than that for SPECT I (74% and 59% for the Goldraich ($p = 0.1$) and Howard ($p < 0.01$) scales, respectively. However, there was non statistically significant difference ($p = 0.3$) between proportions of patients with scarring as seen in the SPECT II and planar modes (85% and 78% for the Goldraich and Howard classification, respectively).

When the degree of scarring (from 0 to IV grade) was analysed, both for patients and individual kidneys, the agreement, defined as the proportion of concordant results, was poorer than that seen when organs and patients were classified as healthy or affected by scarring (Table 3).

Additional information on the character of interobserver agreement was provided by an analysis of the kappa index. The latter takes into account the rate of chance-corrected agreement and possible disproportion between a positive and negative diagnosis.

The disadvantage of this index lies in its dependence on the frequency of positive rating (presence of scarring), which makes invalid direct comparisons with similar data reported in the literature [21].

It seems important that the values of the kappa index for planar images and for all pairs of evaluations by observers fell in the range 0.44–0.60; this range of agreement is commonly declared as moderate. For tomographic images the values kappa were lower: and may be valued as follows: SPECT I — 0.20–0.53 — "fair" to "moderate" SPECT II 0.18–0.37 "slight" to "fair".

A direct explanation for these differences between the evaluation of planar and SPECT images may be sought in the fact that the frequency of positive findings in SPECT modes (and mostly in SPECT II), was higher. These leads to larger difference between the frequency of positive and negative findings, which in turn reduces the values of the kappa index. Critical remarks in the literature on the validity of the kappa index when applied for quantitative judgements regarding the interobserver agreement [22], has led some authors to the conclusion that the index should be used only for assessment if the observed agreement is better than that due to pure chance only. From this point of view, all the data presented above satisfy this requirement; for all evaluations the kappa index was significantly different from zero ($p < 0.01$).

Discussion

The detection of pathological foci in the cortex of kidneys by means of static scintigraphy is one of the commonly applied procedures in paediatric nuclear medicine. The Scientific Committee for Utilization of Radioisotope in Uronephrology at its meeting in Copenhagen in 1998 reached a consensus regarding renal scintigraphy in children with urinary infection. Recommendations of this committe were concentrated on the application of planar study as a basic diagnostic option [6]. In the meantime, the number of multihead scintigraphic cameras has been rapidly increasing and this has led to more and more common utilization of the tomographic option for detection and rating of renal scarring. There are, however, conflicting opinions regarding the sensitivity of the detection of a pathology in the renal cortex by means of SPECT procedures. Everaert et al. [7] did not find a significant increase in the number of detected foci when SPECT was applied instead of planar scanning. Similarly, Mouratidis et al. [3] were unable to demonstrate a statistically significant increase in the number of detected foci when results of a SPECT procedure, carried out by means of a single-head camera were compared with a high resolution planar scintigraphy of the kidneys. On the other hand, Cook et al. [8] detected a greater number of the foci when SPECT was used, as compared with planar acquisition. In the study reported by these authors, the number of equivocal images decreased when tomographic acquisition and reconstruction were evaluated as compared with the results of the planar technique.

In our studies an observation was made that the number of detected patients and individual kidneys with scars was similar when planar and SPECT I procedures were compared. However, when the reconstruction of tomographically acquired data was modified taking longitudinal axes of individual kidneys as the mode of reconstruction (SPECT II), the frequency of detected scars increased. Our interpretation as to why the increase took place is, that the modification of the reconstruction axes improved imaging of both poles of the kidney as well as enabling the unification of the thickness of analysed slices. Therefore, images from SPECT II procedures also resulted in better interobserver agreement in terms of binary decision healthy/affected, and this appeared true both with regard to patients and individual kidneys . The proportion of concordant evaluation in our study is comparable to that given by Craig et al. [23], who reported overall agreement in 88% and 78% for the planar and tomographic study. However, we were unable to improve by SPECT the agreement in the rating of scarring in semiquantitative classifications.

Interobserver differentiation of assessments may also result from high anatomic variability of kidneys, from persisting fetal lobulation, from evident structures of linking parenchyma, from the doubling of the pyelo-calyceal system etc. [24]. Also, the inability to compare renal scintigrams with images obtained by means of other modalities, e.g. ultrasound may enhance disparities in interpretation of the observed structures and their pathologic modifications.

Another factor in our study that reduced degree of agreement between the observers was the difference in their levels of experience. It was quite obvious that the less experienced observer tended to see more renal scars than those with more experience.

The two classifications applied for rating of the scarring lead one to the conclusion that the distribution of the degree of pathology is different for each of the codes. The Goldraich rating displayed a bimodal distribution of the score, which also confirms observations made by Craig [23]. The scale, as proposed by Goldraich bases the rating on the sum of two separate features: on the number of renal defects found and the character of the radiopharmaceutical uptake. Their frequency of distribution does not seem to mimic each other. In consequence, in each evaluation according to this scale, a bimodal distribution of the score is seen. Howard's classification is based on a single criterion: the magnitude of the affected part of the kidney (the extent of the scarring), and therefore avoids the bimodality of the score. There is little, if any evidence of a similar character from those papers whose authors utilized the Howard's system for the rating. In our material, the interobserver agreement is, however, better for the rating according to Goldraich.

The recommendations issued in 2002 by Paediatric Committee of the European Association of Nuclear Medicine [25] for the use of renal scintigraphy performed in the acute stage of pyelonephritis and used for the prognostication of late effects of the disease, has led to a quantitative method being employed for the analysis of the results of planar and tomographic scintigraphy. Hitzel et al. [26] proposed a method that according to the authors, should permit reliable prognosis of future scarring development when based on quantitative parameters for estimating the cortical uptake ($C_{70\%}$) from results of planar renal scintigraphy. Chiou et al. [27] demonstrated that measuring the volume of inflammatory foci in the acute phase of pyelonephitis, detected by tomographic (SPECT) scintigraphic examination, has a substantial predictive value in the evaluation of the scarring process and allows one to identify patients with a high risk of extensive scarring. All this seems to point to the fact that the modification of the methods used so far for the purpose in question, with the aim of improving overall interobserver agreement, makes sense. Therefore, some positive examples should be quoted: a method proposed by Estorch et al. [28] for the assessment of function of individual kidneys, based on the relative uptake of ^{99m}Tc-DMSA, corrected for the dimensions of the organs, and a report by Caione et al. [29], which makes possible determination of the absolute uptake of the same radiopharmaceutical by the kidneys and therefore, a more accurate assessment of the relative uptake of 99mTc-DMSA by each of them.

It is higher than an improvement of the interobserver agreement of the diagnosis of renal scarring by tomographic scintigraphy as well as an increase in real sensitivity of cortical foci detection may be attained e.g. by the change of acquisition from a full 360° rotation of the camera to 180° from the dorsal side only [30, 31]. Supposedly, this leads to the avoidance of spurious hypoactivity of the upper poles of the kidneys, which results from photon absorption by the liver and spleen. Another attempt at the improvement of diagnostic capacity may lie in a tomographic evaluation of the kidneys (SPECT) accompanied, or verified in parallel, by ultrasound examination.

All in all, what seems to follow from this present study and a survey of the literature is the conclusion that static scintigraphic examination of kidney is still burdened with substantial subjectivity. This follows also from the observation that a lack of experience tends to "facilitate" the detection of scars, particularly in the tomographic version of the method.

The results of present study indicate also that planar scintigraphy is an option that allows one at present, to obtain maximum attainable interobserver agreement, and therefore should be recommended as the basic standard. At the same time, a review of the available literature on possible further modifications of the procedure supports the contention that an improvement in the reproducibility of results seems feasible and should, perhaps, lead to more reliable attempts at prognosis of the course of the disease, contributing therefore to the identification of those patients who run the highest risk of an unfavourable outcome and who require particularly intense follow up and therapy.

Conclusions

1. The results of planar renal scintigraphy displayed better interobserver agreement in the assessment of renal scarring than the tomographic (SPECT) counterpart of the procedure.
2. The evaluation of tomographic scintigraphic images of kidneys using a reconstruction of the data based on the longitudinal axes of each kidney increased the frequency of diagnosed scars than a traditional reconstruction of the images along the single axis of the body. The interobserver agreement in the former procedure was also greater than that resulting from the latter.
3. Rating the extent or intensity of scarring using Goldraich's scale secured better interobserver agreement than the use of Howard's scale.

References

1. Smellie JM, Prescod NP, Shaw PJ et al. Childhood reflux and urinary infection: a follow-up of 10–41 years in 226 adults. Pediatr Nephrol 1998; 12: 727–736.
2. Martinell J, Claesson I, Lidin-Janson G et al. Urinary infection, reflux and renal scarring in females continuously followed for 13–38 years. Pediatr Nephrol 1995; 9: 131–136.
3. Jacobson SH, Eklof O, Eriksson CG et al. Development of hypertension and uraemia after pyelonephritis in childhood: 27-year follow-up. BMJ 1989; 299: 703–706.
4. Hoberman A, Charron M, Hickey RW, et al. Imaging studies after a first febrile urinary tract infection in young children. N Engl J Med 2003; 348: 195–202.
5. Rushton HG, Majd M. Dimercaptosuccinic acid renal scintigraphy for the evaluation of pyelonephritis and scarring: a review of experimental and clinical studies. J Urol 1992; 148: 1726–1732.
6. Piepsz A, Blaufox MD, Gordon I et al. Consensus on renal cortical scintigraphy in children with urinary tract infection. Semin Nucl Med 1999; 29: 160–174.
7. Everaert H, Flamen P, Franken PR et al. ^{99}Tc^{m}-DMSA renal scintigraphy for acute pyelonephritis in adults: Planar and/or SPET imaging? Nucl Med Commun 1996; 17: 884–889.
8. Cook GJR, Lewis MK, Clarke SEM. An evaluation of ^{99}Tc^{m}-DMSA SPET with three-dimensional reconstruction in 68 patients with varied renal pathology. Nucl Med Commun 1995; 16: 958–967.
9. Mouratidis B, Ash JM, Gilday DL. Comparison of planar and SPECT ^{99}Tc^{m}-DMSA scintigraphy for the detection of renal cortical defects in children. Nucl Med Commun 1993; 14: 82–86.
10. Goldraich NP, Ramos OL, Goldraich IH. Urography versus DMSA scan in children with vesicoureteric reflux. Pediatric Nephrol 1989; 3: 1–5.
11. Benador D, Benador N, Slosman DO, et al. Cortical scintigraphy in the evaluation of renal parenchymal changes in children with pyelonephritis. J Pediatr 1994; 124: 17–20.
12. Howard RG, RoebuckDJ, Yeung PA et al. Vesicoureteric reflux and renal scarring in Chinese children. Br J Radiol 2001;74: 331–334.
13. Elmore JG, Wells CK, Lee CH, et al. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994; 331: 1493–1499.
14. Carlos RC, Stanley JC, Stafford-Johnson DS et al. Interobserver variability in the evaluation of chronic mesenteric ischemia with gadolinium-enhanced MR angiography. Acad Radiol 2001; 8: 879–887.
15. Lin E, Connolly LP, Zurakowski D et al. Reproducibility of renal length measurements with ^{99m}Tc-DMSA SPECT. J Nucl Med 2000; 41: 1632–1635.
16. Patel K, Charron M, Hoberman A et al. Intra- and interobserver variability in interpretation of DMSA scans using a set of standardized criteria. Pediatr Radiol 1993; 23: 506–509.
17. de Guevara DL, Franken P, de Sadeleer C et al. Interobserver reproducibility in reporting on ^{99m}Tc-DMSA scintigraphy for detection of late renal sequelae. J Nucl Med 2001; 42: 564–566.
18. Jaksic E, Beatovic S, Zagar I et al. Interobserver variability in ^{99m}Tc-DMSA renal scintigraphy reports: multicentric study. Nucl Med Rev 1999; 2: 28–32.
19. Gacinovic S, Buscombe J, Costa DC et al. Inter-observer agreement in the reporting of ^{99}Tc^{m}-DMSA renal studies. Nucl Med Commun 1996; 17: 596–602.
20. Oginski M, Rembelska M. ^{99m}Technetium-unithiol complex, a new pharmaceutical for kidney scintigraphy. Nuklearmedizin. 1976; 15: 282–286.
21. Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003; 228: 303–308.
22. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993; 46 (5); 423–429.
23. Craig JC, Irwig L, Ford M, et al. Reliability of DMSA for the diagnosis of renal parenchymal abnormality in children. Eur J Nucl Med 2000; 27: 1610–1616.

24. Yeh HC. Some misconceptions and pitfalls in ultrasonography. Ultrasound Q 2001; 17: 129–155.
25. Biggi A. Impact of acute cortical scintigraphy in children. Nucl Med Commun 2003; 24: 13–14.
26. Hitzel A, Liard A, Dacher JN et al. Quantitative analysis of ^{99m}TC-DMSA during acute pyelonephritis for prediction of long-term renal scarring. J Nucl Med 2004; 45: 285–289.
27. Chiou YY, Wang ST, Tang MJ et al. Renal fibrosis: Prediction from acute pyelonephritis focus volume measured at ^{99m}Tc Dimercaptosuccinic Acid SPECT. Radiology 2001; 221: 366–370.
28. Estorch M, Torres G, Camacho V et al. Individual renal function based on ^{99m}Tc dimercaptosuccinic acid uptake corrected for renal size. Nucl Med Commun 2004; 25: 167–170.
29. Caione P, Ciofetta G, Collura G, et al. Renal damage in vesico-ureteric reflux. BJU Int 2004; 93: 591–595.
30. Peng NJ, Kwok G, Chiou YH, et al. Posterior 180^{0} ^{99m}Tc-dimercaptosuccinic acid renal SPECT. J Nucl Med 1999; 40: 60–63.
31. Peng NJ, Liu RS, Chiou YH, et al. ^{99}Tc^{m}-dimercaptosuccinic acid renal scintigraphy for detection of renal cortical defects in acute pyelonephritis: posterior 180° SPECT versus planar image and 360° SPECT. Nucl Med Commun 2001; 22: 417–422.