



Published in final edited form as:

J Pediatr Urol. 2017 December ; 13(6): 616.e1–616.e6. doi:10.1016/j.jpuro.2017.03.040.

Interobserver Variability for Interpretation of DMSA Scans in the RIVUR Trial

Tej K. Mattoo^{a,*}, Steven J. Skoog^b, Lisa Gravens-Mueller^c, Russell W. Chesney^d, Alejandro Hoberman^e, Ranjiv Mathews^f, Marva Moxey-Mims^g, Anastasia Ivanova^c, Saul P. Greenfield^h, and Myra A. Carpenter^c For the RIVUR Trial Investigators

^aNephrology and Hypertension, Children's Hospital of Michigan, Wayne State University School of Medicine, Detroit, MI, USA

^bDivision of Pediatric Urology, Oregon Health & Science University, Portland, OR, USA

^cDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

^dDepartment of Pediatrics, University of Tennessee Health Science Center, Memphis, TN, USA (deceased)

^eDepartment of Pediatrics, University of Pittsburgh School of Medicine, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA, USA

^fDepartment of Urology, Johns Hopkins School of Medicine, Baltimore, MD, USA

^gNational Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health Bethesda, MD, USA

^hDivision of Pediatric Urology, Women & Children's Hospital of Buffalo, State University of New York at Buffalo School of Medicine & Biomedical Sciences, Buffalo, NY, USA

Summary

Introduction—Technetium-99m dimercaptosuccinic acid (DMSA) renal scans are used in the diagnosis of renal scarring. In the Randomized Intervention for Children with Vesicoureteral Reflux (RIVUR) trial that randomized 607 children, DMSA renal scans were used for evaluating the presence and the severity of renal scarring.

Objective—The aim was to determine interobserver variability in reporting of DMSA renal scans in the RIVUR trial.

Study design—We compared DMSA renal scan reports for renal scarring and acute pyelonephritis from all non-reference local radiologists (ALRs) at study sites with adjudicated as

*Corresponding author: Children's Hospital of Michigan, Wayne State University School of Medicine, 3010 Beaubien Blvd, Detroit, MI, 48201. Tel.: +1 313 745 5604; Fax: +1 313 966 0039. tmattoo@med.wayne.edu (T.K. Mattoo).

Conflict of interest

None.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

well as non-adjudicated reports from two reference radiologists (RRs) of the RIVUR trial. Two-way comparisons of concordant and discrepant responses were analyzed using an unweighted kappa statistic between the ALR and the adjudicated RR interpretations. All analyses were performed using SAS v 9.4 (SAS institute 2015) and significance was determined at the 0.05 level.

Results—Of the 2,872 kidneys evaluated, adjudicated RR reports had 119 (4%) kidneys with renal scarring compared with 212 (7%) by the ALRs. For 79% kidneys the grading for scarring reported by ALRs was either upgraded (24%) or downgraded (55%) by RRs. For acute pyelonephritis ($n = 2,924$), adjudicated RR reports had 85 (3%) kidneys with pyelonephritis compared with 151 (5%) by the ALRs. For 85% kidneys, the grading for pyelonephritis reported by the ALRs was either upgraded (28%) or downgraded (57%) by the RRs. A three-way comparison revealed that all three (RR1, RR2, and ALR) agreed over presence of renal scarring in 19% cases and two of the three agreed in 80% cases (Table). The respective numbers for pyelonephritis were 13% and 84%. The agreement rate for all DMSA scan reports between the RRs and the ALRs was 93%.

Discussion—The study revealed significant interobserver variability in the reporting of abnormal DMSA renal scans compared with the previously published studies. A noteworthy limitation was a lack of uniformity in local reporting of the scans.

Conclusions—Our study highlights the need for optimizing the clinical yield of DMSA renal scans by more specific guidelines, particularly for standardized and uniform interpretation.

Introduction

Renal scarring because of pyelonephritis can occur with or without vesicoureteral reflux (VUR), but the risk is higher in the presence of VUR, particularly high-grade VUR [1,2]. Scarring can result from a single episode of acute pyelonephritis or may take several years to develop [3,4]. Renal scarring can also be due to abnormal renal development resulting in focal renal hypoplasia/dysplasia [5]. Depending on its severity, the diagnosis of renal scarring can be important for clinical management and patient counseling for potential long-term complications that include hypertension, proteinuria, and progression to end-stage renal failure [6].

Renal scintigraphy using the technetium-99m dimercaptosuccinic acid (DMSA) renal scan is the current standard for identifying renal scars in clinical practice [6,7]. However, inconsistencies in methodology and inter- and intraobserver variability in the interpretation of DMSA scans have been an ongoing concern. Published studies have reported a variable rate of reproducibility in scan interpretation, ranging from poor [7,8] to high [7,9–12]. Various guidelines have attempted to improve the quality of DMSA scans and interpretation of the results. These include the Society of Nuclear Medicine guidelines published by Mandell et al. [13], the international consensus by the Scientific Committee of Radionuclides in Nephrourology by Piepsz et al. [14] and the European guidelines by the Pediatric Committee of the European Association of Nuclear Medicine by Piepsz et al. [15].

The Randomized Intervention for Children with Vesicoureteral Reflux (RIVUR) trial evaluated the role of antimicrobial prophylaxis in the prevention of recurrent urinary tract infection (UTI) and renal scarring in 607 children with VUR [16,17]. During the study

period of 2 years, each study participant underwent three DMSA renal scans, offering a unique opportunity to evaluate interobserver agreement in our study cohort. The main objective of the present report was to determine interobserver agreement in the reporting of DMSA scans, particularly abnormal scans, by two reference radiologists (RRs) of the RIVUR trial and all non-reference local radiologists (ALRs) who initially read the scans at study sites.

Patients and methods

The RIVUR trial was a multicenter, randomized, placebo-controlled trial in 607 children aged 2–71 months with grade I–IV primary VUR diagnosed after a first or second febrile or symptomatic UTI. Study participants received trimethoprim/sulfamethoxazole (TMP/SMZ) or a closely matched placebo and were followed for 2 years. Details on methods for patient selection, data collection, and statistical considerations for the trial have been described previously [16,18].

Study participants in the RIVUR trial were scheduled for three DMSA renal scans. The baseline scan was obtained within 2 weeks of randomization and 112 days from the index UTI. Second and third DMSA scans were obtained around the 12- and 24-month follow-up visits. For 41 children, the final DMSA scan was done approximately 4 months after meeting the criteria for treatment failure, which were defined as identification of new or worsening renal scarring, two febrile UTI recurrences, one febrile and three symptomatic recurrences, or four symptomatic recurrences over the course of the study. For participants diagnosed with pyelonephritis at baseline (first), the DMSA scan was not repeated prior to the end of the first year unless the participant met the criteria for treatment failure. Only patients whose initial DMSA showed grade 3 or higher (> 4 kidney segments) scarring in either kidney and had a recurrence of UTI during the study had repeat scan before the end of the first year. All scans were performed at study participating sites. A pilot study was conducted to ensure uniformity in methodology and imaging quality for DMSA renal scans [19] and written guidelines were provided to each site regarding methodology for the scans.

All DMSA renal scans were reviewed independently by two blinded RRs (RR1 and RR2) and all differences were adjudicated by phone conversation to reach a consensus reading between the two. Methods for obtaining the scan and the definitions for their interpretation by RRs has been described previously [1]. The severity (grade) of scarring and pyelonephritis were reported as none, mild, moderate, and severe. Scarring could also be rated as global atrophy. All DMSA renal scans also had local reporting by a non-RR. Grading of renal involvement was assessed by dividing the renal cortex into 12 segments, and the severity was defined by the number of segments involved (Fig. 1). Scarring was distinguished from pyelonephritis on the basis of decreased radiotracer uptake associated with loss of contour or cortical thinning, and decreased uptake that was seen in preceding scan and not in the follow-up scan was considered as APN and not scarring.

For this report, two investigators (T.M. and S.S.) reviewed together the local radiologists' DMSA renal scan reports. Data were extracted from the local scan reports regarding severity (grade) of scarring and pyelonephritis and were translated into the same format as for the

RIVUR trial. DMSA renal scans ($n = 51$) that indicated a doubtful interpretation by the local radiologists themselves and the scans whose reporting was indefinable during the study review were categorized as “doubtful” and excluded from analysis. Also excluded from analysis were any scans for which a local and/or a reference assessment was not available and one child who was determined to have a horseshoe kidney. Altogether, 158 kidney units were excluded from the renal scarring analysis and 106 kidney units were excluded from the pyelonephritis analysis.

We compared DMSA scan assessments from all local radiologists as a collective single group with those of the two RIVUR RRs, both individually and as a single adjudicated (consensus) observer. Two-way comparisons of concordant and discrepant responses were analyzed using an unweighted kappa statistic between the local reports and the adjudicated RRs’ interpretations. All analyses were performed using SAS v 9.4 (SAS institute 2015) and significance was determined at the 0.05 level.

Results

RRs (RR1 and RR2) and ALRs evaluated a total of 2872 kidneys for renal scarring, 1,432 left kidneys and 1,440 right kidneys. The corresponding numbers for pyelonephritis were 2,924 kidneys, 1,461 left kidneys, and 1,463 right kidneys. The agreement rates for all DMSA scan reports between the RRs and ALRs was 93%.

Renal scarring

The comparison of adjudicated DMSA renal scan assessment of renal scarring between two RRs and ALRs is shown in Table 1. Of the 2,872 total kidney units evaluated, 119 (4%) were reported as having renal scarring by RRs compared with 212 (7%) by the ALRs. For the right kidney ($n = 1440$), 50 (3%) were reported as having renal scarring by RRs compared with 86 (6%) by ALRs. Corresponding numbers for the left kidney ($n = 1432$) were 69 (5%) and 126 (9%).

A comparison of non-adjudicated reporting between RRs and ALRs revealed that RR1 reported renal scarring in 50 (3%) right kidneys ($n = 1440$) compared with 86 (6%) by the ALRs; the corresponding numbers for left kidneys ($n = 1432$) were 68 (5%) and 126 (9%), respectively. RR2 reported renal scarring in 45 (3%) of the right kidneys ($n = 1439$) compared with 86 (6%) by the ALRs; the corresponding numbers for the left kidneys ($n = 1431$) were 69 (5%) and 126 (9%), respectively.

Grades of renal scarring as reported by ALRs and finally upgraded or downgraded by two RRs are shown in Table 2. For 51 (21%) kidneys there was no change in grading. Of the remaining 192 (79%), 59 (24%) were upgraded for scarring by the RRs and 133 (55%) were downgraded by the RRs for renal scarring.

Pyelonephritis

The comparison of adjudicated DMSA renal scan reports for pyelonephritis between two RRs and ALRs is shown in Table 1. Of the 2,924 kidneys units, 85 (3%) were reported to have pyelonephritis by the RRs compared with 151 (5%) by the ALRs. Of the 1,463 right

kidneys, 37 (3%) were reported to have pyelonephritis by the RRs compared with 63 (4%) by the ALRs; the corresponding numbers for the left kidney ($n = 1461$) were 48 (3%) and 88 (6%).

A comparison of non-adjudicated reporting between RRs and ALRs revealed that RR1 reported pyelonephritis in 37 (3%) right kidneys ($n = 1463$) compared with 63 (4%) by the ALRs; the corresponding numbers for left kidneys ($n = 1461$) were 48 (3%) and 88 (6%), respectively. RR2 reported pyelonephritis in 32 (2%) of the right kidneys ($n = 1463$) compared with 63 (4%) by the ALRs; the corresponding numbers for the left kidneys ($n = 1460$) were 47 (3%) and 88 (6%), respectively.

Grades of pyelonephritis as reported by the ALRs and finally upgraded or downgraded by two RRs are shown in Table 2. For 29 (15%) kidneys there was no change in grading. Of the remaining 166 (85%), 55 (28%) were upgraded for pyelonephritis by the RRs, and 111 (57%) were downgraded by the RRs for pyelonephritis.

Three-way comparison

The three-way comparison of independent DMSA reporting of all abnormal scans for renal scarring and pyelonephritis by the two RRs and the ALRs is shown in Table 3. Overall, for DMSA scans with renal scarring, all three (RR1, RR2, and ALRs) agreed in 46 (19%) cases, two agreed in 195 (80%) cases, and all three disagreed in three (1%) cases. For DMSA scans with pyelonephritis, all three agreed in 26 (13%) cases, two agreed in 167 (84%) cases, and all three disagreed in five (3%) cases.

The concordance rates for reporting of abnormal DMSA scans with scarring by the ALRs and adjudicated reporting by the RRs were 22% for right kidneys and 20% for left kidneys. For pyelonephritis, the corresponding numbers were 15% for the right and the left kidneys (Table 4).

Discussion

The clinical usefulness of the DMSA renal scan is dependent on the precision of its methodology and consistency in interpretation. The former may be somewhat easier to achieve by following well-established protocols. Interpretation, however, is reader dependent and can vary between and within interpreters when viewing scans on different occasions. De Sadeleer et al. [9] reported a study on 42 scans where the median percentage of agreement among 42 observers was 92%: 93.5% for normal and 90.5% for abnormal scans. Guevera et al. reported that in 96 kidneys from 46 children, interobserver agreement between three observers was seen in 75%, 78%, and 77% cases for an early DMSA scan during acute pyelonephritis, a late DMSA scan 6 months later, and a late DMSA with an earlier scan for comparison, respectively [7]. Patel et al. [10] reported a study on 57 scans, each read twice by two experts. High levels of intra- (95.9% and 90.6%, $p < 0.05$) and interobserver agreement (84.4%, $p < 0.05$) were demonstrated when using a standardized rating scale. However, all of these studies evaluated a small number of scans and included normal scans in the analyses.

In our study, the agreement rates for all DMSA scan reports between the RRs and ALRs was 93%. A three-way comparison of only abnormal scans for renal scarring and pyelonephritis by the two RRs and the ALRs revealed that all three radiologists agreed only in 19% and 14% of cases respectively. Our study also revealed that nearly twice as many renal units (left > right) were reported as having scarring by the ALRs compared with the adjudicated reports by the RRs (119 vs. 212). However, the difference was not significant when normal scans were included in the analysis (RR 4% and ALR 7%). Also, in nearly half of the instances, the grades of renal scarring reported by the ALRs were downgraded by the RRs. These numbers don't take into account the DMSA scan reports from the ALRs that were categorized as doubtful and excluded from the study, the inclusion of which would have increased the difference even further.

A DMSA renal scan is integral to the diagnosis and tracking of renal scarring. In children with VUR and UTI, DMSA scans have been used in a "bottom-up approach," following a voiding cystourethrogram (VCUG). Sometimes, the DMSA renal scan has been recommended as an initial investigation following UTI to assess renal injury and the possible need for a subsequent VCUG in selected children (top-down approach). The latter has been advocated by some guidelines in the management of UTI and VUR in children [20]. An erroneous interpretation of an acute DMSA renal scan in such cases could lead to an unwarranted VCUG or no VCUG with potential clinical consequences from missing a high-grade VUR.

Clinicians rely heavily on renal imaging results in the management of children with UTI/VUR. Abnormal findings influence short- and long-term clinical management, which may include additional renal imaging and surgical intervention. For patients wrongly diagnosed with pyelonephritis on a DMSA renal scan, the tendency is to follow up with another DMSA renal scan to confirm its resolution or progression to renal scarring. Those wrongly diagnosed with renal scarring may also have additional follow-up renal imaging, surgical correction of VUR, monitoring for hypertension and proteinuria, and counseling at the time of transfer to adult programs for potential pregnancy-related complications. In this study, local readings reported an excess of 93 (212–119) renal units with scarring and 66 (151–85) renal units with pyelonephritis compared with the central reading. Furthermore, 158 renal units that were reported to have scarring and 106 to have pyelonephritis on local reading and excluded from the study were reported as normal on central reading. We do not know how many patients could potentially have been affected if the clinical management was based entirely on local readings because clinical management varies from one practitioner to another and from place to place.

Various guidelines have been published to help improve the quality of DMSA renal scans and reduce variability in interpretation [14,15,21]. Some studies have shown that following the agreed guidelines for data interpretation does not significantly improve the consistency of reports [7]. Others have reported improvement in interobserver reproducibility and objectivity by using standardized criteria [8,10]. In the RIVUR trial, we conducted a pilot study to assess inconsistencies in imaging methodology at participating study sites, and standardized methodology for the procedure, interpretation of scans, and the process of adjudication between the two RRs [22]. The pilot study reported that despite guidelines,

considerable variation existed in the dose administered, methods of image acquisition, and image quality between the institutions. We do not know the impact of our pilot study on the reporting of DMSA renal scans in the RIVUR trial.

The strength of our report is the number of DMSA renal scans evaluated, which is significantly much more than any previous study. Also, the methods used for central reading of DMSA scans by two RRs, including the blinding and adjudication process, helped establish the gold standard for comparison with local readings by non-RRs. However, the study does have some limitations. The two investigators who reviewed the local reports for this study did not examine the DMSA scans. We could not confirm how compliant the study sites were with the methodology provided by RIVUR investigators. A lack of uniformity in local reporting made their review difficult and led to the exclusion of many scan reports from the study. Many reports by local radiologists lacked specificity and standardization, reflecting their clinical purpose. The reports often included statements like “possible mild cortical malfunction of the right upper pole,” “slightly diminished tubular volume at the right,” “if clinically indicated, a gallium scan is useful in distinguishing active pyelonephritis from renal scan,” “cortical thinning suggestive of chronic inflammation,” “may represent possible scar versus normal physiologic cortical variation,” “decreased uptake likely developmental secondary to reflux nephropathy,” and “findings compatible with scar or pyelonephritis.”

In conclusion a DMSA renal scan is an expensive and somewhat invasive investigation that is done mostly to evaluate renal parenchymal injury and differential renal function so that appropriate clinical decisions can be made. Many factors seem to affect inter- and intraobserver variability including the techniques used for the procedure, inadequate emphasis on following published guidelines for detailed reporting, expertise of the radiologist reading the scan, severity of renal injury, and the presence of any previous comparative study. Unlike many previous studies, our study revealed significant interobserver variability in the reporting of abnormal DMSA scans and highlights the need for optimizing the clinical yield of DMSA renal scans by standardized and uniform interpretation. In the meantime, clinicians taking care of such patients need to be aware of this limitation with the DMSA renal scans

Acknowledgments

The authors thank the RIVUR participants, their families and participating physicians, investigators and staffs for making this research possible. We also thank Lena Peschansky and Catherine Klida at Wayne State University for their help.

Funding

This research was supported by grants U01 DK074059, U01 DK074053, U01 DK074082, U01 DK074064, U01 DK074062, U01 DK074063 from the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health.

References

1. Mattoo TK, Chesney RW, Greenfield SP, et al. Renal Scarring in the Randomized Intervention for Children with Vesicoureteral Reflux (RIVUR) Trial. *Clin J Am Soc Nephrol*. 2016; 11(1):54–61. [PubMed: 26555605]
2. Faust WC, Diaz M, Pohl HG. Incidence of post-pyelonephritic renal scarring: a meta-analysis of the dimercapto-succinic acid literature. *Journal Urol*. 2009; 181:290–7. discussion 7–8.
3. Ransley PG, Risdon RA. Reflux nephropathy: effects of antimicrobial therapy on the evolution of the early pyelonephritic scar. *Kidney Int*. 1981; 20(6):733–42. [PubMed: 7038262]
4. Shindo S, Bernstein J, Arant BS Jr. Evolution of renal segmental atrophy (Ask-Upmark kidney) in children with vesicoureteric reflux: Radiographic and morphologic studies. *J Pediatr*. 1983; 102(6): 847–54. [PubMed: 6854447]
5. Patterson LT, Strife CF. Acquired versus congenital renal scarring after childhood urinary tract infection. *J Pediatr*. 2000; 136:2–4. [PubMed: 10636963]
6. Mattoo TK. Vesicoureteral reflux and reflux nephropathy. *Adv Chronic Kidney Dis*. 2011; 18:348–54. [PubMed: 21896376]
7. 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; 17:596–602. [PubMed: 8843119]
8. Jaksic E, Beatovic S, Zagar I, et al. Interobserver variability in 99mTc-DMSA renal scintigraphy reports: Multicentric study. *Nucl Med Rev Cent East Eur*. 1999; 2:28–33. [PubMed: 14600997]
9. De Sadeleer C, Tondeur M, Melis K, et al. A multicenter trial on interobserver reproducibility in reporting on 99mTc-DMSA planar scintigraphy: A Belgian survey. *J Nucl Med*. 2000; 41(1):23–6. [PubMed: 10647601]
10. 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. [PubMed: 8309749]
11. Craig JC, Irwig LM, Howman-Giles RB, et al. Variability in the interpretation of dimercaptosuccinic acid scintigraphy after urinary tract infection in children. *J Nucl Med*. 1998; 39(8):1428–32. [PubMed: 9708522]
12. 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; 29:87–91. [PubMed: 24761059]
13. Mandell GA, Egli DF, Gilday DL, et al. Procedure guideline for renal cortical scintigraphy in children. *Society of Nuclear Medicine. J Nucl Med*. 1997; 38:1644–6. [PubMed: 9379207]
14. Piepsz A, Blafox MD, Gordon I, et al. Consensus on renal cortical scintigraphy in children with urinary tract infection. *Scientific Committee of Radionuclides in Nephrourology. Semin Nucl Med*. 1999; 29(2):160–74. [PubMed: 10321827]
15. Piepsz A, Colarinha P, Gordon I, et al. Guidelines for 99mTc-DMSA scintigraphy in children. *Eur J Nucl Med*. 2001; 28(3):BP37–41. [PubMed: 11315615]
16. Keren R, Carpenter MA, Hoberman A, et al. Rationale and design issues of the Randomized Intervention for Children With Vesicoureteral Reflux (RIVUR) study. *Pediatrics*. 2008; 122(Suppl 5):S240–50. [PubMed: 19018048]
17. Hoberman A, Greenfield SP, et al. Investigators RT. Antimicrobial prophylaxis for children with vesicoureteral reflux. *N Engl J Med*. 2014; 370(25):2367–76. [PubMed: 24795142]
18. Carpenter MA, Hoberman A, Mattoo TK, et al. The RIVUR Trial: Profile and Baseline Clinical Associations of Children With Vesicoureteral Reflux. *Pediatrics*. 2013; 132:e34–45. [PubMed: 23753091]
19. Ziessman HA, Majd M. Importance of methodology on (99m)technetium dimercapto-succinic acid scintigraphic image quality: Imaging pilot study for RIVUR (Randomized Intervention for Children With Vesicoureteral Reflux) multicenter investigation. *J Urol*. 2009; 182:272–9. [PubMed: 19450818]
20. National Institute for Health and Clinical Excellence (NICE). *Urinary tract infection in children*. London: NICE; 2007.

21. Mandell GA, Cooper JA, Leonard JC, et al. Procedure guideline for diuretic renography in children. Society of Nuclear Medicine. J Nucl Med. 1997; 38:1647–50. [PubMed: 9379208]
22. Ziessman HA, Majd M. Importance of methodology on (99m)technetium dimercaptosuccinic acid scintigraphic image quality: imaging pilot study for RIVUR (Randomized Intervention for Children With Vesicoureteral Reflux) multicenter investigation. J Urol. 2009; 182:272–9. [PubMed: 19450818]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

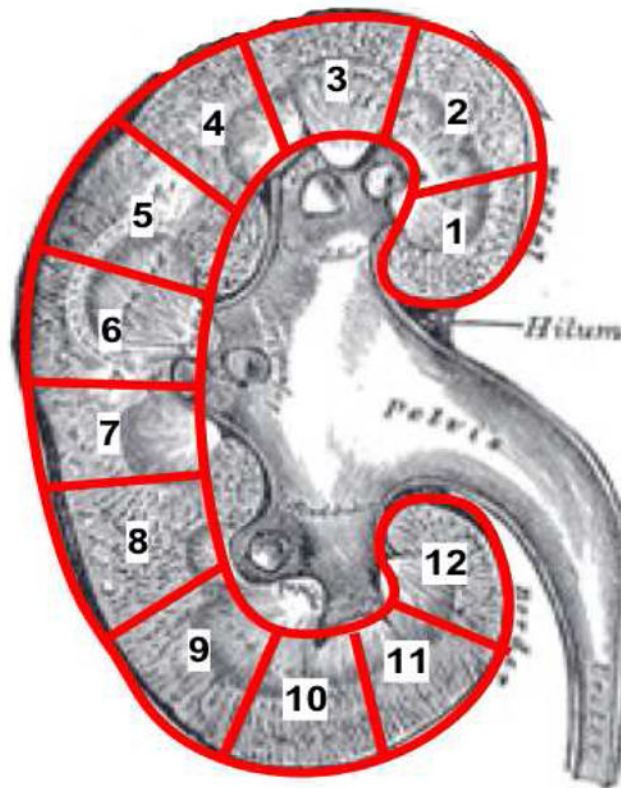


Figure 1. Kidney segments for DMSA renal scan interpretation

Grading system for characterizing the severity of renal involvement by pyelonephritis or renal scarring. Mild (Grade 1): 1–2 kidney segments affected; Moderate (Grade 2): 3–4 kidney segments affected; Severe (Grade 3): >4 kidney segments affected; Global atrophy (Grade 4): Diffusely scarred and shrunken kidney

Table 1

Cross-tabulation by radiologist.

Scarring grade		Reference radiologist adjudication									
		Right kidney					Left kidney				
Non-reference radiologists		None	Mild	Moderate	Severe	Global atrophy	None	Mild	Moderate	Severe	Global Atrophy
None		1342	8	3	1	0	1287	11	1	3	4
Mild		37	9	5	2	0	58	13	2	6	2
Moderate		11	3	4	5	0	12	5	9	3	1
Severe		0	0	0	6	1	0	0	1	4	1
Global Atrophy		0	0	0	0	3	6	0	0	0	3

Pyelonephritis grade		Reference radiologist adjudication									
		Right kidney					Left kidney				
Non-reference radiologists		None	Mild	Moderate	Severe	Global atrophy	None	Mild	Moderate	Severe	Global Atrophy
None		1382	13	4	1	1	1347	19	3	3	4
Mild		40	11	3	2	2	57	16	1	1	3
Moderate		4	1	1	1	1	6	0	1	1	1
Severe		0	0	0	0	0	3	0	0	0	0

Table 2

Change of grading by reference radiologist for scans reported with renal scarring or pyelonephritis by nonreference local radiologists.

Non-reference radiologist' DMSA grade reported as	<i>N</i> (%)
Scarring no different than RIVUR read	51 (21)
Pyelonephritis no different than RIVUR read	29 (15)
Upgraded by RIVUR reference radiologist	
One scarring grade lower than RIVUR read	36 (15)
Two scarring grades lower than RIVUR read	13 (5)
Three to four scarring grades lower than RIVUR read	10 (4)
One pyelonephritis grade lower than RIVUR read	38 (19)
Two to three pyelonephritis grades lower than RIVUR read	17 (9)
Downgraded by RIVUR reference radiologist	
One scarring grade higher than RIVUR read	104 (43)
Two scarring grades higher than RIVUR read	23 (10)
Three to four scarring grades higher than RIVUR read	6 (2)
One pyelonephritis grade higher than RIVUR read	98 (50)
Two to three pyelonephritis grades higher than RIVUR read	13 (7)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Agreement between non-reference radiologists, reference radiologist 1 and reference radiologist 2 (3-way comparison) on abnormal DMSA scans with scarring or pyelonephritis.

	All 3 radiologists agree N (%)	2 radiologists agree, and 1 disagrees N (%)	All 3 radiologists disagree N (%)
Scarring			
Right kidney (<i>n</i> = 1,439)	18 (18%)	79 (81%)	1 (1%)
Left kidney (<i>n</i> = 1,431)	28 (19%)	116 (80%)	2 (1%)
Pyelonephritis			
Right kidney (<i>n</i> = 1,462)	12 (15%)	65 (80%)	4 (5%)
Left kidney (<i>n</i> = 1,460)	14 (12%)	102 (87%)	1 (1%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Agreement between radiologists on abnormal DMSA scans with scarring or pyelonephritis.

	Concordant response no. (%)	Discrepant response no. (%)	Kappa (95% CI)
Non-reference radiologists compared with adjudicated reference radiologists			
Scarring			
Right kidney	22 (22)	76 (78)	0.02 (-0.10 to 0.14)
Left kidney	29 (20)	116 (80)	-0.01 (-0.10 to 0.08)
Pyelonephritis			
Right kidney	12 (15)	69 (84)	-0.30 (-0.44 to -0.15)
Left kidney	17 (15)	97 (85)	-0.30 (-0.42 to -0.17)
RIVUR radiologist 1 compared to RIVUR radiologist 2			
Scarring			
Right kidney	43 (84)	8 (16)	0.7869 (0.6575 to 0.9164)
Left kidney	65 (90)	7 (10)	0.8673 (0.7752 to 0.9594)
Pyelonephritis			
Right kidney	27 (73)	10 (27)	0.5334 (0.2959 to 0.7709)
Left kidney Level	42 (82)	9 (18)	0.6575 (0.4611 to 0.8539)