

# Evaluation of interobserver variability of parenchymal phase of Tc-99m mercaptoacetyltriglycine and Tc-99m dimercaptosuccinic acid renal scintigraphy

Zeynep Erdoğan, Ümmühan Abdülrezzak<sup>1</sup>, Güler Silov, Ayşegül Özdal, Özgül Turhal

Department of Nuclear Medicine, Kayseri Training and Research Hospital, Kayseri, <sup>1</sup>Nuclear Medicine, Erciyes University School of Medicine, Kayseri, Turkey

## ABSTRACT

**Objective:** The aim of this study was to investigate the variability in the interpretation of parenchymal abnormalities and to assess the differences in interpretation of routine renal scintigraphic findings on posterior view of technetium-99m dimercaptosuccinic acid (pvDMSA) scans and parenchymal phase of technetium-99m mercaptoacetyltriglycine (ppMAG3) scans by using standard criteria to make standardization and semiquantitative evaluation and to have more accurately correlation. **Materials and Methods:** Two experienced nuclear medicine physicians independently interpreted pvDMSA scans of 204 and ppMAG3 scans of 102 pediatric patients, retrospectively. Comparisons were made by visual inspection of pvDMSA scans, and ppMAG3 scans by using a grading system modified from Itoh *et al.* According to this, anatomical damage of the renal parenchyma was classified into six types: Grade 0-V. In the calculation of the agreement rates, Kendall correlation (tau-b) analysis was used. **Results:** According to our findings, excellent agreement was found for DMSA grade readings (DMSA-GR) (tau-b = 0.827) and good agreement for MAG3 grade readings (MAG3-GR) (tau-b = 0.790) between two observers. Most of clear parenchymal lesions detected on pvDMSA scans and ppMAG3 scans identified by observers equally. Studies with negative or minimal lesions reduced correlation degrees for both DMSA-GR and MAG3-GR. **Conclusion:** Our grading system can be used for standardization of the reports. We conclude that standardization of criteria and terminology in the interpretations may result in higher interobserver consistency, also improve low interobserver reproducibility and objectivity of renal scintigraphy reports.

**Keywords:** DMSA, MAG3, observer variability, renal scintigraphy

## INTRODUCTION

Renal scintigraphy using technetium-99m dimercaptosuccinic acid (DMSA) has become reference diagnostic test for renal cortical damage.<sup>[1,2]</sup> DMSA is taken up into proximal and distal renal tubular cells actively, and then approximately 2 h after injection, 40-65% of injected activity is concentrated in the cortex for a sufficiently long time to enable detailed scintigraphic evaluation<sup>[3]</sup> and is considered to be the agent of choice for diagnosis of cortical scarring, information regarding relative renal size because of high kidney-to-background ratio, lack of

activity in the collecting system, and lack of liver and bowel activity.<sup>[4,5]</sup>

Although technetium-99m mercaptoacetyltriglycine (MAG3) is a renal plasma flow agent secreting in the proximal tubules used mainly for assessment of renal drainage. Initial part of the study, the parenchymal phase, reflects the distribution of functional parenchyma and can give similar information to DMSA with regard to parenchymal damage, with sensitivity and specificity calculated as 88-89% and 88-100%, respectively.<sup>[6,7]</sup> Due to its high extraction rate, MAG3 accumulates rapidly in the cortex during the first few minutes after injection (40-60% per pass), while the background activity is also rapidly decreasing. Therefore, MAG3 provides sufficient image quality with a high kidney-to-background ratio and acceptable resolution at 2-4 min after injection and so, MAG3 has been also considered in the evaluation of renal parenchymal disorders.<sup>[6,8]</sup>

It is clear that, the value of a diagnostic test essentially depends on good intra- and interobserver reproducibility. Investigations

### Access this article online

#### Quick Response Code:



#### Website:

www.ijnm.in

#### DOI:

10.4103/0972-3919.130288

**Address for correspondence:** Dr. Zeynep Erdoğan, Department of Nuclear Medicine, Kayseri Training and Research Hospital, 38010 Kayseri, Turkey. E-mail: zeynep\_erdogan@yahoo.com

aiming specifically at the evaluation of agreement in the interpretation of renal scans, usually confirm the existence of interobserver variability. Nevertheless, while some authors point to sufficiently good agreement between observers,<sup>[9,10]</sup> others advocate that the agreement is poor.<sup>[11,12]</sup> Our current study was designed to determine possible sources of variability in the interpretation of degree, extent of parenchymal abnormality and to assess the differences in interpretation of routine renal scintigraphic findings by two nuclear medicine specialists. Posterior view of DMSA cortical (pvDMSA) scans and parenchymal phase of MAG3 (ppMAG3) scans were evaluated using standard criteria.

**MATERIALS AND METHODS**

Two experienced nuclear medicine physicians independently interpreted pvDMSA scans of 204 (seven patients with single kidney) and ppMAG3 scans of 102 (four patients with single kidney) pediatric patients (mean age of 7 years; range 1-16 years) who had undergone as part of their clinical workup retrospectively. The nuclear medicine physicians were blind to the laboratory results, other diagnostic tests, and patients’ clinical diagnosis prior to the evaluation.

Cortical scan with DMSA was performed approximately 2-4 h after the patients had been injected intravenously (IV) by an activity of “adult dose (MBq) × body weight (kg)/70”. Examinations were performed in planar mode. Posterior, anterior, left, and right posterior oblique static images (approximately 400k counts) were taken, using a gamma camera equipped with parallel hole low energy high resolution collimator (Siemens ECAM; Siemens Medical Systems, Hoffman Estates, IL, USA, 1999).

After sufficient hydration, MAG3 (2 MBq/kg) was injected IV while the patient was in a supine position. Simultaneous dynamic images were taken with the same gamma camera for 30 min. During the acquisition period, patients who had renal pelvic and/or pelvicalyceal activity accumulation were injected IV furosemide.

Comparisons were made by visual inspection of pvDMSA scans and ppMAG3 scans by using a grading system modified from Itoh *et al.*<sup>[13]</sup> According to this, anatomical damage of the renal parenchyma was classified into six types: Normal (Grade 0), one or two focal/relative decreased activity sides and/or single renal contour defect (Grade I), two renal contour defects but remnant areas of normal renal parenchyma and normal sized kidney (Grade II), diffuse reduction in uptake throughout the whole kidney with or without multiple renal contour defects (Grade III), small or shrunken kidney (Grade IV), indistinct margins of kidney (Grade V).

The results of all evaluations were entered in Statistical Package for Social Sciences (SPSS) 21.0. The degree of agreement between the readers was measured with the Kendall’s tau-b correlation coefficient as for ordinal-level variables. Values less

than 0.2 are associated with very poor agreement, 0.2-0.40 slight agreement, 0.4-0.6 moderate agreement, 0.6-0.8 substantial (good, high) agreement, and values greater than 0.8 are associated with excellent (almost perfect) agreement.<sup>[14]</sup>

**RESULTS**

According to our findings, excellent agreement was found for DMSA grade readings (DMSA-GR) (tau-b = 0.827) and good agreement for MAG3 grade readings (MAG3-GR) (tau-b = 0.790) between two observers. Agreement between DMSA-GR for left kidneys (tau-b = 0.851) and for right kidneys (tau-b = 0.803) were almost perfect. MAG3-GR for left and right kidneys among two observers were good (left: tau-b = 0.795; right: tau-b = 0.778). Also given these findings, interobserver reproducibility was better for DMSA-GR than MAG3-GR and was better for left kidneys than right among two observers.

The results showed that most of clear parenchymal lesions (Grade III-V) detected on pvDMSA scans and ppMAG3 scans identified by observers equally. Studies with negative or minimal lesions reduced correlation degrees for both DMSA-GR and MAG3-GR [Tables 1 and 2].

An analysis of grade readings by two observers, as shown in Tables 1 and 2, led to the conclusion that observer I had an obvious tendency to see abnormality or scars that were not perceived by remaining colleague. This tendency was especially pronounced between Grades “0” and “I”. Figure 1a and 2a show perfect agreement on grade of severity, but Figure 1b and 2b show those for which interpretations differed.

**Table 1: Comparison of MAG3-GR among two observers**

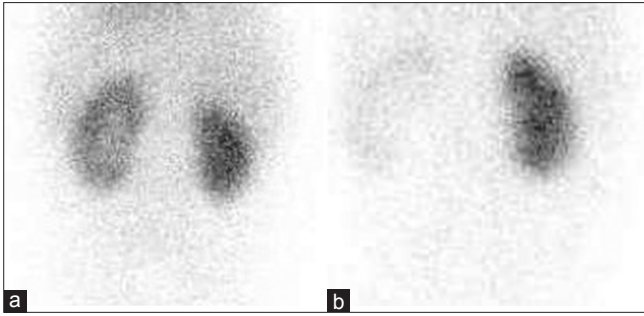
Observer I	MAG3 grade						Total
	Observer II						
	0	I	II	III	IV	V	
0	30	4	0	0	0	0	34
I	35	59	0	1	0	0	95
II	1	9	12	1	0	0	23
III	1	1	1	24	0	0	27
IV	0	0	0	0	15	0	15
V	0	0	0	0	0	6	6
Total	67	73	13	26	15	6	200

MAG3-GR: Mercaptoacetyltriglycine grade readings

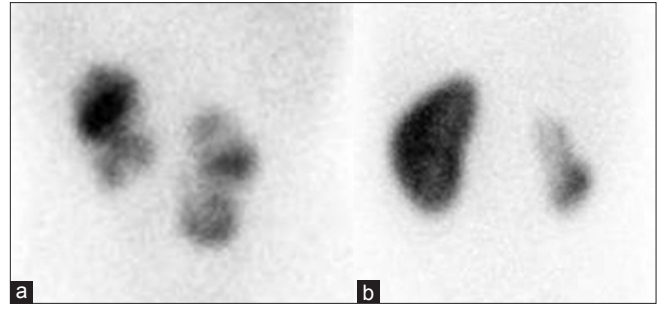
**Table 2: Comparison of DMSA-GR among two observers**

Observer I	DMSA grade						Total
	Observer II						
	0	I	II	III	IV	V	
0	98	6	0	0	0	0	104
I	52	136	4	1	0	0	193
II	0	6	14	0	0	0	20
III	1	3	6	37	0	0	47
IV	0	0	0	1	29	0	30
V	0	0	0	0	0	7	7
Total	151	151	24	39	29	7	401

DMSA-GR: Dimercaptosuccinic acid grade readings



**Figure 1:** (a) ppMAG3 scan of a 1-year-old girl with left Grade V vesicoureteral reflux. According to our grading system, left kidney's MAG3-GR was "III" and right kidney's MAG3-GR was "I" for both observers. (b) ppMAG3 scan of a 2-year-old boy with posterior ureteral valve and vesicoureteral reflux. Both observers reported left kidney as Grade "V", right kidney as Grade "I" for first observer and as Grade "0" for second observer



**Figure 2:** (a) pvDMSA scan of a 8-year-old girl with left Grade V vesicoureteral reflux. According to our grading system, left kidney's DMSA-GR was "III" and right kidney's DMSA-GR was "III" for both observers. (b) pvDMSA scan of a 6-year-old girl with urinary tract infection. Both observers reported right kidney as Grade "IV", left kidney as Grade "I" for first observer and as Grade "0" for second observer

## DISCUSSION

In clinical practice, DMSA scan considered the standard reference method in the assessment of renal cortical lesions. Interpretation is usually only qualitative and differences in reproducibility have been reported previously.<sup>[9,15-18]</sup> MAG3 is mainly a tubular renal imaging agent and used for planar dynamic studies to assess renal parenchymal flow and function and to determine drainage adequacy of the kidneys.<sup>[19,20]</sup> However, due to its high extraction efficiency, MAG3 provides high resolution parenchymal images 1-4 min after injection and may be suitable for renal cortical scintigraphy. Some investigators have evaluated the use of planar dynamic MAG3 scan for investigating the renal parenchymal lesions and variable results were reported in different studies.<sup>[1,2,8]</sup>

The diagnostic value and usability of a method beside accuracy, sensitivity, specificity, and positive and negative predictive values of the method is repeatability and is also associated with high degrees of intra- and interobserver agreements. High level of agreement means that the method is less dependent on the reader.

DMSA and MAG3 scan results affect dramatically diagnosis, treatment type, duration, and follow-up of patients. Correct identification of the scintigraphic results will play an important role in determining the next step. Therefore, the reliability of the obtained results is mandatory. One of the important indicators of the reliability of a test is intra- and interobserver agreement. For this purpose, there are many studies which investigate intra- and interobserver agreement with evaluation of planar images of DMSA scans, but there are not enough reliable references in the literature on scintigraphic reports in terms of objectivity and standardization. In studies researching agreement between observers by evaluating planar images of DMSA scan have presented different results. While some authors represented good agreement between observers and others suggest that the agreement is poor.<sup>[9-12]</sup> Investigations aiming specifically at the evaluation of agreement in the interpretation of renal scintigraphies usually verify the presence of interobserver variability. The use of standard evaluation criteria is known to increase compliance, therefore, in our study which investigated

interobserver agreement, standardized criteria used for assessment of DMSA scan and the results for each parameter was evaluated independently. The same standardized criteria were applied for MAG3 scan. In the literature, there was no comparative study of inter- and intraobserver agreements of MAG3 scan evaluating the renal parenchyma with standard criteria.

Good reproducibility was reported when two to four observers had to choose two or three different parameters between normal, abnormal, or equivocal.<sup>[9,18]</sup> In contrast, poor correlation was found when six to seven observers had to quantify the number of scars or to analyze seven different parameters.<sup>[11,12]</sup>

In a study designed by Craig *et al.*,<sup>[21]</sup> the kidneys divided into three regions and these regions were evaluated according to Goldraich's grading system and for DMSA planar images high level of agreement was found similar to our study. Goldraich's graduation displayed a bimodal score distribution which also confirms observations made by Craig *et al.* The scale, as proposed by Goldraich bases the grading on the sum of two separate features: Radiopharmaceutical uptake characteristics and the number of renal defects. In a previous study assessing intra- and interobserver variability in the interpretation of DMSA scan using standard criteria by Patel *et al.*, high levels of intra- (95.9%) and interobserver agreement (84.4%) were demonstrated<sup>[9]</sup> and there were minor differences in inconsistencies between two kidneys or different kidney zones. In our study, excellent agreement was found among two observers for DMSA-GR and high level for MAG3-GR. The reason for why agreement was better among DMSA-GR than MAG3-GR ( $\tau\text{-}b = 0.827$  and  $\tau\text{-}b = 0.790$ , respectively) will be associated with higher background activity and a little excretion of MAG3 through the liver because agreement level was higher in the left kidney than the right. Interobserver variability, especially between Grades 0 and I was more than the other grades. Presence of normal variants of the evaluated images may act as an important factor in the differentiation of normal to abnormal. So, interobserver differentiation of assessments may result from high anatomic variability of kidneys, from persisting fetal lobulation, evident structures of linking parenchyma, the doubling of the

pyelocalyceal system, etc., Differences in the levels of experience among the observers will be another factor that may reduce the degree of agreement. It was obvious that the less experienced observer tended to see more renal scars than those with more experienced. Furthermore, we used only pvDMSA scans for comparison. With addition of oblique views will provide a firm basis for clinical decision-making, particularly in kidneys which have photon deficient areas due to dilated calices or unclear contours and with only posterior views, amount of abnormal scans will be overestimated.

The shape, severity, and extent of the renal lesions in the patient groups are also the factors that may affect the level of the concordance of assessment. While severity of lesion increases and becomes significant, agreement also increases. In our study, with increasing grades, agreement moved towards complete agreement. The number of criteria used for evaluation in a study affects the statistical correlation, correlation decreases with increasing numbers of criteria as can be seen in our study.<sup>[14]</sup>

In a study with DMSA, IV urography and ultrasonography in 27 children at risk for renal scarring, the percentages of agreement in DMSA scan interpretation for three observers were 90 and 95% for interobserver and intraobserver comparisons, respectively. But, no detail was given how these values determined.<sup>[15]</sup>

In a recent survey reported by De Sadeleer *et al.*, the overall reproducibility was excellent among a large number of nuclear medicine physicians. The effect of being from the same or different centers was investigated, despite the relatively high agreement among those with the same center, but there was no statistically significant difference between them.<sup>[16]</sup> Current study provides data about the reproducibility only for a single center and both observers trained in the same center. Further studies are needed to assess for evaluating reproducibility of different centers with the standard criteria.

In evaluation of scintigrams, routine application of standard criteria like grading systems, for patients especially who need follow-up, the difference between a previous or subsequent scans can be clearly shown semiquantitatively and can reduce the dependence on the reader. Also standardization can be achieved in interpretations between different centers and observers. To minimize differences and to obtain sufficient image quality, examinations of patients should be performed in optimal conditions with optimal devices. Because final reports will severely affect the patient's treatment protocol, images should be evaluated systematically and in appropriate time and to act in conjunction with clinicians will be useful. For a better assessment of the differences in interpretation between observers, such studies with larger series of patients should be performed.

## CONCLUSION

Although poor correlation was found in the literature when the observers had to quantify the number of scars and parenchymal

lesions or to analyze more different parameters, in our study interobserver reproducibility was excellent for pvDMSA scans and good for ppMAG<sub>3</sub> scans among two observers. Our modified grading system can be used for standardization of reports. Such as normal variants, some congenital abnormalities and small defects probably constitute the main cause of disagreement among observers in this study. Disagreement among observers could be reduced by taking the normal variants into account. We conclude that standardization of criteria and terminology in interpretations may result in higher interobserver consistency, improve low interobserver reproducibility and objectivity of renal scintigraphy reports, and improve the quality of reporting of studies. We propose quantitative measures that reflect some common parameters observed by specialists in the renal scintigraphic studies, attributing numerical values to these parameters, and potentially reducing the subjectivity in the interpretation of visual findings.

## REFERENCES

- Gordon I. Indications for technetium-99m dimercaptosuccinic acid scan in children. *J Urol* 1987;137:464-7.
- 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;148:1726-32.
- Mandell GA, Eggl DF, Gilday DL, Heyman S, Leonard JC, Miller JH, *et al.* Procedure guideline for renal cortical scintigraphy in children. *J Nucl Med* 1997;38:1644-6.
- Rossleigh MA. Renal cortical scintigraphy and diuresis renography in infants and children. *J Nucl Med* 2001;42:91-5.
- Majd M, Rushton HG. Renal cortical scintigraphy in the diagnosis of acute pyelonephritis. *Semin Nucl Med* 1992;22:98-111.
- Gordon I, Anderson PJ, Lythgoe MF, Orton M. Can technetium-99mmercaptoacetyl triglycine replace technetium-99m-dimercaptosuccinic acid in the exclusion of a focal renal defect? *J Nucl Med* 1992;33:2090-3.
- Bair HJ, Becker W, Schott G, Kühn RH, Wolf F. Is there still a need for Tc-99m DMSA renal imaging? *Clin Nucl Med* 1995;20:18-21.
- Piepz A, Pintelon H, Verboven M, Keuppens F, Jacobs A. Replacing 99m-Tc-DMSA for renal imaging. *Nucl Med Commun* 1992;13:494-6.
- 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;23:506-9.
- Ladron De Guevara D, Franken P, 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;42:564-6.
- Jaksić E, Beatović S, Zagar I, Punković N, Stefanović A, Ajdinović B, *et al.* Interobserver variability in 99mTc-DMSA renal scintigraphy reports: Multicentric study. *Nucl Med Rev Cent East Eur* 1999;2:28-33.
- 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.
- 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;9:23-8.
- Altman DG. *Practical statistics for medical research.* London: Chapman and Hall; 1991, p. 285.
- Shanon A, Feldman W, McDonald P, Martin DJ, Matzinger MA, Shillinger JF, *et al.* Evaluation of renal scars by technetium-labeled dimercaptosuccinic acid scan, intravenous urography, and ultrasonography: A comparative study. *J Pediatr* 1992;120:399-403.
- De Sadeleer C, Tondeur M, Melis K, Van Espen MB, Verelst J, Ham H, *et al.* A multicenter trial on interobserver reproducibility in reporting on 99mTc-DMSA planar scintigraphy: A Belgian survey. *J Nucl Med* 2000;41:23-6.
- Stokland E, Hellstrom M, Jacobsson B, Jodal U, Sixt R. Renal damage one year after first urinary tract infection: role of dimercaptosuccinic acid scintigraphy. *J Pediatr* 1996;129:815-20.

18. Everaert H, Flamen P, Franken PR, Peeters P, Bossuyt A, Piepsz A. 99mTc-DMSA renal scintigraphy for acute pyelonephritis in adults: Planar and/or SPET imaging? Nucl Med Commun 1996;17:884-9.
19. Taylor A, Eshima D, Frizberg AR, Christian PE, Kasina S. Comparison of iodine-131 OIH and technetium-99m MAG<sub>3</sub> renal imaging in volunteers. J Nucl Med 1986;27:795-803.
20. Eshima D, Taylor A Jr. Technetium-99m mercaptoacetyl-triglycine: Update on the new (99m-Tc) renal tubular function agent. Semin Nucl Med 1992;22:61-73.
21. Craig JC, Irwig L, Ford M, Willis NS, Howman-Giles RB, Uren RF, *et al.* Reliability of DMSA for the diagnosis of renal parenchymal abnormality in children. Eur J Nucl Med 2000;27:1610-6.

**How to cite this article:** Erdoğan Z, Abdülrezzak Ü, Silov G, Özdal A, Turhal Ö. Evaluation of interobserver variability of parenchymal phase of Tc-99m mercaptoacetyl triglycine and Tc-99m dimercaptosuccinic acid renal scintigraphy. Indian J Nucl Med 2014;29:87-91.

**Source of Support:** Nil. **Conflict of Interest:** None declared.