
Reproducibility of Renal Length Measurements with ^{99m}Tc -DSMA SPECT

Eva Lin, Leonard P. Connolly, David Zurakowski, James DiCanzio, Laura Drubach, Karl Mitchell, Tracy Tetrault, Stephen P. Laffin, and S. Ted Treves

Division of Nuclear Medicine, Department of Radiology, Children's Hospital, Harvard Medical School, Boston, Massachusetts

Renal length measurements are used in evaluating several abnormalities of the pediatric genitourinary tract. This study assesses reproducibility of renal length measurements obtained with ^{99m}Tc -dimercaptosuccinic acid (DMSA) SPECT. **Methods:** The lengths of 98 kidneys of 51 children (age range, 1–16 y; mean age, 5.4 y) who underwent ^{99m}Tc -DMSA SPECT were measured independently by 2 observers. Renal length was calculated by converting pixels between points at the superior and inferior renal margins on a summated coronal image to centimeters. Lengths were measured for kidneys as they appeared in situ and after realignment along their long axes. SPECT reconstruction, choice of display parameters, positioning of points used for measuring, and alignment were performed independently by each observer. Interobserver variability, interobserver correlation, and mean differences between observers' measurements (expressed as measurement of observer 2 – measurement of observer 1) were calculated. **Results:** Correlation between the observers' measurements was highly significant for both nonaligned and aligned studies ($r = 0.95$ and 0.97 , respectively; both, $P < 0.0001$). Interobserver variability expressed as 1 SD was 3.6 mm for nonaligned studies and 2.8 mm for aligned studies. The mean difference between the 2 observers' measurements for nonaligned studies was 2.0 ± 4.8 mm ($P < 0.0001$) with a range of -11 to 14 mm. For aligned studies the mean difference between the 2 observers' measurements was -0.1 ± 4.0 mm ($P = 0.88$) with a range of -20 to 10 mm. Differences between observers were not dependent on absolute renal length ($P = 0.68$ for nonaligned studies; $P = 0.40$ for aligned studies). **Conclusion:** The variability in renal length measurements determined by ^{99m}Tc -DMSA SPECT is similar to that reported previously using sonography. Because the interobserver differences in renal length are similar to annual renal growth rates during childhood, caution should be applied when incorporating renal length measurements determined by ^{99m}Tc -DMSA SPECT into management algorithms. Additional studies are required to further establish interobserver variability, to assess intraobserver variability, and to evaluate means of improving standardization.

Key Words: kidney; SPECT; ^{99m}Tc -DMSA

J Nucl Med 2000; 41:1632–1635

Assessment of renal size is used in the evaluation of several abnormalities of the pediatric genitourinary tract (1). Sequential measurements of renal length as a child grows are used to determine if renal growth has been appropriate. Measurements of renal length in a child of a given age or overall body size (or both) are compared with the renal length that is predicted by standard nomograms (2–6). Sonography is the imaging modality most commonly used for this purpose, although limitations and pitfalls of sonographically determining renal length are well recognized (7).

^{99m}Tc -dimercaptosuccinic acid (DMSA) scintigraphy has assumed an important role in the evaluation of pediatric genitourinary abnormalities. ^{99m}Tc -DMSA scintigraphy is used primarily to detect cortical abnormalities, for which purpose SPECT is more sensitive than planar imaging (8,9), and to assess differential renal function. ^{99m}Tc -DMSA scintigraphy can also provide renal size measurements, but little is known about the reliability of such measurements. This study assesses the reproducibility of renal length measurements obtained with ^{99m}Tc -DMSA SPECT.

MATERIALS AND METHODS

A retrospective study was performed using 51 studies of 51 consecutively referred children (30 females, 21 males) who had ^{99m}Tc -DMSA scintigraphy with SPECT during a 3-mo period at our institution. The mean age was 5.4 y (range, 1–16 y). The absence of patients younger than 12 mo of age reflects a decision that we have made at our institution to perform pinhole imaging rather than SPECT in such patients on the basis of limitations in resolution related to the proximity with which the heads of our triple-detector gamma camera approach small children. A total of 98 kidneys were studied (only 1 kidney was visualized in 4 cases).

All children received a weight-adjusted dose of ^{99m}Tc -DMSA intravenously (1.85 MBq/kg). SPECT was performed 4 h after injection using a Multispect 3 system (Gamma Sonics; Siemens Medical Systems, Hoffman Estates, IL) equipped with an ultrahigh-resolution collimator. Images were acquired on a 128×128 matrix. A noncircular 360° orbit with 120 stops (40 stops per detector) was used. A Butterworth filter was used (0.3–0.35 cycle/cm cutoff frequency).

For renal length measurement, the tips of the superior and inferior poles of each kidney were defined manually on the basis of visual inspection of summated coronal SPECT images. The pixel distance between these 2 points was automatically calculated and converted to centimeters using a calibration factor determined for

Received Oct. 4, 1999; revision accepted Feb. 16, 2000.

For correspondence or reprints contact: Leonard P. Connolly, MD, Division of Nuclear Medicine, Children's Hospital, 300 Longwood Ave., Boston, MA 02115.

the gamma camera. Measurement was performed for each kidney as it appeared in its anatomic position (Figs. 1A and B). A separate measurement was performed for each kidney after computer-assisted alignment of the kidney along its long axis (Figs. 1C and D). Two observers independently performed all steps involved in renal measurement and realignment. They were unaware of each other's results.

Interobserver variability, interobserver correlation, and mean differences between observers' measurements (expressed as measurement of observer 2 - measurement of observer 1) were calculated. Pearson coefficients (r) were calculated to evaluate the interobserver correlation for nonaligned and aligned studies. Paired t tests were used to assess differences between observers and between methods. The 95% confidence intervals for interobserver differences were determined by the Student t distribution (10). The Bland and Altman method (11) was used to determine whether differences between observers were dependent on absolute renal length for nonaligned and aligned studies. Statistical analysis was performed using version 9.0 of the SPSS software package (SPSS Inc., Chicago, IL). Analysis was conducted on measurements of 98 kidneys for nonaligned studies and measurements of 97 studies for aligned studies. The lower number of kidneys aligned reflects the fact that 1 observer found 1 kidney to be too poorly visualized to be aligned accurately.

RESULTS

Correlation between observers was highly significant for both nonaligned and aligned studies ($r = 0.95$ and 0.97 , respectively; both, $P < 0.0001$). Interobserver variability, expressed as 1 SD, was 3.6 mm for nonaligned studies and 2.8 mm for aligned studies. The smaller variability for aligned studies compared with nonaligned studies was statistically significant ($P = 0.004$). The mean difference between the 2 observers' measurements for nonaligned studies was 2.0 ± 4.8 mm ($P < 0.0001$) with a range of -11 to 14 mm. For aligned studies the mean difference between the 2 observers' measurements was -0.1 ± 4.0 mm

($P = 0.88$) with a range of -20 to 10 mm. Bland and Altman plots (11) are presented in Figure 2. Slope tests based on linear regression indicated that differences between observers were not dependent on absolute renal length ($P = 0.68$ for nonaligned studies; $P = 0.40$ for aligned studies). Renal length measurements increased significantly ($P < 0.0001$) for both observers when studies were aligned with 1 observer recording a mean increase of 5.3 mm and the other a mean increase of 3.2 mm. The results are summarized in Table 1.

DISCUSSION

Judicious clinical use of a length measured by an imaging test requires understanding the degree to which measurement error or variability might affect that measurement. From a clinical standpoint, interobserver variability and mean interobserver difference in measurements are more meaningful gauges of interobserver reproducibility than is interobserver correlation, which can be high even when interobserver agreement is poor (11).

Interobserver variability and mean interobserver difference in renal length measurements of children must be considered in the context of expected annual renal growth during childhood. Renal growth rates of 1–1.5 mm/mo during the first year of life and 2–4 mm/y thereafter until puberty have been described using ^{99m}Tc -DMSA scintigraphy (12) and sonography (5). With this as a frame of reference, the interobserver variability of 3.6 mm for nonaligned studies and 2.8 mm for aligned studies as well as the mean difference (\pm SD) between observers' measurements of 2.0 ± 4.8 mm for nonaligned studies and of 0.1 ± 4.0 mm for aligned studies can be considered sizable. Failure of a kidney to grow for 2–3 y could easily go undetected if one were to use limits equal to 2 SDs (95%

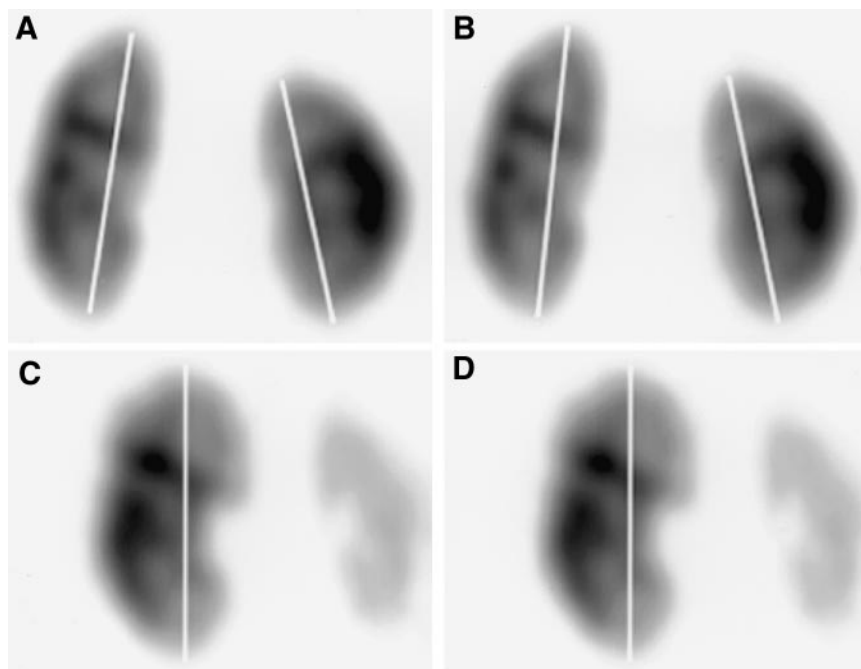


FIGURE 1. Renal length measurements using coronal SPECT displayed as posterior projection. On nonaligned study, observer 1 recorded length measurements of 80 mm for left kidney and 70 mm for right kidney (A), whereas observer 2 recorded measurements of 82 mm for left kidney and 71 mm for right kidney (B). After alignment of left kidney along its long axis, observer 1 (C) and observer 2 (D) both measured kidney as being 83 mm in length.

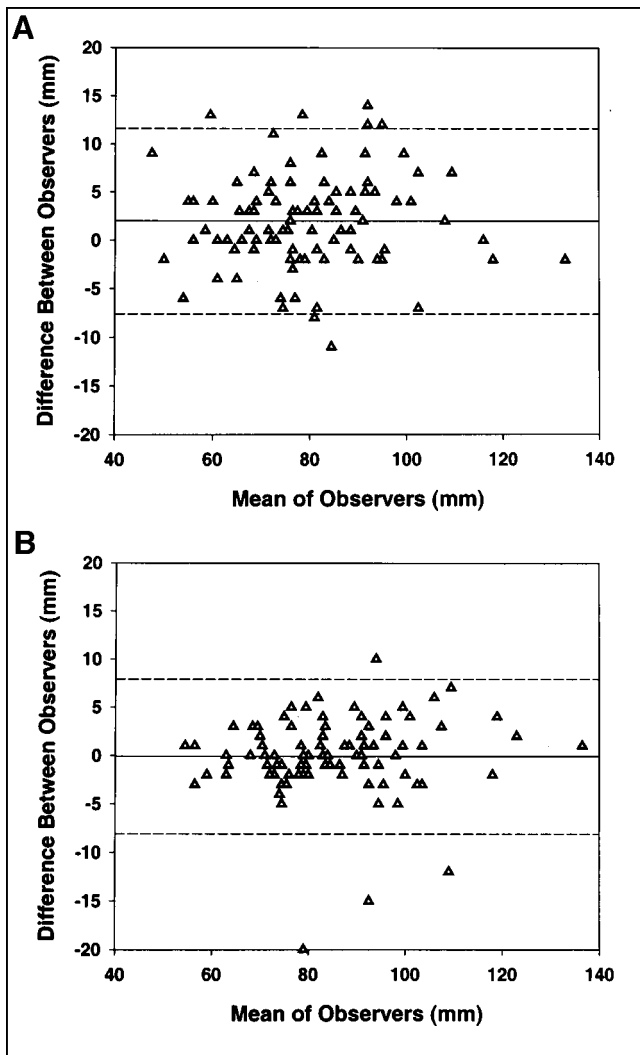


FIGURE 2. Difference between observers' measurements plotted against mean of measurements for nonaligned (A) and aligned (B) studies. For each plot, solid line represents mean interobserver difference, and dashed lines indicate 95% confidence interval for individual studies.

confidence interval) of the interobserver variability and differences reported here.

Limited reproducibility of renal length measurements has also been reported using sonography. Schlesinger et al. (13) reported that the mean (\pm SD) of the absolute values of the

TABLE 1
Summary of Results

Kidneys	Nonaligned	Aligned
Number	98	97
Mean length \pm 1 SD (mm)		
Observer 1	78.4 \pm 15.0	83.7 \pm 15.3
Observer 2	80.4 \pm 15.2	83.6 \pm 15.6
Interobserver correlation	0.95	0.97
Mean interobserver variability \pm 1 SD (mm)	3.6	2.8
Mean interobserver difference \pm 1 SD (mm)	2.0 \pm 4.8	-0.1 \pm 4.0

differences between 2 renal length measurements performed on the same kidney by 2 sonologists ranged from 3.87 ± 2.39 mm to 5.49 ± 3.27 mm. Sargent and Wilson (14) found that the mean difference (\pm SD) of 2 renal length measurements performed by 2 sonologists ranged from 0.21 ± 4.81 mm to 0.96 ± 3.95 mm for kidneys considered as normal and from 0.39 ± 6.6 mm to 1.83 ± 6.58 mm for kidneys considered as abnormal. The magnitude of the differences between measurements obtained by 2 sonologists in a series of Carrico et al. (15) ranged from 3.1 to 3.6 mm.

This study does not assess the precision of renal length measurements with ^{99m}Tc -DMSA SPECT. In animal models, measurement by ^{99m}Tc -DMSA planar imaging has been shown to overestimate actual renal length despite the foreshortened appearance of the kidneys associated with their anatomic tilt that would be expected on anterior or posterior images. Wallin et al. (16) reported that renal length in piglets was overestimated by 1.7% compared with measurements of excised kidneys and by 3.9% compared with measurements of in vivo kidneys. Rossleigh et al. (17) found that renal length measurement by ^{99m}Tc -DMSA planar imaging was higher by 6% than was measurement on pathologic examination. Scintigraphic measurements using SPECT on a kidney phantom have overestimated length by 5% (16). In contrast, animal models have shown sonography to underestimate renal length by as much as 22% (17,18). This indicates a potential pitfall in comparing renal length determined sonographically with renal length measured by ^{99m}Tc -DMSA scintigraphy.

The establishment of standard nomograms for kidney length on the basis of ^{99m}Tc -DMSA SPECT is also not addressed by this study. We doubt that valid standard nomograms are likely to be developed using ^{99m}Tc -DMSA scintigraphy. Performing ^{99m}Tc -DMSA scintigraphy on children without known or suspected genitourinary abnormalities to obtain a normal database is not justifiable from an ethical standpoint. The use of kidneys contralateral to kidneys with a known or suspected abnormality is not suitable for establishing norms because growth alterations of a kidney contralateral to a kidney affected by genitourinary pathology is a well-known occurrence. Derivation of nomograms from ^{99m}Tc -DMSA studies interpreted as normal in children being evaluated for suspected renal disease is 1 strategy that has been applied with planar imaging (19). Although this is probably the most suitable alternative, it is subject to error because of renal damage that is below the resolution of scintigraphy or below the interpretive threshold of the individuals evaluating the studies. Standard nomograms derived from sonography (2-6) are not directly applicable to measurements obtained with ^{99m}Tc -DMSA scintigraphy because of the factors discussed above regarding precision. Standard nomograms derived from intravenous urography (20) are also not directly applicable to renal length measurements determined by ^{99m}Tc -DMSA SPECT because numerous factors, particularly increase in renal size

associated with the osmotic effect of iodinated contrast agents, affect renal length measurement by intravenous urography. Any standard nomogram reflects the norms only for the population for which it was derived and may not be appropriate for use in individuals of different ethnicity, different geographic location, or different socioeconomic status. This problem has been encountered using standard nomograms derived from sonography (7). An additional problem regarding nomograms of renal length in children relates to the expected yearly growth of kidneys and the wide variation of renal lengths encountered in normal children. Using a conventional normal range of 2 SDs, the same kidney size can be considered normal across a range of 6 or more years with 1 widely used nomogram (7).

These findings suggest that sequential measurements based on ^{99m}Tc -DMSA SPECT are likely to be more meaningful for a particular patient than is a single measurement at any given age. Therefore, improving reproducibility of renal length measurements may be more important than assessing their precision or establishing a normal database. Greater reproducibility of renal length measurements after alignment of a kidney along its long axis in our patients suggests that alignment is 1 step that can be taken toward this end. Because placement of the points used for length determination at the margins of the kidney is a likely source of interobserver variability, enhanced standardization of this step may be beneficial. For example, using an edge detection algorithm to identify the renal margins might decrease interobserver variability and differences below what we attained measuring renal length by visual inspection of the images. Another possible strategy might be to select an image showing the maximum length of each kidney by viewing rotating volume-rendered images between the cursors used to set the distance over which the transaxial planes will be reconstructed.

This study is limited in that it shows interobserver differences for only 2 observers at 1 institution. The applicability of the results to other institutions or even to additional observers at the same institution is not established. Intraobserver variability is not assessed. Despite these limitations, the results provide insight into problems involved in using renal length measurements in managing children with genitourinary tract abnormalities. On the basis of our findings, the caution that others (13,14) have advised regarding incorporation of sonographic renal measurements into management algorithms should be extended to ^{99m}Tc -DMSA SPECT renal length measurements.

Another limitation of this study is that it does not assess how measurements obtained by planar techniques differ from those obtained with SPECT. Although the similarity between the previously cited results comparing planar with pathologic measurements (17) and SPECT measurements with phantom size (16) suggests that measurements obtained with planar imaging and SPECT might not differ significantly, that has not been established. In the absence of data indicating higher precision or reproducibility of measurements obtained with SPECT relative to those obtained with planar imaging, individual practitioners may opt to supplement SPECT with a single planar image to be

used for length assessment rather than perform the steps described and suggested here.

CONCLUSION

Interobserver differences in measurement of renal length with ^{99m}Tc -DMSA SPECT are similar to those that have been reported previously for sonographically determined renal length measurements. Interobserver differences impose significant limitations on the use of sequential renal length measurements in evaluating renal growth during childhood. Additional studies are required to further establish interobserver variability, to assess intraobserver variability, and to evaluate means of improving standardization. Only with such studies can the appropriateness of assessing longitudinal renal growth with ^{99m}Tc -DMSA SPECT during childhood be determined.

REFERENCES

1. Barnewolt CE, Paltiel HJ, Lebowitz RL, Kirks DR. Genitourinary tract. In: Kirks DR, ed. *Practical Pediatric Imaging: Diagnostic Radiology of Infants and Children*. 3rd ed. Philadelphia, PA: Lippincott-Raven; 1997:1009–1170.
2. Blane CE, Bookstein FI, DiPietro MA, Kelsch RC. Sonographic standards for normal infant kidney length. *AJR*. 1985;145:1289–1291.
3. Dinkel E, Ertel M, Dittrich M, Peters H, Berres M, Schulte-Wisserman H. Kidney size in childhood: sonographical growth charts for kidney length and volume. *Pediatr Radiol*. 1985;15:38–43.
4. Han BK, Babcock DS. Sonographic measurements and appearance of normal kidneys in children. *AJR*. 1985;145:611–616.
5. Rosenbaum DM, Korngold E, Teele RL. Sonographic assessment of renal length in normal children. *AJR*. 1984;142:467–469.
6. Schlesinger AE, Hedlund GE, Pierson WP, Null DM. Normal standards for kidney length in premature infants: determination with US. *Radiology*. 1987;164:127–129.
7. Zerlin JM, Blane CE. Sonographic assessment of renal length in children: a reappraisal. *Pediatr Radiol*. 1994;24:101–106.
8. Tarkington MA, Fildes RD, Levin K, Zeissman H, Harkness B, Gibbons MD. High resolution single photon emission computed tomography (SPECT) ^{99m}Tc -dimercaptio-succinic acid renal imaging: a state of the art technique. *J Urol*. 1990;144:598–600.
9. Majd M, Rushton G, Chandra R, Andrich MP, Tardif CP, Rashti F. Technetium- ^{99m}Tc -DMSA renal cortical scintigraphy to detect experimental acute pyelonephritis in piglets: comparison of planar (pinhole) and SPECT imaging. *J Nucl Med*. 1996;37:1731–1734.
10. Armitage P, Berry G. *Statistical Methods in Medical Research*. Oxford, UK: Blackwell Science; 1994.
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
12. Bajc M, Wallin L. ^{99m}Tc -DMSA renal scintigraphy during kidney maturation. *Clin Nucl Med*. 1995;20:211–214.
13. Schlesinger AE, Hernandez RJ, Zerlin JM, Marks TI, Kelsch RC. Interobserver and intraobserver variation in sonographic renal length measurements in children. *AJR*. 1991;156:1029–1032.
14. Sargent MA, Wilson BPM. Observer variability in the sonographic measurement of renal length in childhood. *Clin Radiol*. 1992;46:344–347.
15. Carrico CWT, Zerlin JM. Sonographic measurement of renal length in children: does the position of the patient matter? *Pediatr Radiol*. 1996;26:553–555.
16. Wallin L, Thorne J, Bajc M. Kidney size estimation in piglets using dimercaptio-succinic acid (DMSA) scintigraphy. *Clin Physiol*. 1997;17:591–597.
17. Rossleigh MA, Farnsworth RH, Leighton DM, Yong JLC, Rose M, Christian CL. Technetium- ^{99m}Tc -dimercaptio-succinic acid scintigraphy studies of renal cortical scarring and renal length. *J Nucl Med*. 1998;39:1280–1285.
18. Ferrer FA, McKenna PH, Bauer MB, Miller SF. Accuracy of renal ultrasound for predicting actual renal size. *J Urol*. 1997;157:2278–2281.
19. Sisayan RM, Rossleigh MA, Mackey DWJ. Nomograms of renal length in children obtained from DMSA scintigraphy. *Clin Nucl Med*. 1993;18:970–973.
20. Lebowitz RL, Colodny AH. Urinary tract infection in children. *CRC Crit Rev Clin Radiol Nucl Med*. 1973;4:457–475.