How Accurate Is Dimercaptosuccinic Acid Scintigraphy for the Diagnosis of Acute Pyelonephritis? A Meta-Analysis of Experimental Studies

Jonathan C. Craig, Danielle M. Wheeler, Les Irwig, and Robert B. Howman-Giles

Center for Kidney Research and Department of Nuclear Medicine, The New Children's Hospital, Sydney; Department of Public Health and Community Medicine, and Department of Paediatrics and Child Health, University of Sydney, Sydney, Australia

The purpose of this study was to evaluate the performance of dimercaptosuccinic acid (DMSA) scintigraphy in the diagnosis of acute pyelonephritis and to compare the test performance of the standard technique, planar DMSA, with the newly introduced technique, SPECT DMSA. **Methods:** All published animal studies in which DMSA scintigraphy was compared with histopathology, the reference standard for acute pyelonephritis, were identified using a comprehensive search strategy with the MEDLINE and EMBASE databases. Test performances of all DMSA methods and SPECT versus planar DMSA were analyzed using summary receiver operating characteristic (sROC) curves. **Results:** Seven studies were identified, including 2 of SPECT DMSA. Problems in study design or reporting were common, with numerical errors in 4 studies. Overall, at a sensitivity of 86%, specificity was estimated to be 91%. Detection of acute pyelonephritis was at a lower threshold for SPECT than for planar DMSA (sensitivity/specificity values of 97%/66% compared with 82%/97%), and the overall test performance of SPECT was not demonstrably better than that of planar DMSA. When applied to a group of children with a prevalence of renal damage of 40%, this means that 98% of children with abnormal planar DMSA scans will have renal damage, whereas only 65% of those with abnormal SPECT scans will have renal damage. Planar and SPECT DMSA will miss 11% and 3% of children with renal damage, respectively. Out of 100 children in the hypothetical group with 40% experiencing renal damage, SPECT will identify 6 extra true cases of renal damage at the expense of 19 extra false positives, when compared with planar DMSA. **Conclusion:** Published studies of DMSA test performance are few in number and have significant methodologic problems that should be avoided in future studies. DMSA, particularly the planar technique, performs well for the diagnosis of acute pyelonephritis. Using test performance criteria, SPECT DMSA alone has not been shown to be preferable to the established planar method and will result in a small number of true-positives at the expense of a larger number of false-positives.

Key Words: dimercaptosuccinic acid; pyelonephritis; diagnosis; systematic review; sensitivity

J Nucl Med 2000; 41:986–993

Received Jan. 19, 1999; revision accepted Sep. 9, 1999.

For correspondence or reprints contact: Jonathan C. Craig, PhD, Center for Kidney Research, The New Children's Hospital, PO Box 3515, Parramatta, NSW 2124, Australia.

Renal tract imaging using ^{99m}Tc-dimercaptosuccinic acid (DMSA) scintigraphy has become a widely recommended and used diagnostic test, largely replacing intravenous urography as the preferred test for identifying children who have sustained damage to the renal parenchyma because of urinary tract infection (UTI) (*1,2*). Two questions must be answered to interpret DMSA scans appropriately. First, what is the performance of this test for the detection of renal damage? In particular, how often will DMSA scintigraphy miss renal damage when present or show a cortical defect when renal damage is not present? Second, is the test performance of DMSA affected by patient and technical factors, and, in particular, are there clinically important differences in the DMSA result depending on whether planar or SPECT is used? SPECT is increasingly used and is advocated as the preferred method for DMSA scintigraphy (*3,4*). Can this be justified by improved test performance relative to the established planar method?

Methods for evaluating diagnostic tests by systematic reviews (and by meta-analysis when appropriate) have been formulated (*5,6*). A systematic review uses explicit methods to identify, select, critically appraise, and summarize research that is relevant to a clearly formulated question. Primary data may also be pooled (meta-analysis). Meta-analysis of diagnostic tests can be used for 2 purposes. First, it can provide an overall summary of diagnostic test performance. Second, meta-analysis can be used to determine whether observed differences of test performance among the primary studies are explained by methodologically or clinically important factors, such as the quality of the primary studies and characteristics of the patient and test.

The purpose of this study was to obtain an estimate of the overall test performance of DMSA scintigraphy for the diagnosis of acute pyelonephritis, using meta-analytical techniques. We also explored whether the test performance of DMSA scintigraphy was affected by characteristics of the study population or the way in which the test was performed. Our primary comparison was between SPECT and planar

images. Although several reviews of cortical scintigraphy of UTI in children have been published, the methods used to find and appraise all studies of DMSA for the diagnosis of acute pyelonephritis have not been explicitly given or systematically applied, a summary estimate of overall accuracy has not been calculated, and the effects of patient and test factors of accuracy have not been fully explored (*7,8*). This study is designed to provide clinicians with a systematic review of the test performance of DMSA for the diagnosis of acute pyelonephritis, in a form that is readily usable at the bedside.

MATERIALS AND METHODS

Inclusion and Exclusion Criteria

It is not ethical to compare DMSA scan results in humans with histopathology, which is the reference standard for acute pyelonephritis, because this would require nephrectomy for histological examination. We therefore included any study in which there was a comparison of DMSA scan results and histopathology in animals. To identify eligible studies, we performed a literature search using the MEDLINE (National Institutes of Health, Bethesda, MD; 1966–March 1998) and EMBASE (Elsevier Science, New York, NY; 1988–March 1998) databases, with the following search strategies in OVID: MEDLINE, pyelonephritis in major MESH (medical subject heading) heading (exploded) or pyelonephritis as a textword, and succimer in major MESH heading (exploded) or dimercaptosuccinic acid as a textword or DMSA as a textword; EMBASE, pyelonephritis (exploded) or acute pyelonephritis (exploded) or pyelonephritis as a textword, and succimer in major MESH heading (exploded) or dimercaptosuccinic acid as a textword or DMSA as a textword. All abstracts (total, n = 190; overlapping, n = 31) were reviewed online by 1 author, and 9 full articles were retrieved and analyzed only if the abstracts suggested that there was a comparison of histopathology and DMSA scintigraphy results, or if abstracts were unavailable or ambiguous. Six original articles were identified through this search strategy. Searches using other terms, including combinations of sensitivity, specificity (using exploded major MESH headings and textwords truncated with the wildcard $), and pyelonephritis (with and without subheadings of radionuclide imaging and diagnosis), did not identify any additional studies. Manual searching through reference lists and the collected reprints of content experts yielded 1 additional article. A manual search through proceedings of nephrology, nuclear medicine, and pediatric conferences yielded 1 additional abstract suitable for inclusion.

Studies were excluded if (a) no definition of DMSA or histopathology criteria for acute pyelonephritis was given or shown; (b) there was a stated period of more than 1 wk between DMSA scintigraphy and histopathology; (c) sensitivity and specificity could not be calculated directly from the data given; or (d) the results of a study were published more than once. If study duplication occurred because of publication in abstract and full paper form, the results given in the full paper were used. Using all 4 criteria, only 1 of the 8 eligible studies was excluded, because of abstract and paper duplication.

Data Extraction and Critical Appraisal

All 7 eligible studies were analyzed, and the data on study and test characteristics and results were extracted by 2 independent reviewers. Disagreement was resolved by consensus, with resolution of outstanding differences by a third reviewer. Readers were not excluded from details of authorship. Each study was also critically appraised using a checklist of potential factors considered to bias the true estimate of test performance (*6*).

Statistical Analysis

The overall test performance of DMSA for the detection of acute pyelonephritis was analyzed using a summary receiver operating characteristic (sROC) curve, plotting sensitivity (on the y-axis) against 1-specificity (on the x-axis) using data points from each primary study (*5,6*). Values for sensitivity (true-positive rate), specificity (true-negative rate), and prevalence were calculated directly from the raw data rather than using the values calculated by the study authors. An sROC curve graphically represents the trade-off between sensitivity and specificity. As sensitivity increases, with fewer affected cases missed (a reduction in false-negatives), the specificity decreases, so that more subjects are classified wrongly as affected when they in fact are unaffected (an increase in false-positives). The sROC curve is estimated from the sensitivity and specificity data of the primary studies by regression methods applied to the logistic transformation of the sROC axes. This measure involves predicting the log-odds ratio by test threshold. If the coefficient for test threshold is near 0 and not significant, then the odds ratio is constant at all thresholds. The regression equation and the sROC plot were obtained after adding 0.5 to the numerator and 1.0 to the denominator of both the true-positive and false-positive rates for each study, so that any 0 cells did not result in undefined transformations. The sROC curve was constructed with individual study points of equal weighting and weighting by the inverse of the variance. Weighting individual studies by variance or sample size before incorporation into the overall model made no appreciable difference to the result, so only unweighted results are included here.

Because of considerable variation in the methods by which portions (areas or zones) of the kidney were assigned in the individual studies, the unit of analysis for this systematic review was complete kidneys (renal units). Within each animal, we assumed that the DMSA results for each kidney were independent.

Factors considered to affect the test performance of DMSA, such as variations in study populations (piglet or rat), DMSA methods (SPECT or planar), and study design (time from infection to scanning), were analyzed using an unpaired *t* test for significance testing with the odds ratio as the summary estimate of test performance (odds ratio = odds of test positivity in the diseased kidneys divided by the odds of test positivity in nondiseased kidneys). This was appropriate, because the odds ratio was found to be constant across thresholds. The preferred method of analysis, incorporating these factors individually into the regression analysis using the sROC method, was not feasible, because the number of studies was too small to obtain stable estimates. True-positive rates and false-positive rates from each study, not adjusted for potential predictors of test performance, were therefore included in the sROC model. The sROC curve was displayed graphically, with studies sharing a characteristic resulting in a lower test performance tending to appear below the sROC line and studies sharing a characteristic resulting in better DMSA test performance tending to appear above the sROC line.

RESULTS

Critical Appraisal

The results of our critical appraisal are given in Tables 1–3. Of the 7 studies (*9–15*), 6 used a refluxing pig model (*9–11,13–15*), and 1 used a rat model of acute pyelonephritis (*12*). Generally, the studies provided good detail on the cortical scintigraphy methods and the criteria used to define acute pyelonephritis histologically and for DMSA. For all studies, it was possible to extract the data directly from the tables provided to calculate sensitivity and specificity.

Of the 7 studies, 3 contained mathematical errors (*10,12,13*). In 1 study sensitivity was calculated using the formula for specificity and vice versa (*10*), and in 2 studies the calculated sensitivity, specificity, or both, were incorrect (*12,13*). Accordingly, in 3 abstracts erroneous values for sensitivity, specificity, or both, were given. In a different paper, specificity was not calculated for the reader; only values for sensitivity and "accuracy" were given instead (*15*). Accuracy (the sum of true-positives and true-negatives divided by the total number of subjects) is prevalence dependent and not a good measure of test performance (*16*). In only 2 of 7 studies was it specifically stated that the observers reporting the DMSA scan result were not aware of the histopathology result and the histopathologist was not aware of the DMSA result (*9,15*). In 3 studies, animals were selected for testing by histopathology, based on the DMSA result, which would potentially result in a biased (better) estimate of sensitivity (*10,11,13*). An example of this potential selection bias was the exclusion of animals from further analysis when DMSA scintigraphy was abnormal, before acute pyelonephritis was induced.

Measuring Overall Test Performance

From the sROC plotted by using the sensitivity and specificity from all the primary studies (Fig. 1), sensitivity and specificity at various points of the curve can be read. The average sensitivity of all tests over all studies was 86%, corresponding to a specificity of 91% on the sROC (area under the curve = 0.96).

The effect of different study factors on DMSA test performance are given in Table 4. No factor was found to significantly influence test performance. The power to detect

TABLE 1
^{99m}Tc-DMSA Scintigraphy Methods

Method	Rushton et al. (*9*) (1988)	Parkhouse et al. (*10*) (1989)	Arnold et al. (*11*) (1990)	Wikstad et al. (*12*) (1990)	Giblin et al. (*13*) (1993)	Risdon et al. (*14*) (1994)	Majd et al. (*15*) (1995, planar)	Majd et al. (*15*) (1995, SPECT)
Definition of pyelonephritis	Photon deficient area	Photon deficient area	NS (examples given)	NS (examples given)	Cortical thinning/cortical mottling/ focal cortical defect	Photon deficient area	Focal or diffuse photon deficient area	Focal or diffuse photon deficient area
Planar/ SPECT	Planar	Planar	Planar	Planar	SPECT	Planar	Planar	SPECT
Views	Posterior and posterior-oblique	Posterior and posterior-oblique	Anterior and posterior	Anterior	NA	Posterior and posterior-oblique	NA	Posterior and posterior-oblique
Dosimetry	7.4 MBq/kg	3–4 MBq/kg	11 MBq/kg	60–120 MBq/kg	NS	3–4 MBq/kg	3.7 MBq/kg	3.7 MBq/kg
Counts	NS	250×10^3 (posterior) 250×10^3 (oblique)	100×10^3	NS	$250–450 \times 10^3$ (posterior) 150×10^3 (oblique)	NS	NS	
Camera	Converging collimator	Multipurpose collimator	Multipurpose collimator	Pinhole collimator	Converging collimator	Multipurpose collimator	Pinhole collimator	Dual-head ultrahigh-resolution collimator
Matrix	128×128	NS	64×64	128×128	NS	NS	128×128	NS
Time span* (h)	5	6	16–20	5	NS	6	2–3	2–3
Time for scan (min)	NS	NS	NS	10	NS	NS	NS	NS

*Between injection and scan.
NS = not stated; NA = not applicable.

TABLE 2
Study Methods and Results

	Rushton et al. (*9*) (1988)	Parkhouse et al. (*10*) (1989)	Arnold et al. (*11*) (1990)	Wikstad et al. (*12*) (1990)	Giblin et al. (*13*) (1993)	Risdon et al. (*14*) (1994)	Majd et al. (*15*) (1995, planar)	Majd et al. (*15*) (1995, SPECT)
Animal	Pig	Pig	Pig	Rat	Pig	Pig	Pig	Pig
Sample size								
Animals	22	33	60	36	17	42	16	16
Kidneys	44	37	120	72	34	46	32	32
Prevalence of abnormality (%)	36	73	28	74	59	78	75	75
Time from UTI to DMSA	7 and 21 d	13 d (median)	7 and 21 d	5 and 21 d	Variable	Variable	1, 2, 3, and 10 d	1, 2, 3, and 10 d
Time between DMSA and pathology	0	NS	<48 h	NS	NS	NS	0	0
Method of UTI	Reflux + wax + *E. coli*	Reflux + wax + *E. coli*	Reflux + wax + *E. coli*	*E. coli*	Reflux + wax + *E. coli*	Reflux + wax + *E. coli*	Reflux + wax + *E. coli*	Reflux + wax + *E. coli*
Antibiotic treatment for some	No	No	No	Yes	No	No	No	No
Only animals with VUR included	Yes	Yes	No	No	No	Yes	Yes	Yes
True-positives	13	24	29	32	20	34	20	22
False-positives	1	0	3	0	6	0	0	2
True-negatives	27	10	83	19	8	10	8	6
False-negatives	3	3	5	21	0	2	4	2
Sensitivity	81	89	85	60	100*	94	83	92*
Specificity	96	100	97	100	57*	100	100	75*
Odds ratio	71	147	128	59	54*	290	77	23*

*SPECT = 1st and 3rd highest sensitivity, 1st and 3rd lowest specificity, and 2 lowest odds ratios.
VUR = vesicoureteral reflux; NS = not stated.

TABLE 3
Study Quality

Study design feature	Number of studies/ total studies (references)
Reference standard used	
Macroscopic appearance	1/7 (*11*)
Histopathology	6/7 (*9, 10, 12–15*)
Stated blinding of the person interpreting DMSA to the result of pathologic examination and vice versa	2/7 (*9, 15*)
Subjects stated to be included on a consecutive basis	3/7 (*9, 12, 14*)
SPECT and planar images interpreted independently	1/2 (*15*)
Numerically correct	4/7 (*9, 11, 14, 15*)

TABLE 4
Factors Associated with Increased Test Performance

Predictors	*P* for difference in mean odds ratio
Study design	
Blinding stated	0.9
Numerically correct	0.6
Subjects not included on a consecutive basis	0.2
Reference standard was histopathology	Only 1 nonhistopathology study
Sample characteristics	
Pig studies	Only 1 nonpig study
Antibiotic treatment for some animals	Only 1 study
Time between UTI and DMSA	Details not sufficient for analysis
Only animals with vesicoureteric reflux included	0.5
Stated time between DMSA and pathology	0.4
Sample size	0.8
Prevalence of abnormal kidneys	0.5
DMSA methods	
Planar DMSA	0.07
Posterior and posterior oblique views (planar images only)	0.1
Dosimetry	0.5
Counts	0.1
Multipurpose collimators	0.6
Matrix	0.3
2–6 h between injection and scan	0.9
Time for scan	Only given in 1 study

clinically important differences of study design, sample characteristics, and DMSA methods on DMSA test performance was limited by the small sample size and the extent to which information was given in the individual studies. A sensitivity analysis, excluding the rat study, showed no appreciable effect on overall test performance (sensitivity 87%, specificity 90%, area under the curve = 0.96), and so the rat study was included in the analysis.

Comparing Test Performance of SPECT and Planar DMSA

Analysis of the sROC curve (Fig. 1) demonstrated that there was no significant difference in the test performance of planar and SPECT DMSA scintigraphy for the diagnosis of acute pyelonephritis, with the results of 1 SPECT DMSA study plotted above the line and the other below the line. The odds ratios were lower in the 2 SPECT DMSA studies than in any planar DMSA studies (Table 2), but this difference did not reach statistical significance (Table 4).

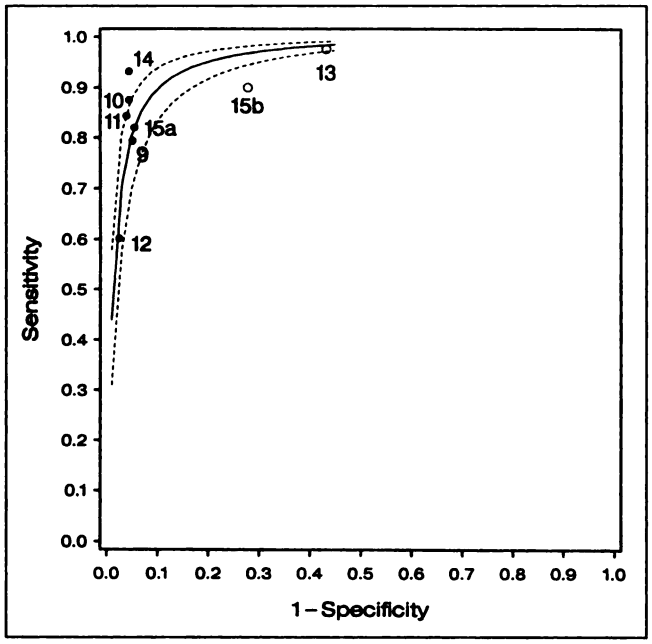

FIGURE 1. Overall summary receiver operating characteristic curve (sROC) for all studies. Numbers next to plots represent reference numbers for studies. Planar studies are represented by dots. Asterisks represent SPECT studies.

A threshold effect is evident from the sROC curve, with the 2 SPECT studies plotted at a region of higher sensitivity but lower specificity than the 6 planar studies. The average sensitivities of SPECT and planar DMSA were 97% (95% confidence limits; range, 62%–100%) and 82% (range, 70%–95%), respectively. The average specificities of SPECT and planar DMSA were 66% (range, 0%–100%) and 99% (range, 97%–100%), respectively.

In 2 studies, planar and SPECT were performed on the same animals (*13,15*). This comparison is preferable, because any between-study differences were controlled for. However, results were only given for 1 study in which sensitivity/specificity combinations for planar were 83%/100% and for SPECT DMSA 92%/75% (*15*). A McNemar's test of agreement between the 2 imaging methods was not possible, because results in the same animals were not cross classified.

DISCUSSION

Critical Appraisal

Critical appraisal showed several study design and reporting problems. Generally, the methods by which histopathology and DMSA scintigraphy were performed were adequately reported, but important factors that may have biased the true test performance of DMSA were not considered. Whether this represents faulty design or incomplete reporting is not clear, but future investigators should be aware of methodological issues in the appropriate assessment of new diagnostic tests. For example, interpretation of the new diagnostic test and the reference standard should be independent, and subjects should not be tested by the reference standard depending on the new test result, but consecutively (*6*).

About half of the studies were numerically incorrect and reported erroneously high values for sensitivity and specificity in the abstract. This is particularly misleading for clinicians who read only the abstract and do not have the time to verify the results personally. This finding has implications for future investigators and also for reviewers and journal editors. Compared with randomized controlled trials, for which strict criteria are used in major journals (the Consolidated Standards of Reporting Trials [CONSORT] statement (*17*)), the reporting of diagnostic tests is heterogeneous. A standardized method for diagnostic test assessment and reporting may improve the quality of future studies and has been proposed (*18*).

The strongest study design to measure agreement between 2 tests is a paired comparison whereby the same subjects are tested with both methods. This approach accounts for any between-study differences that inevitably occur, and McNemar's test, the χ^2 equivalent of the paired t test for a binomial distribution, would then be the most powerful significance test (*19*). In 2 studies this approach would have been possible but was not taken.

Animals and Human Subjects

What relevance do animal studies have to clinical decision making? An optimal study design would be the comparison of DMSA scintigraphy with histopathology of kidneys from children. Because of the localized nature of pyelonephritis, such a study is not feasible, and animal studies are the next-best study design. The pig model of acute pyelonephritis has been extensively used for the past 30 y because of anatomical detail similar to that in humans. The assumption that the test performance of DMSA in pigs is similar to humans seems reasonable, in view of anatomical similarity and the similarity of the observed DMSA defects in children and pigs. The clinical applicability of the rat model (used once) is perhaps more questionable. Analysis suggested that there was a threshold effect, with the rat model having low sensitivity/high specificity rather than having an overall effect on test performance (Fig. 1). Sensitivity analysis did not show any appreciable effect on the area under the sROC curve. Accordingly, the rat study was not excluded.

Overall Test Performance

Overall, DMSA scintigraphy performed very well in the detection of abnormal renal cortex after UTI. At a mean sensitivity of 84%, an overall specificity of 88% was found. Although there was a wide range of test performance reported, with the sROC ranging from a sensitivity/specificity combination of 60%/100%–100%/57%, most studies clustered in a range of sensitivity values <80% with specificity <95%.

It could be suggested that a sensitivity of around 80% is too low, resulting in too many missed cases. This statement assumes that the value of detecting extra true cases exceeds the value of false-positive cases, which is uncertain. The benefit of intervention in extra true cases must be weighed against the harms of anxiety, labeling, extra testing, and extra intervention in the false-positives. Criteria for judging an appropriate threshold for case definition should include known prognostic value (what is the outcome for patients with the diagnosis of interest using the defined threshold?), or improved patient-centered outcomes from interventions given to those with the diagnosis of interest, or both. In the context of UTI and specifically pyelonephritis, we would suggest that a higher threshold for case identification is appropriate. For example, acute pyelonephritis is common after UTI (about 40% using DMSA scintigraphy and probably about 50% if histopathology was done [*20*]), but hypertension (<1%) and chronic renal failure (<0.001%) are rare (*21*). Similarly, the diagnosis of acute pyelonephritis does not confer any advantage to the patient, because no proven interventions are routinely made once the diagnosis is made.

Because of the small number of studies and the ways in which results were reported, we could not explore the effect of disease spectrum on test performance. Disease spectrum is a term that describes such features as disease duration and severity that may influence test performance (*22*). Many authors reported that DMSA mainly missed the small areas of pyelonephritis. Certainly the sensitivity of many other diagnostic tests is known to increase as the spectrum of disease worsens. Accordingly, the sensitivity of DMSA is likely to be >80% for those children with more severe pyelonephritis, and these are the children who clinicians would be most concerned about identifying.

Like all meta-analyses of published research, this study is subject to publication bias, and so may be an overestimate of the true test performance of DMSA. With the small number of studies, graphical or numerical estimates of publication bias, such as funnel plots, were not possible.

Planar Versus SPECT DMSA

Two important comparisons were noted between the results of planar and SPECT DMSA scintigraphy. First, the threshold for detecting acute pyelonephritis was lower for

SPECT than for planar DMSA. Second, the overall test performance of SPECT was not demonstrably better than planar DMSA. SPECT DMSA has higher true-positive and false-positive rates than planar DMSA. That is, SPECT DMSA detects more areas of pyelonephritis at the expense of an increased number of abnormal test results that do not represent areas of pyelonephritis. The effects of these differences are shown in Tables 5 and 6 as applied to published prevalences of abnormal renal parenchyma in children after UTI. Sensitivity and specificity generally remain constant when applied to populations with different prevalence of disease, but predictive values vary widely (*16*). For example, assuming a prevalence of acute pyelonephritis of 40% in children presenting with UTI (*20*), 98% of those with an abnormal planar DMSA scan will have abnormal kidneys (2% will not) and 65% of those with an abnormal SPECT scan will have abnormal kidneys (35% will not). In the same group of children, 11% of those with a normal planar DMSA will have acute pyelonephritis and 3% of those with a normal SPECT DMSA will have pyelonephritis (Table 5). In other words, for every 100 children with acute UTI, compared with planar DMSA, SPECT DMSA will detect 6 extra true cases at the expense of classifying 19 extra children with disease when they have normal kidneys. When applied to children with a 5% risk of renal damage (*23*), for every extra child correctly identified with renal parenchymal abnormality, 27 extra children will be incorrectly diagnosed as diseased.

Second, in addition to the threshold effect, the overall test performance of SPECT DMSA was not superior to planar DMSA and may even have been worse. Because of the small number of studies, particularly the small number of SPECT DMSA studies, any conclusion on which test showed better performance is uncertain. There are few data on which SPECT DMSA test performance can be evaluated (2 studies with a total animal number of 33), and these data do not provide a firm justification for SPECT becoming the standard method of DMSA scintigraphy in children.

TABLE 5
Predictive Values of Planar and SPECT DMSA Scintigraphy with Sensitivity/Specificity Combinations of 82%/99% and 97%/66%, Respectively, in Different Prevalences of Kidney Damage

Prevalence of disease (%)	Probability of kidney damage given abnormal DMSA scan (%)		Probability of kidney damage given normal DMSA scan (%)	
	Planar	SPECT	Planar	SPECT
5*	80	13	0.01	0.002
40†	98	65	11	3
80‡	99	92	42	15

*Prevalence of persistent renal damage following UTI (*23*).
†Prevalence of renal damage at the time of UTI (*21*).
‡Prevalence of renal damage in children with febrile UTI and at the time of infection (*24*).

TABLE 6
Effects of Lower Threshold for Detecting Renal Parenchymal Abnormality for SPECT DMSA and Planar DMSA Scintigraphy Expressed as Extra Number of True-Positives and False-Positives Detected by SPECT in 100 Children at Varying Prevalences of Renal Parenchymal Abnormality

Prevalence of disease (%)	Extra true-positives detected by SPECT compared with planar DMSA scintigraphy	Extra false-positives detected by SPECT compared with planar DMSA scintigraphy
5*	1	27
40†	6	19
80‡	12	6

*Prevalence of persistent renal damage following UTI (*23*).
†Prevalence of renal damage at the time of UTI (*21*).
‡Prevalence of renal damage in children with febrile UTI and at the time of infection (*24*), based on sensitivity/specificity combinations of 82%/97% for planar and 97%/66% for SPECT DMSA scintigraphy.

Apart from test performance, does SPECT have other advantages over planar DMSA that warrant this method becoming routinely used in the care of children with UTI? SPECT is in fact more expensive ($391 versus $285 in Australian currency) and requires children to be restrained for at least twice as long to obtain the required image clarity (25 min versus 12 min). As discussed above, in our view the better test is the one that confers more benefit than harm. The provision of accurate prognostic information is a benefit, but currently no information is available on long-term clinical outcomes for children who have been diagnosed using any form of cortical scintigraphy. Substantially more information, however, is available on short-term imaging and clinical outcomes for children diagnosed using abnormal planar DMSA scintigraphy than is available for SPECT DMSA (*7,8,20,24*). This will probably change as more SPECT is performed.

Unfortunately, we could not compare the incremental value of SPECT over planar DMSA alone, even though this method is in widespread clinical use, because published studies have not addressed the test performance of this combination.

Why then is SPECT DMSA being accepted as the preferred technique for cortical scintigraphy in children, when there are no clear advantages over the old test? This is not a unique occurrence. The introduction of new health technologies is common even before they are proven to be safe, effective, or better than the old method. The reasons for this have been explored in detail elsewhere (*25*). It should be the responsibility of health care providers and individual clinicians to appraise new health care technologies critically before introducing them in the clinical setting.

CONCLUSION

Published studies of DMSA test performance are few and have significant methodological problems that should be

avoided in future studies. In animal studies, we can conclude that DMSA, particularly the planar technique, performs well for the diagnosis of acute pyelonephritis. Using test performance criteria, SPECT DMSA alone has not been shown to be preferable to the established planar method and may be worse. Assuming these results are transferable to humans, SPECT DMSA has a lower threshold for pyelonephritis detection, resulting in more true-positives and false-positives, which is probably disadvantageous in the diagnosis of UTI in children.

ACKNOWLEDGMENTS

The authors thank Elisabeth Hodson and Paul Roy for reviewing earlier drafts and Petra Macaskill for assistance with generating the sROC curve. This study was supported by a grant from the Children's Hospital Fund, Sydney, Australia.

REFERENCES

1. Lavocat MP, Granjon D, Allard D, Gay C, Freycon MT, Dubois F. Imaging of pyelonephritis. *Pediatr Radiol.* 1997;27:159–165.
2. Goldraich NP, Ramos OL, Goldraich IH. Urography versus DMSA scan in children with vesicoureteric reflux. *Pediatr Nephrol.* 1989;3:1–5.
3. O'Connor JP. DMSA scanning: a pediatric urologist's point of view. *Pediatr Radiol.* 1995;25:S50–S51.
4. Yen TC, Chen WP, Chang SL, Liu RS, Yeh SH, Lin CY. Technetium-99m-DMSA renal SPECT in diagnosing and monitoring pediatric acute pyelonephritis. *J Nucl Med.* 1996;37:1349–1353.
5. Moses LE, Shapiro D. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293–1316.
6. Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med.* 1994;120:667–676.
7. Mackenzie JR. A review of renal scarring in children. *Nucl Med Commun.* 1996;17:176–190.
8. 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.
9. Rushton HG, Majd M, Chandra R, Yim D. Evaluation of ^{99m}technetium-dimercaptosuccinic acid renal scans in experimental acute pyelonephritis in piglets. *J Urol.* 1988;140:1169–1174.
10. Parkhouse HF, Godley ML, Cooper J, Risdon RA. Renal imaging with ^{99m}Tc-labelled DMSA in the detection of acute pyelonephritis: an experimental study in the pig. *Nucl Med Commun.* 1989;10:63–70.
11. Arnold AJ, Brownless SM, Carty HM, Rickwood AMK. Detection of renal scarring by DMSA scanning: an experimental study. *J Pediatr Surg.* 1990;25:391–393.
12. Wikstad I, Hannerz L, Karlsson A, Eklof A-C, Olling S, Aperia A. ^{99m}Technetium dimercaptosuccinic acid scintigraphy in the diagnosis of acute pyelonephritis in rats. *Pediatr Nephrol.* 1990;4:331–334.
13. Giblin JG, O'Connor KP, Fildes RD, et al. The diagnosis of acute pyelonephritis in the piglet using single photon emission computerized tomography dimercaptosuccinic acid scintigraphy: a pathological correlation. *J Urol.* 1993;150:759–762.
14. Risdon A, Godley ML, Parkhouse HF, Gordon I, Ransley PG. Renal pathology and the ^{99m}Tc-DMSA image during the evolution of the early pyelonephritic scar: an experimental study. *J Urol.* 1994;151:767–773.
15. Majd M, Rushton HG, Chandra R, Andrich MP, Tardif CP, Rashti F. Technetium-99m-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.
16. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology, a Basic Science for Clinical Medicine.* 2nd ed. Boston, MA: Little, Brown; 1991.
17. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276:637–639.
18. Bruns DE, Huth EJ, Majid E, Young DS. Toward a checklist of studies of diagnostic accuracy. *Clin Chem.* 1997;43:2211.
19. Armitage P, Berry G. *Statistical Methods in Medical Research.* 3rd ed. Oxford, UK: Blackwell Scientific Publications; 1994.
20. Stokland E, Hellstrom M, Jacobsson B, Jodal U, Lundgren P, Sixt R. Early ^{99m}Tc dimercaptosuccinic acid (DMSA) scintigraphy in symptomatic first-time urinary tract infection. *Acta Paediatrica.* 1996;85:430–436.
21. Kramer MS, Tange SM, Drummond KN, Mills EL. Urine testing in young febrile children: a risk-benefit analysis. *J Pediatr.* 1994;125:6–13.
22. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA.* 1995;274:645–651.
23. Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their first referral for urinary tract infection. *BMJ.* 1997;315:918–919.
24. Escobar-Billing R, Fituri O, Karlsson A, Linne T, Wikstad I, Aperia A. ^{99m}Tc-DMSA scintigraphy as diagnostic tool in acute pyelonephritis. *Pediatr Nephrol.* 1990;4:44–64.
25. Grimes DA. Technology follies. The uncritical acceptance of medical innovation. *JAMA.* 1993;269:3030–3033.