

**A critical appraisal of the evaluation of children with
urinary tract infection (UTI) by diagnostic imaging**

A report for the Consultative Committee on Diagnostic Imaging Research Program
(CCDIRP)

June 3rd 2002

Investigators

Robert Howman-Giles MBBS DDU MD FRACP¹

Jonathan C. Craig MBChB DCH MM(Clin Epi) FRACP PhD^{2,3}

Les Irwig MBBCh FFPHM PhD^{3,4}

Danielle Wheeler BSc²

Gabrielle Williams BSc PhD²

¹Department of Nuclear Medicine, The Children's Hospital at Westmead, Sydney, Australia

²Centre for Kidney Research, The Children's Hospital, at Westmead, Sydney, Australia

³Department of Public Health and Community Medicine, University of Sydney, Australia

⁴Department of Paediatrics and Child Health, University of Sydney, Sydney, Australia

Abstract:

Objectives:

1. To evaluate the test performances of commonly used renal tract imaging tests for the evaluation of children following urinary tract infection for the diagnosis of important clinical and imaging outcomes, such as vesicoureteric reflux and renal parenchymal abnormality .
2. To determine whether renal tract imaging reduces the frequency of adverse clinical outcomes in children following urinary tract infection.
3. To do an economic evaluation of the alternative diagnostic imaging strategies.
4. To identify unanswered research questions about the test performance and/or clinical significance of renal tract imaging.

Methodology: All published studies comparing renal tract imaging results of one test with another in children who have had at least one positive urine culture were identified from a comprehensive MEDLINE search strategy. Test performances of each method were analyzed using summary receiver operating characteristic curves and relative diagnostic odds ratios.

Results: Of 1013 titles screened, 74 publications were eligible and included, 37 examined the outcome vesicoureteric reflux, 18 kidney damage, 1 obstruction and 18 looked at both vesicoureteric reflux and kidney damage. There was significant heterogeneity between the majority of studies thus point estimates for sensitivity and specificity were not used. Quality assessment analysis showed that the ten quality items examined had no significant affect on the diagnostic odds ratios.

Conclusions: There is considerable published data comparing the various renal tract imaging tests. The studies are heterogeneous and of variable quality. None of the ten quality items assessed had a significant affect on the study outcome. Detection of vesicoureteric reflux and renal parenchymal abnormality by each of the major renal

tract imaging tests is summarised in the two following tables.

Detected VUR					Missed VUR		
Preval	Cases	US	DMSA	MCU*	US	DMSA	MCU
1%	10	3.6	8.3	10	6.4	1.7	0
10%	100	36	83	100	64	17	0
25%	250	90	207.5	250	160	42.5	0

*assumes MCU is a perfect reference standard

Detected Renal parenchymal abnormality					Missed renal parenchymal abnormality		
Preval	Cases	US	MCU	DMSA*	US	MCU	DMSA
1%	10	5.5	3.5	10	4.5	6.5	0
10%	100	55	35	100	45	65	0
40%	400	220	140	400	180	260	0

* assumes DMSA is a perfect reference standard

Ultrasound of the renal tract is often used as a screen for deciding to carry out an MCU or DMSA. If this practice were followed, assuming a 25% prevalence of vesicoureteric reflux, ultrasound would detect 90 of 250 cases and miss 160. For renal parenchymal abnormality, assuming a 40% prevalence, ultrasound would detect 220 of 400 cases and miss 180 cases.

Abbreviations:

UTI	-	urinary tract infection
DMSA	-	dimercaptosuccinic acid
MCU	-	micturating cystourethrogram
US	-	ultrasound
IVP/U	-	intravenous pyelogram/urogram
sROC	-	summary receiver operating characteristic curve
CRP	-	C reactive protein

Introduction:

Urinary tract infection is an important and common clinical problem in childhood. About 8% of girls and 2% of boys will develop at least one UTI by 11 years of age (Hellstrom *et al* 1991). Renal tract abnormalities in children presenting with UTI are common. A previous study by this Centre found 40% of a cohort with UTI had renal parenchymal abnormalities, 30% had vesicoureteric reflux and 1% obstructive uropathy (Craig *et al* 1998).

An ideal diagnostic test should; 1. accurately distinguish diseased from non-diseased individuals, 2. have little or no morbidity, 3. be affordable, 4. be easily accessed by the relevant population, and most importantly, 4. should be performed on the basis that the identification of the disease in question will result in better treatment and long-term outcomes. Renal tract imaging in children following UTI is performed on the assumption that the identification of children with renal tract abnormalities and the treatment of these abnormalities improves patient outcome by reducing the risk of urinary tract infection, kidney damage, hypertension and chronic renal failure. These assumptions have uncertain validity. The uncertainty surrounding the place of renal tract imaging in children with UTI is reflected by widely inconsistent recommendations by experts on the mode and timing of these tests (Downs, 1999).

The aim of this study was to obtain an estimate of the overall test performances of each renal tract imaging study in children with culture proven UTI using meta-analytical techniques. Tests considered include all those performed in this clinical setting – micturating cystourethrography (using direct and indirect methods, with both contrast and isotope methods), renal tract ultrasonography, renal scintigraphy,

intravenous urography, and diuretic renography. Outcomes considered include all clinical and imaging outcomes of interest, ESRD, chronic renal failure, hypertension, renal parenchymal abnormality, renal tract obstruction and vesicoureteric reflux. This study is designed to provide clinicians with a systematic review of the test performance of each test used in clinical practice for the diagnosis of urinary tract infection in children.

Methods:

Inclusion criteria:

Studies were included if the results of two or more renal tract imaging results were compared in the same child following at least one positive urine culture or if the results of one renal tract imaging test were cross tabulated with a clinically important outcome such as ESRD or hypertension. Studies were excluded if results from children and adults were unable to be separated, UTI was not confirmed with urine culture, sensitivity and specificity were not able to be calculated directly from the data given, or if the results of a study were published more than once. If studies were published twice results from the paper with the more detailed and complete methods and results was used.

Search strategy:

MEDLINE (1966-December 2000) was searched using kidney failure, urinary tract infection, vesico-ureteral reflux, obstruction, dialysis, end stage renal disease, and scarring as exploded medical subject headings and textwords terms, which were then combined with exploded medical subject headings and textwords for diagnosis, radiography, radionuclide imaging and ultrasonography (Appendix 1). The search was

age limited by excluding studies limited to adults only, and then limited to English language studies. All titles were reviewed online by one author (DW) and potentially relevant articles were identified. Abstracts of these were reviewed and articles that appeared relevant to this study were obtained for full paper assessment. Eligible papers underwent data extraction.

Data extraction and critical appraisal

Eligible studies were analysed and the data on study and test characteristics, and results were extracted by one of two independent reviewers (GW & DW). Twelve of 74 papers (16%) were chosen at random to undergo data extracted by the two reviewers to assess consistency. There was excellent agreement between the two reviewers, agreement on quality items 91.0% (95% CI 86.0-96.0), agreement on numerical items 100% (constant). Disagreement was resolved by discussion with a third reviewer (JC). Readers were not blinded to details of authorship. Each study was also critically appraised using a checklist of potential factors considered to bias the true estimate of test performance (Lijmer *et al* 1999). Authors used different units of measure, that is, children, kidneys, renal units and ureters. Data on kidneys, renal units and ureters were pooled using the term kidney, data for children were analysed separately.

Quality assessment and sources of heterogeneity

Plausible modifiers of test performance were considered under three headings: study quality (consecutive patient selection, prospective data collection, blinding of observers to the results of all other tests, avoidance of verification bias), patient characteristics (age, gender, sample size, prevalence of clinical outcome), and test

details (description provided, positive and/or negative test results defined). Each paper had the details of these items extracted and classified as follows. Spectrum was classified as clinical population, case control, or not stated. Verification bias was designated complete (100% of patients had both tests), partial (<100% had both tests), not stated, or a different reference test. Reference test blinding was classified as yes if it was carried out without knowledge of the comparative tests results, no if authors stated it was not done blind, or not stated if there was no information given. Comparative test blinding was classified similarly to above. The selection of patients was classified as consecutive, non consecutive or not stated. Data collection was designated prospective, retrospective or not stated if authors failed to state the collection method. Reference test details and comparative test details were designated adequate if they defined at least two of the following three variables, definition of a positive test result, definition of a negative test result, description of test methodology. Population details were designated adequate if they provided three of the following four items, age range, median or mean age, gender ratio, symptoms at presentation. Time between tests was classified as stated or not stated.

Statistical analysis

In cases where there were 5 or more papers comparing the same test and units, the overall test performance of each imaging examination was analyzed using forest plots of sensitivity and specificity, estimating the diagnostic odds ratios, and generating summary receiver operator characteristic (sROC) curves using the methods of Littenberg and Moses (1993) using weighted and unweighted data. sROC curves plot sensitivity (y-axis) against 1-specificity (x-axis) using data points from each primary study. Values for sensitivity (true positive rate), specificity (true negative rate), and

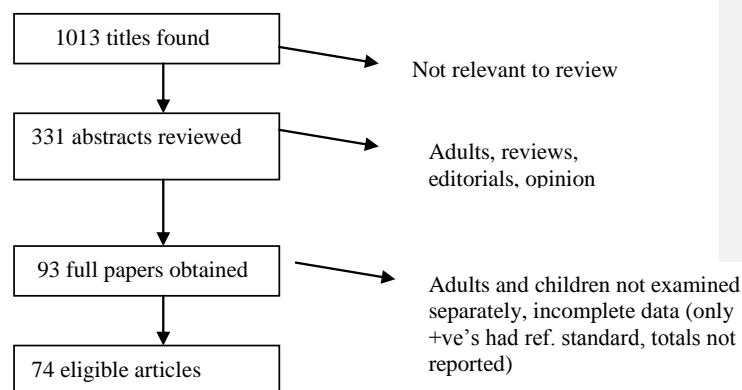
prevalence were calculated directly from the raw data rather than using the values calculated by the study authors. The regression equation and the sROC plot were obtained after adding 0.5 to all cells for each study so that any zero cells did not result in undefined transformations (Haldane 1955). Heterogeneity was explored by modelling candidate test, population and quality items as explanatory variables for the \log_e transformed diagnostic odds ratio (D) in the sROC equation. If threshold terms were close to zero and not significant, and diagnostic odds ratios were homogeneous, data were summarised using a pooled odds ratio using a random effects model. Study homogeneity was tested using the Q statistic generated from meta-analysis in a random effects model of the diagnostic odds ratio. If heterogeneous sROCs were simply plotted. Meta-test® and SAS® were used for all analyses.

Economic Analysis

Imaging tests were costed according to the Medicare Benefits schedule as at 1st November 2000 and cost effectiveness ratios were calculated (data on utilities were unavailable)

Results

Literature search;



No study was found which evaluated the association between a renal tract imaging test and a clinically important outcome (ESRD, chronic renal failure, hypertension).

37 papers examined VUR only, 1 obstruction and 18 kidney damage only. Eighteen papers examined both VUR and kidney abnormality. Comparisons with less than five publications are included in Tables 1-3 but were not analysed. Analysis of sensitivity and specificity data for differentiation between different grades of reflux was not possible as published information gave only agreement on presence or absence of VUR not agreement on grade of reflux.

SECTION I - Tests for vesicoureteric reflux (MCU as the reference

standard) ~~For each number of studies (children/kidneys), give refs to table of s-p, figures of s-p, sroc and table of number of tps, fns, etc at range of prevalences, summarise data by:~~

~~-S-p ranges~~

~~-LogDOR and DOR medians and iqr, means with 95%~~

~~-Q for logDOR~~

~~-Sources of heterogeneity found~~

~~-Comment about sroc~~

~~Comment about fps/fns etc~~

Intravenous pyelography

Eleven papers examined the accuracy of intravenous pyelogram for the detection of

Formatted: Bullets and Numbering

vesicoureteric reflux (table 3.1, figures 1.1 to 1.4). In seven studies children (n=1182) and in four studies kidneys (n=793) were the unit of study. For the studies using children as the measure, the shape of the sROC shows a trade off between the true positive rate (TPR) or sensitivity and false positive rate (FPR). As sensitivity increases and more true disease cases are identified there is an increase in false positive cases detected. There was significant heterogeneity between studies, $Q=45.80$, $P<0.001$. At the average FPR of 0.10 sensitivity or TPR is approximately 0.40. At a prevalence of 30% IVP would detect 120 of 300 cases of VUR and would falsely classify 70 children as having VUR when they did not (Table 4). For the four papers that used kidneys as the unit of study, the shape of the sROC shows that the true positive rate or sensitivity remains relatively constant while the FPR varies substantially. There was significant heterogeneity between studies, $Q=23.72$, $P<0.001$. At the average FPR of 0.22, sensitivity or TPR is approximately 0.49. At a prevalence of 30% IVP would detect 147 of 300 cases of VUR and would falsely classify 154 kidneys as having VUR when they did not (Table 4)

Renal tract ultrasonography

Fifteen studies evaluated the accuracy of ultrasound to detect vesicoureteric reflux (table 3.2 and figures 2.1-2.4). In seven studies, children (n=1239) and in eight studies, kidneys (n=2517) were the unit of study. For children as the unit of study, the shape of the sROC and the combined DOR (mean DOR 4.48) suggests that the accuracy of ultrasound is no better than chance for the detection of VUR. There was no significant heterogeneity between studies $Q=11.32$, 7df, $0.25<P>0.1$. At the average FPR of 0.35, sensitivity or TPR is approximately 0.34. At a prevalence of 30% US would detect 84 of 300 cases of VUR and would falsely classify 196 children

as having VUR when they did not (Table 4). For studies with kidneys as the unit of measure there was significant heterogeneity between studies, $Q=128.04$, 7df $P<0.001$. The sROC for these studies is better than for the studies using children as the unit of measure. The mean DOR is 127.20. At the average FPR of 0.10, sensitivity or TPR is approximately 0.34. At a prevalence of 30% US would detect 102 of 300 cases of VUR and would falsely classify 70 kidneys as having VUR when they did not (Table 4)

Dimercaptosuccinic acid scan

Thirteen papers examined DMSA for the detection of VUR (table3.3 and figures 3.1-3.4). In five studies, children (n=402) and in eight, kidneys (n=1146) were the unit of study. For children as the unit of study, the shape of the sROC was flat, sensitivity or TPR varied very little across studies while the false positive rate ranged from 0.156-0.917. There was significant heterogeneity between studies, $Q=11.07$, 4df, $P<0.05$. At the average FPR of 0.58, sensitivity or TPR is approximately 0.84. At a prevalence of 30% DMSA would detect 252 of 300 cases of VUR and would falsely classify 48 children as having VUR when they did not (Table 4). For papers using kidneys as the unit of study, the shape of the sROC curve shows that this test is only slightly better than chance for detecting VUR. There was significant heterogeneity between studies, $Q=12.03$, 7 df, $P<0.1$. At the average FPR of 0.25, sensitivity or TPR is approximately 0.51. At a prevalence of 30% DMSA would detect 153 of 300 cases of VUR and would falsely classify 147 kidneys as having VUR when they did not (Table 4).

Additional comparisons with VUR as the outcome

23 papers examined comparisons of novel forms of MCU. These were unable to be combined for analysis as too few studies examined the same combination of novel MCU using the same units or there was too little information to determine the exact technique utilised for the study. Results of sensitivity, specificity and diagnostic odds ratios for these studies can be found in Table 3.4. These included the use of indigo carmine dye as a contrast medium, indirect radionuclide cystourethrography, retrograde vs antegrade cystography, MCU with colloidal sulphur, colour flow doppler ultrasound, and cyclic MCU.

SECTION II - Tests for renal parenchymal abnormality

IVP compared to ultrasound

Seven papers examined IVP and ultrasound, six used children as the unit of measure (n=369) and one used kidneys as the unit of study (n=160). These papers were published between 1984 and 1989, before DMSA became the more frequently used and more sensitive test for kidney damage. In these comparisons IVP is treated as the reference standard for kidney abnormality, however it is an imperfect measure. There was no significant heterogeneity between studies, $Q=4.68$, 6 df, $0.5 < P < 0.75$. The range of values for sensitivity or TPR was large, 0.276-0.957. This was less marked for specificity, 0.688-0.992. The sROC curve for these studies shows a clear relationship between increasing sensitivity and a higher false positive rate. At the average FPR, 0.06, sensitivity or TPR is approximately 0.59. At a renal parenchymal abnormality prevalence of 10%, US detects 59 of 100 cases and wrongly classifies 54 of 900 children as having the condition when in truth they do not (Table 4).

Intravenous Pyelogram

Eight papers examined IVP compared to the reference standard DMSA, three used children as the unit of study (n=288) and five used kidneys as the study unit (n=977)(Table 3.8). Figure 5.1 shows the Forest plot of sensitivity and specificity for papers using kidneys as the unit of study. Sensitivity varied considerably, 0.269-0.855, but specificity remained relatively constant, especially in 4 papers (0.930-0.984, Farnsworth 1991, Goldraich 1989, Thelle 1985 and Merrick 1980), while the fifth paper, (Jakobsson 1992) showed a more divergent value, 0.746. There was significant heterogeneity between studies, $Q=52.73$, 4 df, $P < 0.001$. Figure 5.2 shows sROC curve for these studies. The four papers (Farnsworth 1991, Goldraich 1989,

Thelle 1985 and Merrick 1980) line up along the y axis and only the Jakobsson (1991) article is plotted a distance from the y axis. At the average FPR, 0.05, sensitivity or TPR is approximately 0.54. At a renal parenchymal abnormality prevalence of 10%, IVP detects 54 of 100 cases and wrongly classifies 45 of 900 kidneys as having the condition when in truth they do not (Table 4).

Analysis of the three papers using children as the unit of study was not carried out since we considered that three was too few studies.

Micturating Cystourethrogram

Thirteen papers examined MCU compared to DMSA (reference standard) for the detection of renal parenchymal abnormality, five used children as the unit of study (n=402) and eight used kidney as the unit of measure (n=1146). Table 3.9 displays the data. Figure 6.1 shows the Forest plot of sensitivity and specificity for the papers using children as the study unit. There was significant heterogeneity between studies, $Q=9.11$, 4df, $P<0.1$ with large ranges for both sensitivity and specificity. Figure 6.2 shows the sROC curve for these studies and highlights the low values for sensitivity across all studies. At the average FPR, 0.18 sensitivity or TPR is approximately 0.38. At a renal parenchymal abnormality prevalence of 10%, MCU detects 38 of 100 cases and wrongly classifies 162 of 900 children as having the condition when in truth they do not (Table 4). Figure 6.3 shows the Forest plot of sensitivity and specificity for papers using kidneys as the unit of study. Here too the range of values for sensitivity and specificity is very large (sensitivity: 0.3 – 0.82, specificity: 0.21 – 1.0). There was heterogeneity between studies, $Q=12.03$, 7df, $0.10<P<0.25$ but it was not statistically significant. Figure 6.4 displays the sROC curve for these studies and shows that MCU is not much better than chance for detecting RPA. At the average

FPR of 0.27, sensitivity or TPR is approximately 0.56. At a renal parenchymal abnormality prevalence of 10%, MCU detects 56 of 100 cases and wrongly classifies 253 of 900 children as having the condition when in truth they do not (Table 4).

Ultrasound

Thirteen papers compared ultrasound with the reference standard DMSA. Six papers used children as the unit of study (n=2469) and seven used kidneys (n=1452). Figure 7.1 shows the Forest plot of sensitivity and specificity for the six papers using children as the units, the range of values is enormous for sensitivity, 0.241-1.0 but less variable for specificity, 0.493-0.97. There was significant heterogeneity between studies, $Q=78.43$, 5df, $P<0.001$. Figure 7.2 shows the sROC curve for these studies and highlights the variation. Two papers results are plotted considerable distance from the curve. At the average FPR, 0.19 sensitivity or TPR is approximately 0.89. At a renal parenchymal abnormality prevalence of 10%, US detects 89 of 100 cases and wrongly classifies 171 of 900 children as having the condition when in truth they do not (Table 4). Figure 7.3 shows the Forest plot of sensitivity and specificity for papers using kidneys as the unit of study. In these studies, again there is considerable variation in sensitivity, 0.234-0.865, but less so for specificity, 0.625-1.00. There is significant heterogeneity between studies, $Q=95.63$, 6df, $P<0.001$. Figure 7.4 shows the sROC curve for these studies, again the curve is not a good reflection of the individual data points. At the average FPR, 0.07 sensitivity or TPR is approximately 0.67. At a renal parenchymal abnormality prevalence of 10%, US detects 67 of 100 cases and wrongly classifies 63 of 900 children as having the condition when in truth they do not (Table 4).

Additional comparisons with kidney abnormality as the outcome

Seven papers examined comparisons of different test for kidney abnormalities. These were unable to be combined for analysis since there were too few that examined the same tests. Results of sensitivity, specificity and diagnostic odds ratios for these studies can be found in table 3.11. Comparisons included variations on standard IVP, planar, spect and pinhole DMSA, Mag3, and CT.

SECTION III - Tests for kidney obstruction

Only one paper examining tests for obstruction was found. There is no established reference standard for identifying kidney obstruction, the single paper identified examined renography compared with diuretic renogram and Whitaker pressure perfusion.

Study Quality Assessment

Ten quality items were assessed for their affect on the best diagnostic odds ratio of each study. Regression analysis showed that none of the items had a significant affect on the diagnostic odds ratio

Economic Assessment

The statistical analysis shows that no one renal tract imaging test can reliably detect all abnormalities with a precision that would make it clinically appropriate or economically viable. An economic evaluation of these tests involves several assumptions, firstly that detecting the abnormality (true positive) leads to beneficial treatment that reduces costs, both economic and patient based, in the long-term. Secondly, that false positive results lead to further testing or inappropriate treatment which increases costs. Third, that true negative cases are not investigated further. Difficulties arise for this evaluation since our first assumption may not be true and also there is no information about the costs and outcomes of cases that were wrongly diagnosed as disease free (false negative). Analysis limited to comparing the false positive rate relative to the detection rate is displayed in Tables 6.1 (a-h) and summarised in Table 7. Clearly MCU is the best option for detecting VUR. Ultrasound is the cheapest option for detecting renal parenchymal abnormality, however the high false positive rate is high.

Discussion

This systematic review of the published literature identified 74 papers that examined at least one renal tract imaging test with another in children with a culture proven UTI. Papers were grouped into those with VUR as the outcome and those with

kidney abnormality as the outcome. Seven different comparisons were analysed as they included more than five studies that examined the same tests, many other comparisons were not analysed due to the small number of studies that examined the same tests and used the same units of measure, ie children or kidneys. Thus of 74 papers only 50 were able to be included in the meta-analysis. Of the seven analysed comparisons, the maximum number of studies was eight. A further problem arising in the study was the very large amount of variability across the different studies. Very few of the comparisons were homogeneous enough to use a summary point estimate for sensitivity and specificity.

The small number of studies that examined the same tests combined with the large heterogeneity made it impossible to give a precise/accurate estimate of how good each imaging study is in comparison with the reference standard.

Analysis of the available data shows that renal ultrasound detects approximately 36% of cases of vesicoureteric reflux and misses 64%. Thus each case of VUR detected by ultrasound costs \$277.50 which is significantly higher than that of MCU which detects all cases (\$130.30 per case detected). Dimercaptosuccinic acid scan has a reasonable detection rate for VUR (83%), however 17% of cases are missed and the cost per detected case of VUR is \$452.35, thus not the optimal balance. For detection of VUR, MCU is the best test clinically and economically.

Renal ultrasound detects 55% of cases of renal parenchymal abnormality and is clearly the cheapest option (\$181.64 per detected case) however it misses 45% of cases. Similarly MCU is a poor test for renal parenchymal abnormality, detecting only 35% of cases, thus missing 65%. Its costs is comparable to the DMSA scan for

each detected case (MCU \$372.29, DMSA \$375.45) but the number of missed cases make MCU a poor imaging option for the detection of renal parenchymal abnormality. DMSA scan is clinically the best option for detection of renal parenchymal abnormality and can be economically justified because of its superior sensitivity.

Table 1. Study characteristics of papers with vesicoureteric reflux only, kidney abnormalities only and both VUR and kidney abnormalities as the outcome of interest

Author	Year	Total	Gender M:F	Age range, yr (mean)	Tests Compared	Time from UTI	Test positive definition	Spectrum UTI
VUR only					MCU vs		VUR	
Amar 1	1964	40	-	-	Indigo carmine dye	-	-	-
Amar 2	1964	60*	12:48	<14	Indigo carmine dye	-	-	-
Smith	1966	145	-	-	Cine MCU	-	-	-
Blaufox	1971	47	17:30	2mo- 15.5	n MCU	6 wks	-	At least 1
Conway	1972	100	41:59	9mo-16	n MCU	-	-	-
Saxena	1975	97	30:67	0-13	IVP	-	-	1 st
Harrison	1976	82	-	-	IVP	-	-	-
Merrick	1977	57	-	-	Indirect voiding study	6wk – 3mo	-	1st and recurrent
Hedman	1978	51	7:55	2-15 (8yr)	DTPA vs r MCU	-	-	Recurrent
Lanning	1979	255	-	0-15	IVP	-	IRS I-V	1 st
Thompson	1979	37	8:29	2-27 (med=6)	Retrograde vs antegrade	Post UTI	I-IV	Recurrent
Pollet	1981	110	-	7mo-21	IVP	-	Lyon et al (1970) Hodson and Wilson (1965)	-
Nasrallah	1982	86	11:75	2-19	n MCU	-	-	-
Brendstrup	1983	30	9:21	5 mo-15	n-MCU	-	I-IV	1 st
Drachman	1984	191	14:177	6 mo-12	IVP	4-6wks	I-IV	1 st
Fretzayas	1984	36	13:23	6mo-14 (4.5)	n MCU	4 wks	IRS I-V	>= 1
Majd	1984	120	30:90	3-16	Direct vs indirect	-	-	-
Redman	1984	200	0:200	0-14	IVP	Post infection	-	History of UTI
Schneider	1984	110	35:75	0-7	US	-	-	80% UTI
Nielson	1985	33	1:32	6-14 (9)	IVP	MCU > 1 month	-	Recurrent
Rizzoni	1985	48	27:21	4days-14	Colloidal MCU vs IVP	-	I-IV (Dwoskin et al 1973)	Recurrent
Martin	1986	53	-	6mo-16	nMCU vs US	-	IRS I-V	Recurrent
Chapman	1988	47	-	5-15	Direct vs indirect	-	IRS I-V	-
Tan	1988	100	2:1	0-11 (1.76)	US	-	I-IV (Smellie)	66% UTI culture
Bergius	1989	124	65:59	0-10+	US	-	IRS I-V	106/124 recurrent
Fettich	1992	428	144:284	4mo-7	Cyclic MCU	-	I-III	-
Rickwood	1992	200	59:141	<10	US	-	-	At least 1
Blane	1993	493	134:359	1wk-19 (4.9)	US	-	I-III	US
Salih	1994	21	7:14	2-16 (6.6)	Colour US	-	I-IV	-
Merrick1	1995	3646	-	-	Indirect voiding study	Within 1 mo	-	>=1
Davey	1997	455	76:379	0-10 (4.2)	US	-	IRS I-V	-
DiPietro	1997	70	6:64	5-15 (7.3)	US	-	-	-
Atala	1998	20	-	-	US	-	IRS I-V	-
Balbay	1998	26	10:16	-	US	-	-	-
Bosio	1998	58	33:25	0-10	Contrast enhanced US, indirect MCU vs ceUS	1mo	IRS I-V	At least 1
Poli Merol	1998	108	24:84	5mo-14	n MCU	UTI culture –ve	-	2+
Mentzel	1999	46	12:34	3wks-14	Contrast enhanced US	-	IRS I-V	54.3% UTI
Abnormality only					Definition of RPA			
Merrick	1980	79	-	1->10	DMSA vs IVP	-	-	-
Leonidas	1983	62	9:53	7 da-13 (4.2)	IVP vs 5 min IVP	-	-	-
Sherwood	1984	50	20:30	>6 mo (7 yr)	IVP vs US	-	-	-
Thelle	1985	87	14:73	5 mo-14	DMSA vs IVP	-	-	-
LeQuesne	1986	33	-	5 mo-14	DMSA vs US	-	-	-
Lindsell	1986	100	29:71	<1 yr - > 5 yr	IVP vs US	-	-	-
Ben-Ami	1989	283	54:229	0-15 (-)	IVP vs US	-	-	-
Goldraich	1989	202	29:173	0-14 (-)	DMSA vs IVP	> 4 wks	Defect in outline, little/no DMSA uptake	-
Kenda	1989	101	34:67	0-6 (-)	IVP vs US	2wk – 3 mo	-	-
Joseph	1990	35	5:30	6 mo-14(4yr)	Planar vs Spect DMSA, planar vs pinhole DMSA, spect vs pinhole DMSA	-	-	-
Jakobss~1	1992	72	13:59	0.1-15.9 yr (1.1 yr)	DMSA vs US, DMSA vs IVP	US, DMSA-< 5 days	Tracer uptakebelow 2 SD of control kidneys	APN

Kass	1992	46	6:40	8 da-10 yr (3.6yr)	DMSA vs US	MCU, IVU-6-20 wk US < 48 hr DMSA < 96 hr < 3 days	Decreased tracer with preserved outline	APN	
Benador	1994	111	51:60	1 wk-16yr (5.5 mo)	DMSA vs US		Focal/diffuse decrease or absence of DMSA uptake	1 st = 87 >1 = 24	
Stokland	1994	25	-	2-16 yr (6.5 yr)	IVP vs US	-	Caliceal deformation, parenchymal reduction	-	
Hansen	1995	90	32:58	0-13	DMSA vs IVP, DMSA vs US IVP vs US	US, MCU dayS-wks of UTI, DMSA 1-3 mo post UTI	<45% uptake, abnormal outline in outer contour	At least 1	
Merrick2	1995	3646	-	0-14yr	DMSA/DTPA vs US	-	-	>=1	
Sreenaras~	1995	50	8:42	2mo-15	US vs glucoheptonate renal scan	GHRs, RUS, IVP 48-96 hrs post admission, MCU post infection	-	Pyelonephritis	
Barry	1998	300	124:176	18 day – 13.6 yr (2.99 yr)	DMSA vs US	US – between 1 and 3 months DMSA – 3 months	-	-	
Both VUR and Abnormality									
Johnson	1986	64	7:57	<13	IVP vs MCU US vs MCU US vs IVP	1 st = 5 days post initiation of antibiotics 2 nd = 4-6wks after 1 st series	-	At least 1	
Bisset	1987	523	0:523	1mo-18	MCU vs IVP	-	VUR; IRS I-V RPA; blunting, clubbing of calices, parenchymal loss, outline abnormality	At least 1	
Meller	1987	25	-	-	DMSA vs MCU	-	-	1 st	
Verber	1988	105	35:80	0-59 mo (1.7 mo)	DMSA vs IVP DMSA vs MCU MCU vs IVP MCU vs US IVP vs MCU	-	VUR; 1-3 Bailey RPA; abnormal outline,	1 st	
Hellstrom	1989	88	23:65	2mo-6 yr	IVP vs MCU	-	VUR; IRS I-V RPA; reduced parenchymal thickness, +/- or caliceal dormity (Hodson 1967)	1 st	
Tappin	1989	102	27:75	0-13 yr	DMSA vs MCU	MCU – 1-30 wks	VUR; reached pelvis(gradeII) RPA; reuced uptake, abnormal outline, differential function > 12%	1 st E.coli	
Bjorgovins~	1991	91	29:62	1 wk-10 yr (2.7yr)	MCU vs US DMSA vs US	<72 hr	VUR; - RPA; decreased uptake, preserved outline	At least 1	
Farnsworth	1991	113	60:53	< 1 yr	DMSA vs IVP	-	VUR; IRS I-V RPA; reduced function+/- or decreased uptake + outline abnormality	-	
Gleeson	1991	94	33:61	<1->5	DMSA vs MCU	-	-	1 st	
Jakobsson2	1992	106	33:73	0-15.9 yr(1.1yr)	DMSA vs CRP DMSA vs US DMSA vs IVP DMSA vs MCU	CRP-on admission for UTI US+DMSA+de smopressin <5 da IVP+MCU – 6-20 wks <72 hr	VUR; IRS I-V RPA; -	Symptomatic	
Majd	1992	94	36:58	2 wk-18.9yr	DMSA vs MCU	-	-	At least 1	
Melis	1992	146	57:89	1 wk-16 yr	DMSA vs P.fim E.coli DMSA vs MCU DMSA vs US	MCU- 3-6 wks	VUR; IRS I-V RPA; -	-	

Pickworth	1992	100	-	6 da-17 yr	MCU vs indirect MCU DMSA vs MAG3 IVP vs MAG3	-	-	-	
Rosenberg	1992	65	42:88	- (21.7mo)	DMSA vs MCU	Within 3wk	VUR; -IRS I-V RPA; >10% differential function+/-or reduced uptake causing abnormal outline	1 st	
Scherz	1994	75	22:53	3wk-12 yr	DMSA vs US	-	VUR; IRS I-V RPA; < 43% renal function +/-or decreased/absent uptake	-	
Smellie	1995	58	12:46	1 mo-13.5 yr (3.3yr)	IVP vs US	Within 3 mo	-	1 st = 29, recurrent = 29	
Lavocat	1997	55	23:32	3 wk-15.5 yr (33.4 mo)	DMSA vs US DMSA vs MCU DMSA vs CT scan	MCU-> 3 wks, DMSA-< 1 wk	VUR; IRS I-V RPA; <45% uptake, 1 or more cortical defects, hypoactivity, small vol, +/-or abnormal outline	-	
Morin	1999	70	27:43	1 mo-17 yr	DMSA vs US DMSA vs MCU MCU vs US	US on admission, DMSA<5 da, M<CU <4 wk	VUR; IRS I-V RPA; decreased uptake	Pyelonephritis	
Obstruction									
Kass	1985	34	-	1 wk-16 yr (-)	Renogram vs diuretic renogram Renogram vs Whitaker	-	Obstructed – a half time >20 mins	Hydronephrosis	

Table 2. Quality characteristics of papers, grouped by outcome of vesicoureteric reflux only, kidney abnormalities only and both vesicoureteric reflux and kidney abnormalities as the outcome.

Author	Year	Consec	Prosp.	Complete verification	Adequate ref test details	Adequate Comp test details	Ref test done blind to test	Comp test done blind to ref test	Blind to clinical info	Population described
VUR only										
Amar 1	1964	-	Yes	Yes	No	Yes	-	-	-	No
Amar 2	1964	Yes	-	Yes	No	No	-	-	-	No
Smith	1966	-	Yes	Yes	No	No	-	-	-	No
Blaufox	1971	-	Yes	Yes	No	No	-	-	-	Yes
Conway	1972	No	Yes	Yes	No	No	-	-	-	Yes
Saxena	1975	-	Yes	-	No	No	-	-	-	Yes
Harrison	1976	-	No	Yes	No	No	-	-	-	No
Merrick	1977	-	No	Yes	No	Yes	-	-	-	No
Hedman	1978	-	Yes	Yes	No	Yes	-	-	-	Yes
Lanning	1979	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	No
Thompson	1979	-	-	Yes	Yes	Yes	-	-	-	Yes
Pollet	1981	-	-	Partial	Yes	Yes	-	-	-	No
Nasrallah	1982	-	-	Yes	No	No	-	-	-	No
Brendstrup	1983	-	-	Yes	Yes	Yes	-	-	-	Yes
Drachman	1984	-	-	Yes	No	No	-	-	-	Yes
Fretzayas	1984	-	-	Yes	Yes	No	-	-	-	Yes
Majid	1984	-	-	Yes	Yes	Yes	-	-	-	Yes
Redman	1984	-	Yes	Yes	No	No	-	-	-	Yes
Schneider	1984	-	Yes	-	No	Yes	-	-	-	Yes
Nielson	1985	Yes	-	Yes	Yes	Yes	-	-	-	Yes
Rizzoni	1985	-	Yes	Yes	Yes	No	Yes	Yes	-	Yes
Martin	1986	-	Yes	Yes	Yes	Yes	-	-	-	No
Chapman	1988	-	Yes	Yes	No	Yes	-	-	-	No
Tan	1988	-	Yes	Yes	Yes	Yes	-	-	-	Yes
Bergius	1989	Yes	Yes	Yes	Yes	Yes	-	-	-	Yes
Fettich	1992	-	Yes	Yes	Yes	Yes	-	-	-	Yes
Rickwood	1992	Yes	Yes	Partial	No	No	-	-	-	Yes
Blane	1993	-	No	Yes	No	No	No	No	No	Yes
Salih	1994	-	Yes	Yes	No	Yes	Yes	Yes	-	Yes
Merrick1	1995	Yes	Prosp ID. Retro review	Yes	No	No	-	-	-	No
Davey	1997	Yes	No	Yes	Yes	Yes	Yes	Yes	-	Yes
DiPietro	1997	-	No	Yes	No	No	-	-	-	Yes
Atala	1998	-	-	Yes	Yes	Yes	-	-	-	No
Balbay	1998	-	Yes	Yes	No	No	-	-	-	Yes
Bosio	1998	-	Yes	Partial	Yes	No	-	-	-	Yes
Poli-Merol	1998	-	Yes	Partial	No	No	Yes	No	-	Yes
Mentzel	1999	-	Yes	Yes	No	Yes	Yes	Yes	-	Yes
Abnormality only										
Merrick	1980	Yes	No	Yes	No	No	Yes	Yes	-	No
Leonidas	1983	-	-	Yes	No	No	-	-	-	Yes
Sherwood	1984	-	Yes	Yes	No	No	Yes	Yes	-	Yes
Telle	1985	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes
LeQuesne	1986	-	No	-	No	No	-	-	-	No
Lindsell	1986	-	-	Yes	No	No	-	-	-	No
Ben Ami	1989	-	No	Yes	No	No	-	-	-	Yes
Goldraich	1989	-	Yes	Yes	Yes	Yes	Yes	Yes	-	No
Kenda	1989	-	Yes	Yes	No	No	-	-	-	No
Joseph	1990	Yes	No	Yes	No	No	No	No	No	Yes
Jakobss-1	1992	-	-	Yes	No	Yes – US No- IVU	-	-	-	Yes
Kass	1992	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Benador	1994	-	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Stokland	1994	-	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Hansen	1995	-	Yes	Yes	Yes	Yes	-	DMSA yes	-	Yes
Merrick1	1995	Yes	Yes	Yes	No	No	-	-	-	No
Sreenaras~	1995	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Barry	1998	-	Yes	Yes	No	Yes	Yes	Yes	-	Yes

Both VUR and abnormality

Johnson	1986	Yes	Yes	Yes 1st series, Partial 2 nd series	Yes	No	Yes	Yes	Yes	Yes
Bisset	1987	-	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Meller	1987	Yes	-	Yes	Yes	Yes	-	-	-	No
Verber	1988	-	-	Partial	Yes	DMSA - yes	-	-	-	Yes
						IVU - yes				
						US - yes				
Hellstrom	1989	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Tappin	1989	-	Yes	Yes	No	No	Yes	Yes	-	No
Bjorgvins~	1991	Yes	Yes	Yes	Yes	Yes	-	Yes	-	Yes
Farnsworth	1991	Yes	Yes	Yes	No	DMSA yes	-	-	-	No
						IVP-yes				
Gleeson	1991	-	-	Yes	No	No	-	-	-	No
Jakobsson2	1992	-	-	Yes	No	No	-	-	-	Yes
Majd	1992	Yes	Yes	Yes	No	No	-	-	-	Yes
Melis	1992	-	No	Partial	Yes	Yes	-	-	-	Yes
Pickworth	1992	Yes	No	Yes	No	Yes	-	-	-	No
Rosenberg	1992	-	Yes	Partial	No	IVP - yes	-	DMSA	-	No
						DMSA - yes		blind to others		
Scherz	1994	-	-	Yes	No	Yes	-	-	-	No
Smellie	1995	-	Yes	Yes	No	No	-	-	-	Yes
Lavocat	1997	Yes	Yes	Yes	No	Yes	-	-	-	Yes
Morrin	1999	-	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Obstructn Kass	1985	-	-	Yes	No	No	-	-	-	No

Outcome VUR

Table 3.1 Sensitivity, specificity and diagnostic odds ratio for **Intravenous Pyelography** as compared with the reference standard micturating cystourethrogram for the diagnosis of vesicoureteric reflux in children and kidneys.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	Ln DOR	Prev %
Saxena	1975	Child	6	34	4	14	0.310	0.885	1.235	34.5
Harrison	1976	Child	34	23	7	18	0.651	0.758	1.765	63.4
Drachman	1984	Child	25	75	0	77	0.248	0.993	3.906	57.6
Redman	1984	Child	8	147	14	31	0.213	0.911	1.008	19.5
Johnson	1986	Child	2	52	1	9	0.208	0.972	2.220	17.19
Bisset	1987	Child	60	308	21	128	0.320	0.935	1.911	36.36
Hellstrom	1989	Child	8	55	19	2	0.773	0.740	2.270	11.9
Lanning	1979	Kidney	50	368	46	45	0.526	0.888	2.175	18.86
Pollet	1981	Kidney	15	40	38	17	0.470	0.513	-0.073	29.1
Nielson	1985	Kidney	11	39	5	11	0.500	0.878	1.971	33.3
Verber	1988	Kidney	23	44	16	25	0.480	0.730	0.9122	44.4

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (P)	Mean DOR (95%CI)	Median DOR (IQR)
7	child	1182	0.208-0.773	0.740-0.993	45.80 (P<0.001)	12.05 (-2.20-26.39)	6.76 (6.24)
4	kidney	793	0.470-0.526	0.513-0.888	23.72 (P<0.001)	4.85 (-1.10-10.80)	4.83 (7.08)

Figure 1.1 Forest plot of **Intravenous Pyelogram** for the diagnosis of vesicoureteric reflux (children)

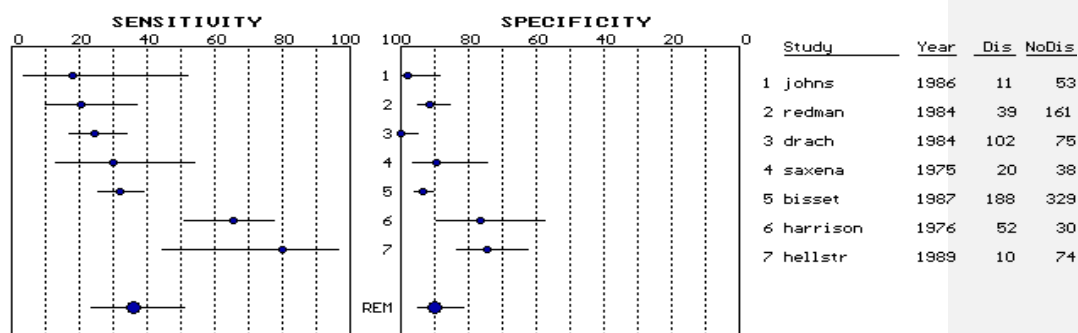


Figure 1.2 Summary receiver operator characteristic curve of **Intravenous Pyelogram** for the diagnosis of vesicoureteric reflux (children)

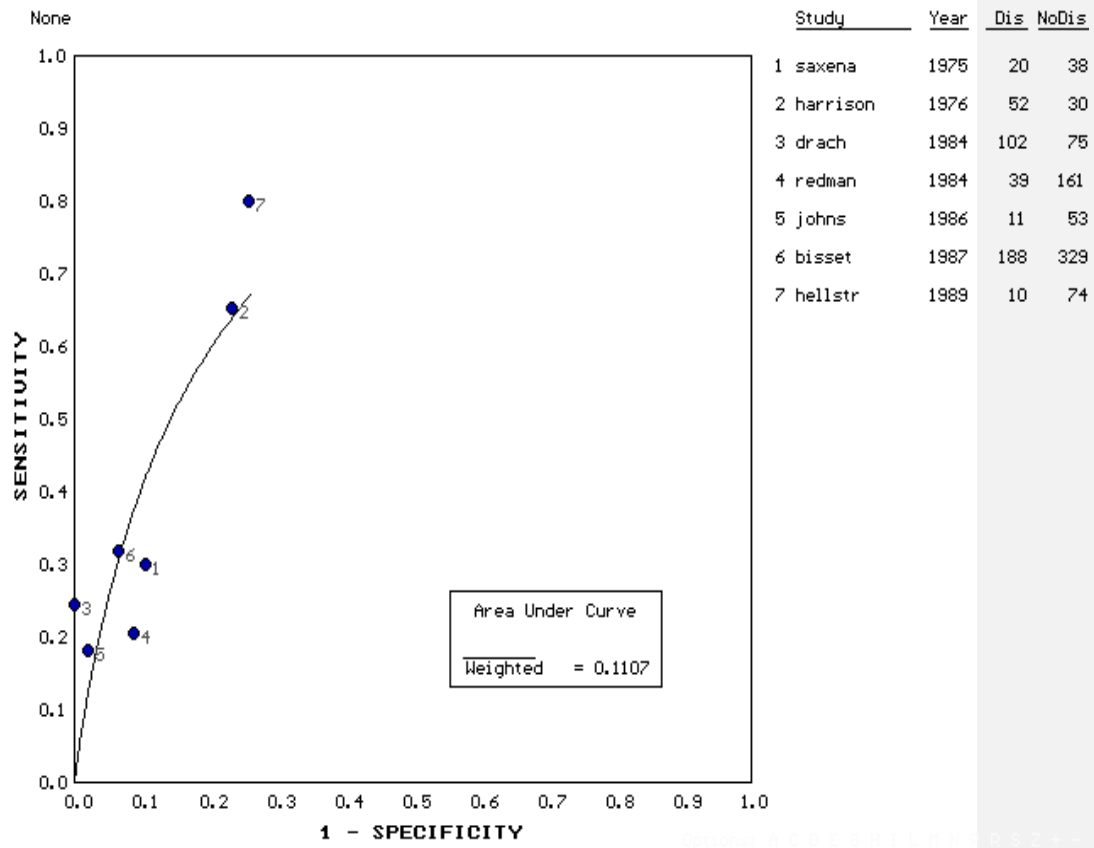


Figure 1.3 Forest plot of intravenous pyelogram for the diagnosis of vesicoureteric reflux (kidneys)

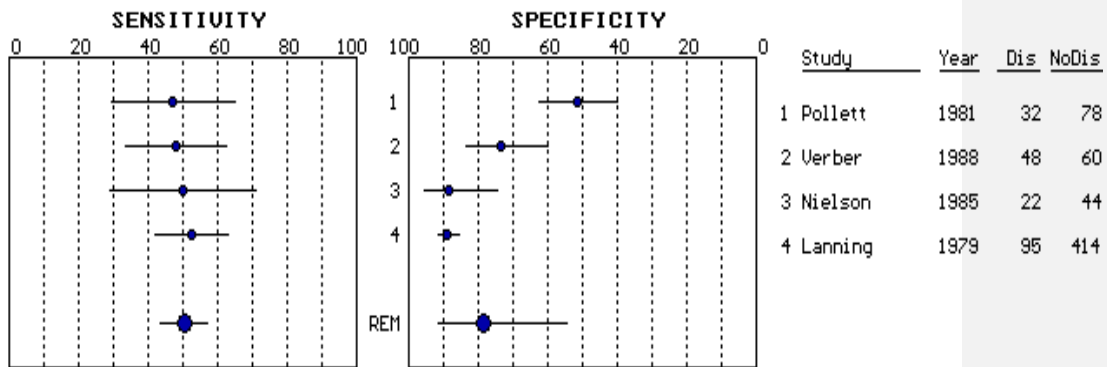


Figure 1.4 Summary receiver operator characteristic curve of intravenous pyelogram for the diagnosis of vesicoureteric reflux (kidneys)

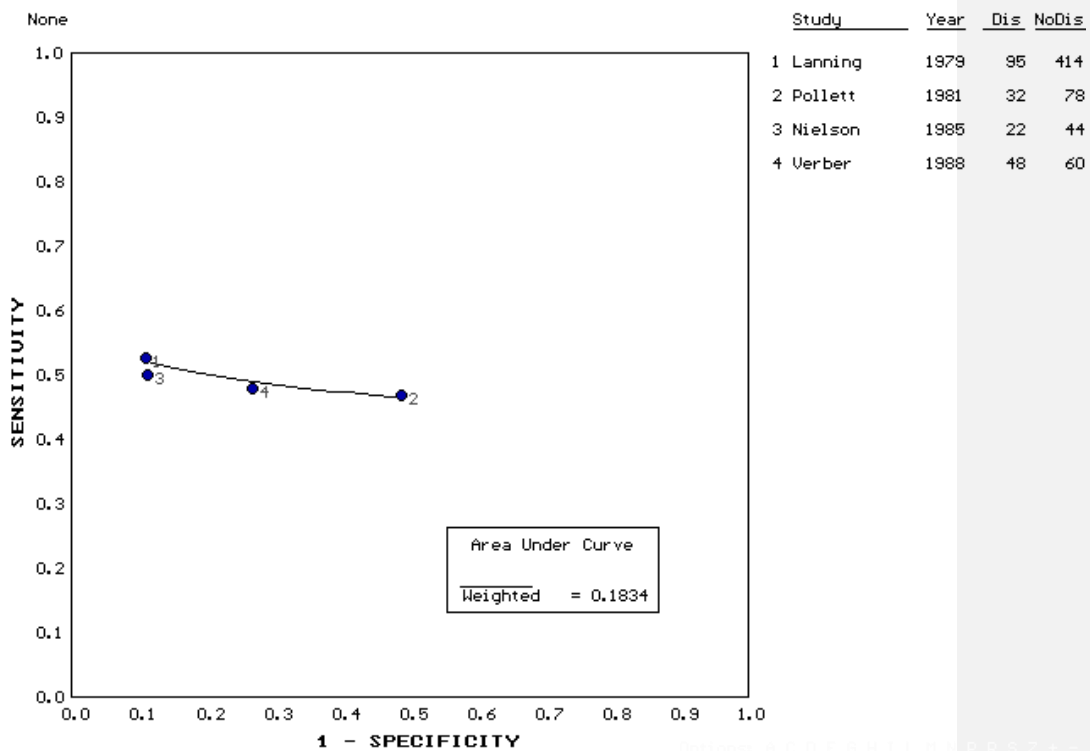


Table 3.2 Sensitivity, specificity and diagnostic odds ratio for **Ultrasound** compared with the reference standard micturating cystourethrogram for diagnosing vesicoureteric reflux in children and kidneys.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	Ln DOR	%Prev
Tan	1988	Child	3	32	6	14	0.194	0.833	0.191	30.9
Rickwood	1992	Child	13	129	22	36	0.270	0.852	0.756	24.5
Smellie	1995	Child	15	10	2	21	0.419	0.808	1.109	75.0
Davey	1997	Child	25	231	50	149	0.144	0.822	-0.248	38.2
DiPietro	1997	Child	2	46	3	19	0.114	0.930	0.531	30.0
Atala	1998	Child	4	3	0	1	0.750	0.875	3.045	62.5
Morrin	1999	Child	20	7	41	2	0.891	0.153	0.392	31.4
Schneider	1984	Kidney	46	141	15	17	0.727	0.901	3.189	28.7
Martin	1986	Kidney	11	5	3	2	0.821	0.611	1.978	61.9
Verber	1988	Kidney	8	25	9	20	0.293	0.729	0.104	45.2
Bergius	1989	Kidney	58	176	2	14	0.801	0.986	5.651	29.0
Blane	1993	Kidney	71	714	0	201	0.262	0.999	6.229	27.6
Davey	1997	Kidney	19	571	71	249	0.071	0.889	-0.494	29.5
Balbay	1998	Kidney	6	43	3	0	0.929	0.926	5.085	11.5
Bosio	1998	Kidney	15	1	1	0	0.969	0.500	3.434	88.2

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (P)	Mean DOR (95% CI)	Median DOR (IQR)
7	child	1239	0.114-0.891	0.153-0.930	11.32 (0.100<P<0.05)	4.48 (-2.29-11.25)	1.70 (1.82)
8	kidney	2517	0.073-0.969	0.500-0.999	128.04 (P<0.001)	127.20 (-26.48-280.90)	27.63 (251.19)

Figure 2.1 Forest plot of **ultrasound** for the diagnosis of vesicoureteric reflux (children)

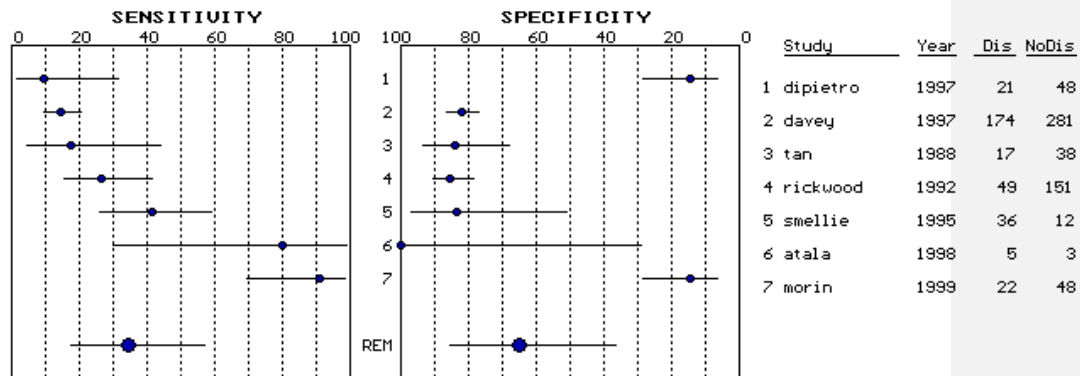


Figure 2.2 Summary receiver operator characteristic curve of ultrasound for the diagnosis of vesicoureteric reflux (children)

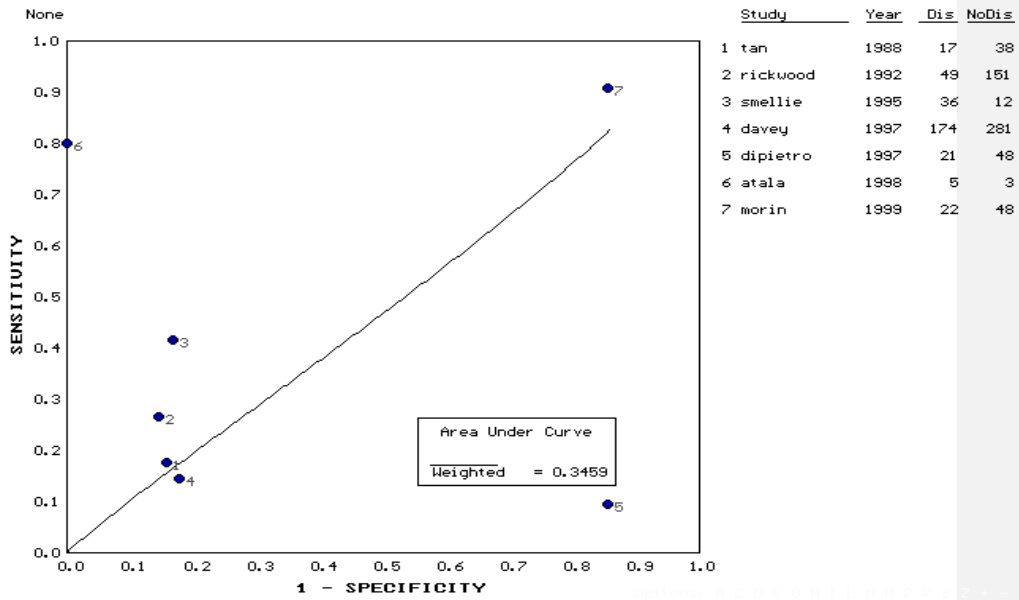


Figure 2.3 Forest plot of **ultrasound** for the diagnosis of vesicoureteric reflux (kidneys)

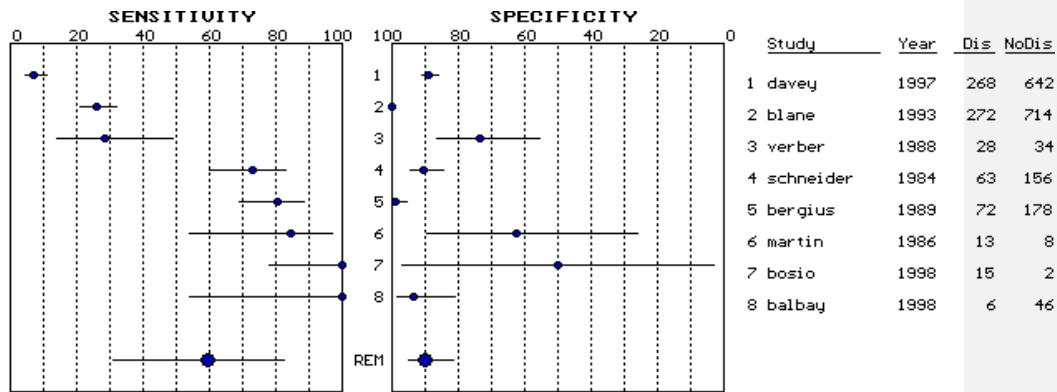


Figure 2.4 Summary receiver operator characteristic curve of **ultrasound** for the diagnosis of vesicoureteric reflux (kidneys)

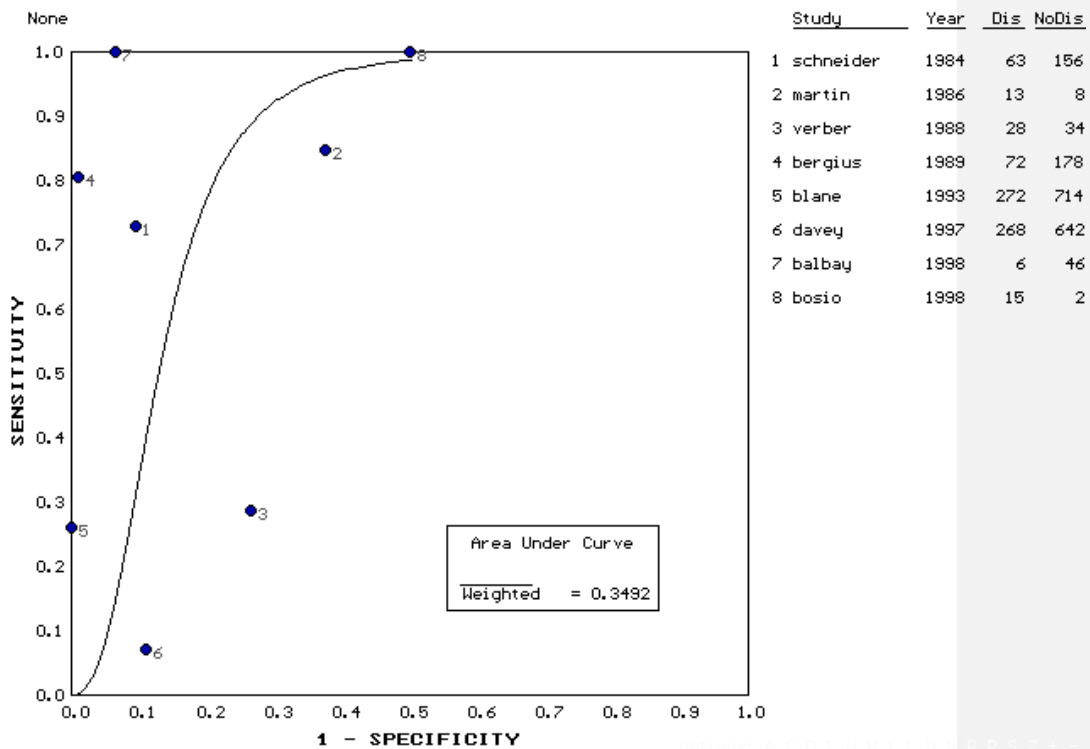


Table 3.3 Sensitivity, specificity and diagnostic odds ratio for **dimercaptosuccinic acid** compared with reference standard micturating cystourethrogram for detecting vesicoureteric reflux in children and kidneys.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	LnDOR	Prev.%
Bjorgvinsson	1991	Child	25	27	30	5	0.455	0.844	1.430	34.5
Jakobsson2	1992	Child	23	20	60	3	0.910	0.560	0.822	21.7
Majd	1992	Child	23	26	30	6	0.783	0.465	1.144	34.1
Rosenberg	1992	Child	10	19	15	1	0.875	0.557	2.176	24.4
Morrin	1999	Child	18	4	44	4	0.818	0.083	-0.892	31.4
Mellar	1987	Kidney	10	35	5	0	0.955	0.866	0.909	20.0
Verber	1988	Kidney	40	44	26	20	0.664	0.627	1.20	46.2
Tappin	1989	Kidney	9	103	21	13	0.409	0.831	1.221	15.1
Farnsworth	1991	Kidney	48	24	9	90	0.349	0.721	0.3221	80.7
Gleeson	1991	Kidney	26	102	16	44	0.373	0.861	1.308	37.2
Melis	1992	Kidney	42	75	69	10	0.802	0.521	1.482	26.5
Scherz	1994	Kidney	23	35	5	92	0.203	86.59	0.495	76.7
Lavocat	1997	Kidney	17	48	39	6	0.729	55.11	1.197	20.9

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (p)	Mean DOR (95% CI)	Median DOR (IQR)
5	child	402	0.455-0.910	0.083-0.844	11.07 (0.025<P<0.05)	3.76 (-0.13-7.66)	3.14 (5.15)
8	kidney	1146	0.203-0.955	0.521-0.866	12.03 (P>0.10)	2.95(2.09-3.81)	3.32 (1.77)

Figure 3.1 Forest plot of **dimercaptosuccinic acid** for the diagnosis of vesicoureteric reflux (children)

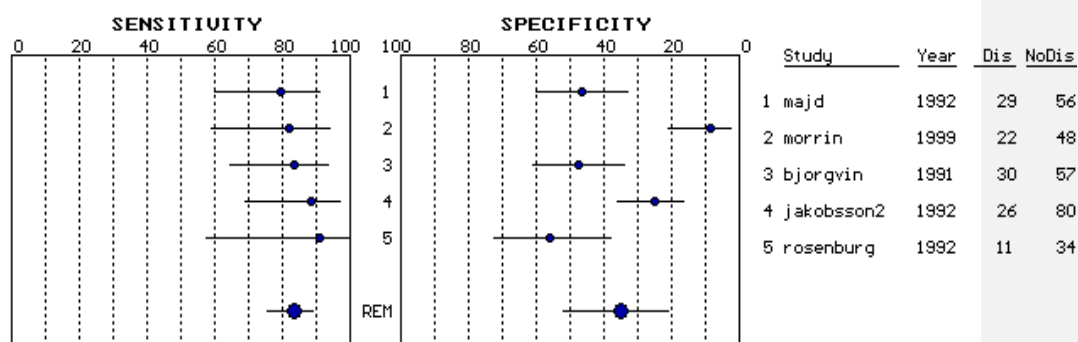


Figure 3.2 Summary receiver operator characteristic curve of Dimercaptosuccinic acid for the diagnosis of vesicoureteric reflux (children)

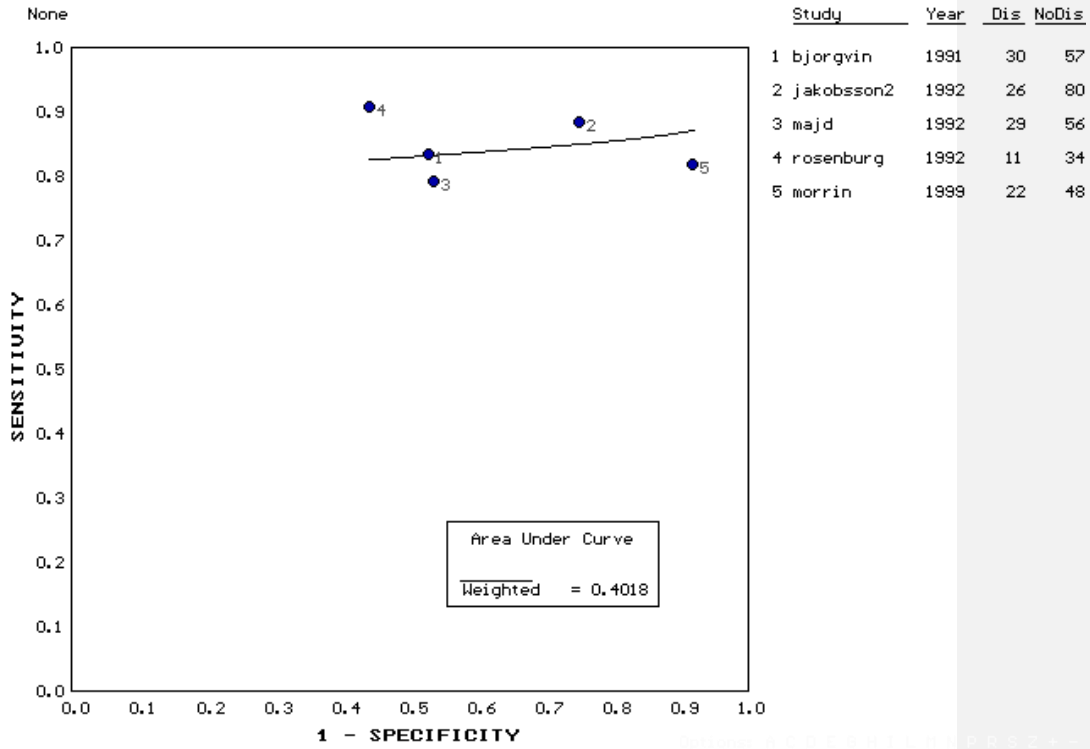


Figure 3.3 Forest plot of Dimercaptosuccinic acid for the diagnosis of vesicoureteric reflux (kidneys)

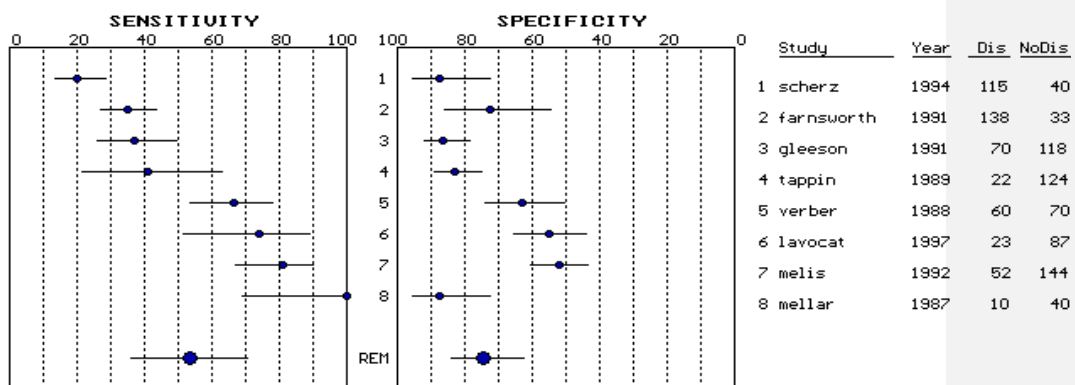
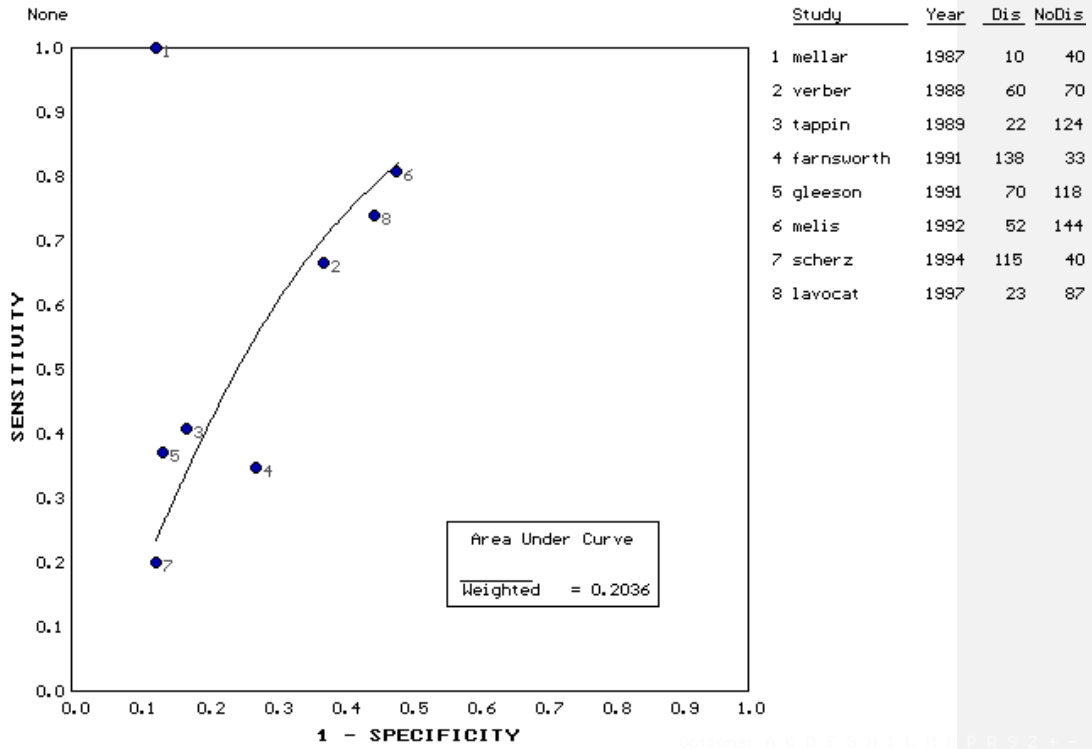


Figure 3.4 Summary receiver operator characteristic curve of **Dimercaptosuccinic acid** for the diagnosis of vesicoureteric reflux (kidneys)



Comparisons that had less than 5 studies and were therefore not included in the meta-analysis

Outcome VUR

Table 3.4. Sensitivity, specificity and diagnostic odds ratios

Author	Year	Ref Test vs Comparison test	Units	TP	TN	FP	FN	Sens.	Spec.	lnDOR	Prev. %
Amar 1	1964	MCU vs Indigo carmine dye	Child	17	21	1	1	0.944	0.955	5.119	45.0
Amar 2	1964	MCU vs Indigo carmine dye	Child	26	29	2	3	0.897	0.935	4.492	48.3
Smith	1966	MCU vs Cine MCU	Child	25	100	9	11	0.694	0.917	3.155	24.8
Blaufox	1971	MCUr vs MCUn	Child	6	35	2	4	0.600	0.946	3.021	21.3
Conway	1972	MCUr vs MCUn	Kidney	44	142	9	5	0.898	0.940	4.799	24.5
Merrick	1977	Indirect vs direct MCU	Kidney	24	62	12	16	0.600	0.838	2.004	35.1
Hedman	1978	MCU vs DTPA	Kidney	13	77	4	8	0.619	0.951	3.309	20.6
Thompson	1979	Retrograde vs antegrade MCU	Kidney	22	28	3	17	0.564	0.903	2.349	55.7
Nasrallah	1982	MCUr vs MCUn	Child	22	45	15	4	0.846	0.750	2.686	30.2
Brendstrup	1983	MCUr vs MCUn	Kidney	25	32	4	3	0.893	0.889	3.963	43.8
Fretzayas	1984	MCUr vs MCUn	Child	15	17	3	1	0.938	0.850	4.459	44.4
Majd	1984	Indirect vs direct MCU	Child	68	20	32	0	1.000	0.385	4.459	56.7
Schneider	1984	MCU vs separated CRCs	Kidney	40	144	12	23	0.635	0.923	2.992	28.8
Rizzoni	1985	MCUr 5hrs vs IVP, MCUr 20hrs vs IVP	Kidney	23 11	13 13	7 6	6 0	0.793 1.000	0.650 0.684	1.873 3.866	59.2 36.7
Chapman	1988	Indirect vs direct MCU	Kidney	30	22	17	16	0.652	0.564	0.867	54.4
Pickworth	1992	Indirect vs direct MCU	Kidney	8	22	4	8	0.500	0.846	1.609	38.1
Fettich	1992	MCU vs Cyclic MCU	Kidney	115	625	70	46	0.714	0.899	3.093	18.8
Salih	1994	MCU vs colour flow doppler US	Kidney	26	12	3	1	0.963	0.800	4.145	64.3
Merrick1	1995	MCU vs indirect MCUn	Child	99	214	88	208	0.322	0.709	0.148	50.4
Poli-Merol	1998	MCUr vs MCUn	Child	13	66	21	8	0.619	0.756	1.591	19.4
Bosio	1998	Indirect MCU vs US	Kidney	20	3	0	12	0.625	1.000	2.441	91.4
Mentzel	1999	MCU vs contrast enhanced US	Kidney	11	67	5	1	0.917	0.931	4.544	14.3

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity
9	child	1248	0.322-1.000	0.385-0.955
13	kidney	1962	0.500-1.000	0.564-1.000

MCUr = radiologic MCU

MCUn = nuclear MCU

Outcome Renal Parenchymal Abnormality

Table 4.1. Sensitivity, specificity and diagnostic odds ratio for **ultrasound** compared with reference standard intravenous pyelography for the diagnosis of renal parenchymal abnormality.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	lnDOR	Prev. %
Sherwood	1984	Child	7	36	3	4	0.636	0.913	2.855	22.0
Lindsell	1986	Child	25	63	0	12	0.676	0.992	5.557	37.0
Ben-Ami	1989	Child	29	29	4	9	0.7.63	0.868	3.01	53.5
Kenda	1989	Child	90	5	2	4	0.957	0.688	3.79	93.1
Hansen	1995	Child	8	59	2	21	0.276	0.960	2.242	32.2
Smellie	1995	Child	9	17	1	11	0.450	0.921	2.266	52.6
Stokland	1994	Kidney	32	94	7	27	0.542	0.931	2.701	36.9

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (p)	Mean DOR (95% CI)	Median DOR (IQR)
6	child	369	0.276-0.957	0.688-0.992	4.68 (0.50>P>0.25)	60.0 (-43.20-163.21)	18.83 (88.37)
1	kidney	160	0.542	0.931	N/A	N/A	N/A

Figure 4.1 Forest plot of **ultrasound** compared to the reference standard intravenous pyelogram for the diagnosis of renal parenchymal abnormality (children)

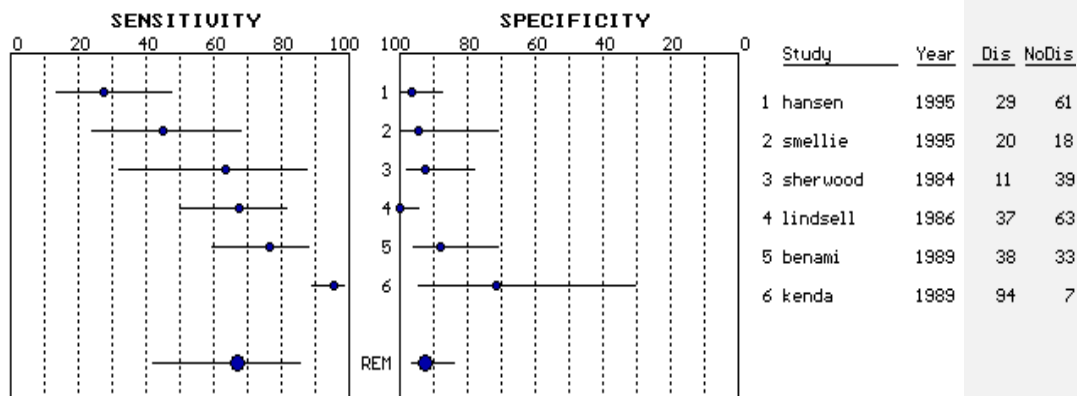
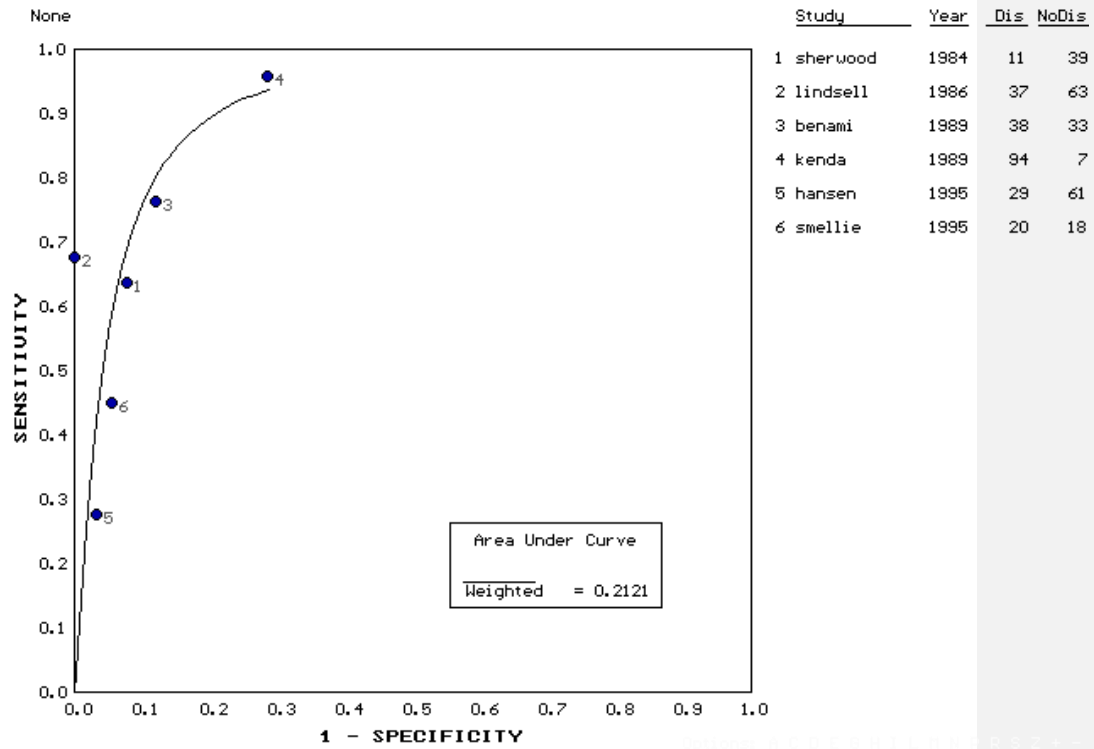


Figure 4.2 Summary receiver operator characteristic curve of ultrasound compared with intravenous pyelography for the diagnosis of renal parenchymal abnormality



(children)

Table 3.8 Sensitivity, specificity and diagnostic odds ratio for **intravenous pyelography** compared with the reference standard dimercaptosuccinic acid for the diagnosis of renal parenchymal abnormality.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	lnDOR	Prev.%
Verber	1988	Child	32	35	2	23	0.582	0.946	2.98	37.0
Jakobsson2	1992	Child	15	21	2	68	0.181	0.913	1.946	76.9
Hansen	1995	Child	11	56	18	5	0.688	0.757	6.386	17.8
Merrick3	1980	Kidney	47	98	2	8	0.855	0.980	5.394	35.5
Thelle	1985	Kidney	26	66	13	5	0.667	0.930	3.167	28.2
Goldraich	1989	Kidney	80	182	3	34	0.702	0.984	4.801	38.1
Farnsworth	1991	Kidney	18	127	3	49	0.269	0.977	2.611	34.0
Jakobsson1	1992	Kidney	24	50	17	33	0.421	0.746	0.747	46.0

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (P)	Mean DOR (95% CI)	Median DOR (IQR)
3	child	288	0.181-0.688	0.757-0.946	48.41 (P<0.001)	206.72 (-625.47-1038.91)	19.69 (NA)
5	kidney	977	0.269-0.855	0.746-0.984	52.73 (P<0.001)	76.23 (-39.79-192.26)	23.74 (163.0)

Figure 5.1 Forest plot of **Intravenous pyelography** compared with the reference standard dimercaptosuccinic acid for the diagnosis of renal parenchymal abnormality (kidney)

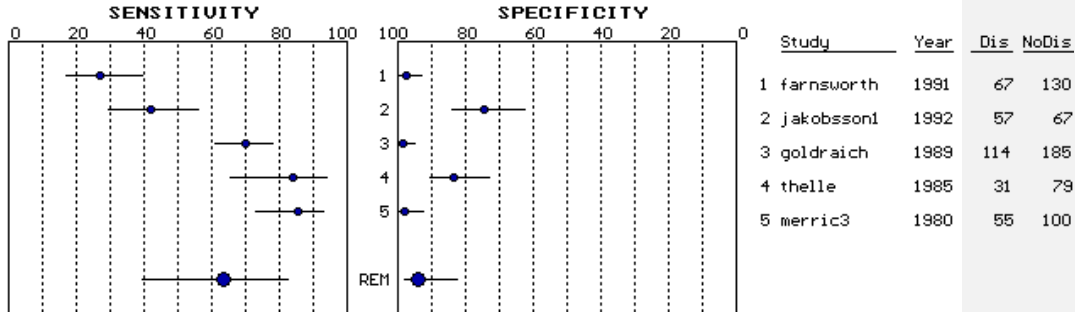


Figure 5.2 Summary receiver operator characteristic curve of intravenous pyelography compared to the reference standard dimercaptosuccinic acid for the diagnosis of renal parenchymal abnormality (kidney)

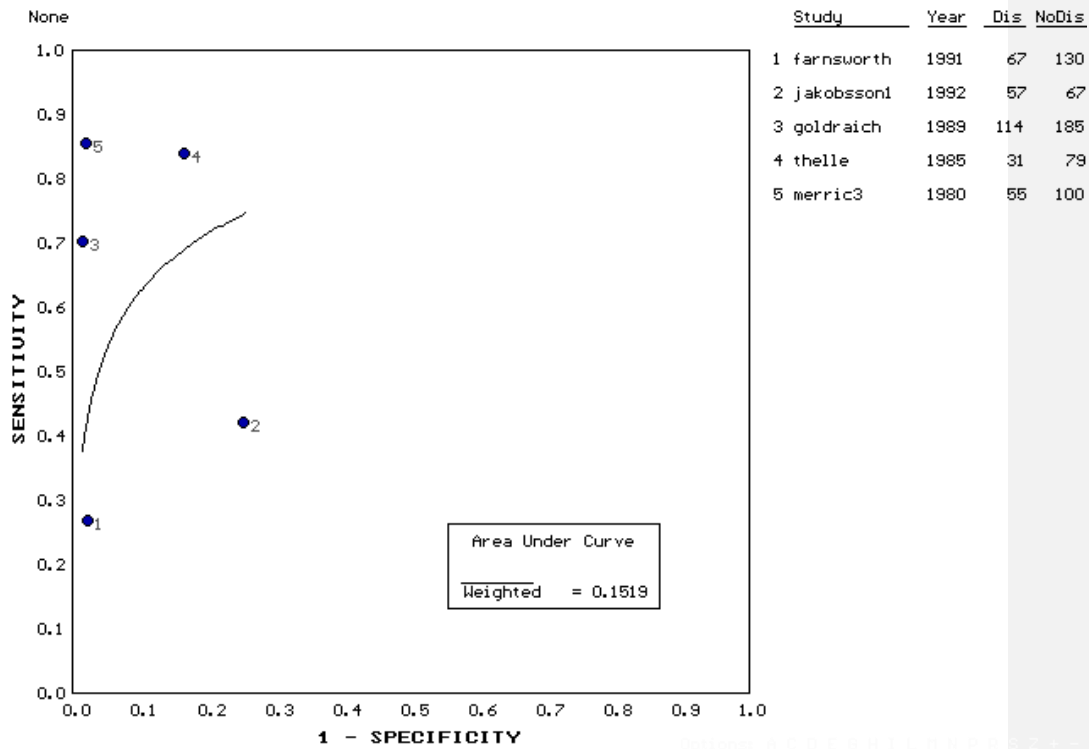


Table 3.9 Sensitivity, specificity and diagnostic odds ratio for **micturating cystourethrogram** compared with reference standard dimercaptosuccinic acid for diagnosing renal parenchymal abnormality.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	lnDOR	Prev. %
Bjorgvinsson	1991	Child	25	27	5	30	0.455	0.844	1.430	63.2
Jakobsson2	1992	Child	23	20	3	60	0.277	0.870	0.824	78.3
Majd	1992	Child	23	26	6	30	0.434	0.813	1.144	72.9
Rosenberg	1992	Child	10	19	1	15	0.400	0.950	2.176	55.6
Morrin	1999	Child	18	4	4	44	0.290	0.500	0.868	88.6
Mellar	1987	Kidney	10	35	0	5	0.667	1.000	4.909	30.0
Verber	1988	Kidney	40	44	20	26	0.606	0.688	1.200	50.8
Tappin	1989	Kidney	9	103	13	21	0.300	0.888	1.221	20.5
Farnsworth	1991	Kidney	48	24	90	9	0.842	0.211	0.322	33.3
Gleeson	1991	Kidney	26	102	44	16	0.619	0.698	1.308	22.3
Melis	1992	Kidney	42	75	10	69	0.378	0.882	1.482	56.6
Scherz	1994	Kidney	23	35	92	5	0.821	0.276	10.495	18.7
Lavocat	1997	Kidney	17	48	6	39	0.304	0.889	1.197	50.9

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (P)	Mean DOR (95% CI)	Median DOR (IQR)
5	child	402	0.277-0.455	0.500-0.950	9.11 (0.1>P>0.05)	4.17 (0.82-7.52)	3.14 (4.13)
8	kidney	1146	0.300-0.821	0.211-1.000	12.03 (0.1>P>0.05)	19.59 (-19.57-58.76)	3.36 (2.09)

Figure 6.1 Forest plot of **micturating cystourethrogram** for the diagnosis of renal parenchymal abnormality (children)

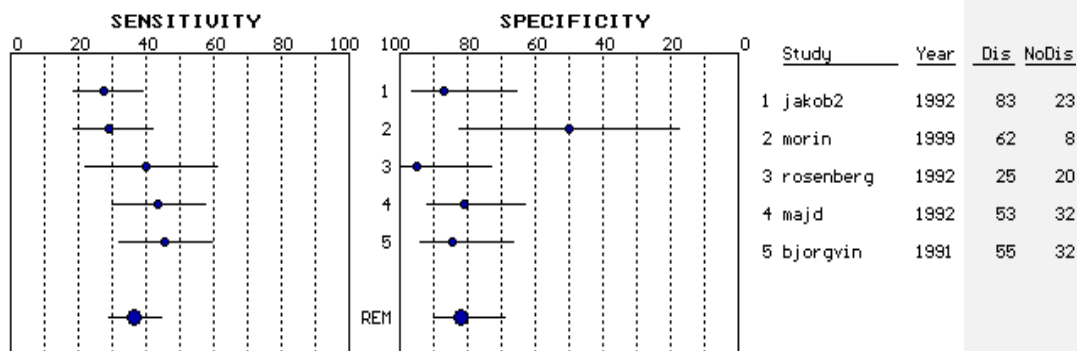


Figure 6.2 Summary receiver operator characteristic curve of micturating cystourethrogram for the diagnosis of renal parenchymal abnormality (children)

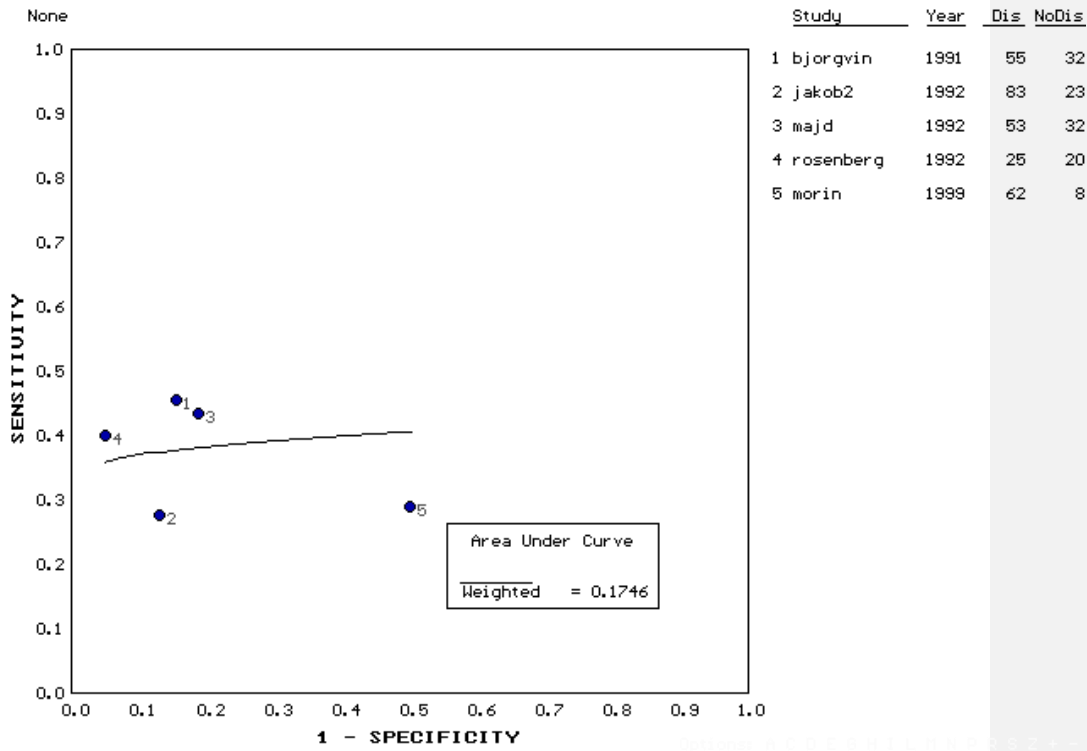


Figure 6.3 Forest plot of micturating cystourethrogram for the diagnosis of renal parenchymal abnormality (kidney)

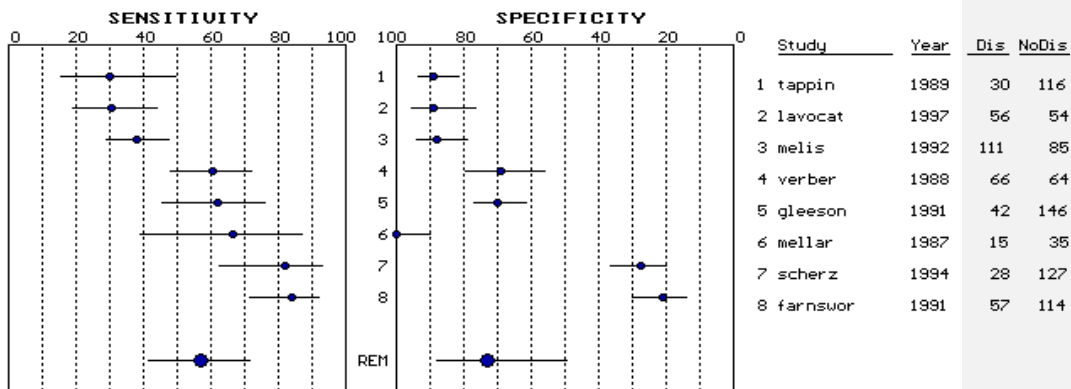
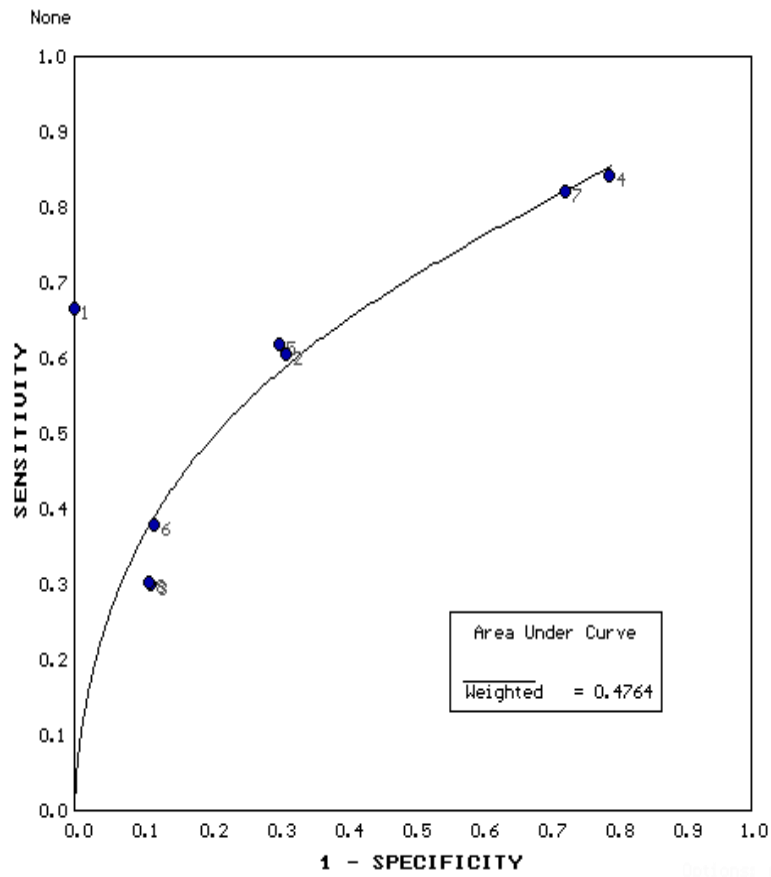


Figure 6.4 Summary receiver operator characteristic curve of micturating

cystourethrogram for the diagnosis of renal parenchymal abnormality (kidneys)



Study	Year	Dis	NoDis
1 mellar	1987	15	35
2 verber	1988	66	64
3 tappin	1989	30	116
4 farnswor	1991	57	114
5 gleeson	1991	42	146
6 melis	1992	111	85
7 scherz	1994	28	127
8 lavocat	1997	56	54

Table 3.10 Sensitivity, specificity and diagnostic odds ratio for **ultrasound compared** with the reference standard dimercaptosuccinic acid for the diagnosis of renal parenchymal abnormality.

Author	Year	Units	TP	TN	FP	FN	Sens.	Spec.	lnDOR	Prev.%
Bjorgvinsson	1991	Child	22	34	35	0	1.00	0.493	3.777	24.2
Jakobsson2	1992	Child	20	18	5	63	0.241	0.783	0.086	78.3
Kass	1992	Child	1	6	4	35	0.0278	0.600	-2.813	78.3
Hansen	1995	Child	6	70	4	10	0.375	0.946	2.272	17.8
Merrick 2	1995	Child	286	1570	48	162	0.638	0.970	4.045	21.7
Morrin	1999	Child	58	5	3	4	0.935	0.625	3.017	88.6
LeQuesne	1986	Kidney	22	34	2	8	0.733	0.944	3.598	45.5
Jakobsson1	1992	Kidney	61	23	13	47	0.565	0.639	0.811	75.0
Melis	1992	Kidney	32	153	0	105	0.234	1.000	4.550	47.2
Benador	1994	Kidney	32	30	7	42	0.432	0.811	1.135	66.7
Scherz	1994	Kidney	12	55	10	6	0.667	0.846	2.318	21.7
Lavocat	1997	Kidney	28	54	0	28	0.500	1.000	4.691	50.9
Barry	1998	Kidney	147	467	11	23	0.865	0.977	5.542	26.2

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity	Q (P)	Mean DOR (95% CI)	Median DOR (IQR)
6	child	2469	0.241-1.000	0.493-0.970	78.43 (P<0.001)	22.01 (-2.71-46.74)	15.06 (46.21)
7	kidney	1452	0.234-0.865	0.625-1.000	95.63 (P<0.001)	72.97 (-1.55-157.51)	36.53 (105.85)

Figure 7.1 Forest plot of **ultrasound** for the diagnosis of renal parenchymal abnormality (children)

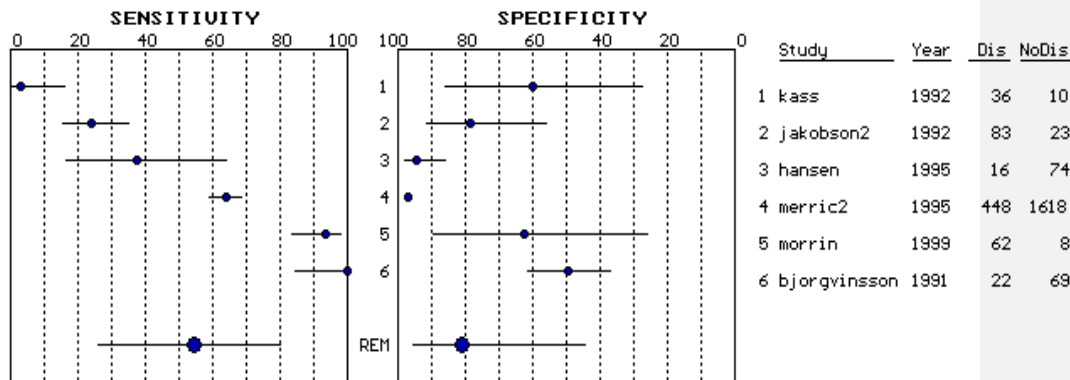


Figure 7.2 Summary receiver operator characteristic curve of **ultrasound** for the diagnosis of renal parenchymal abnormality (children)

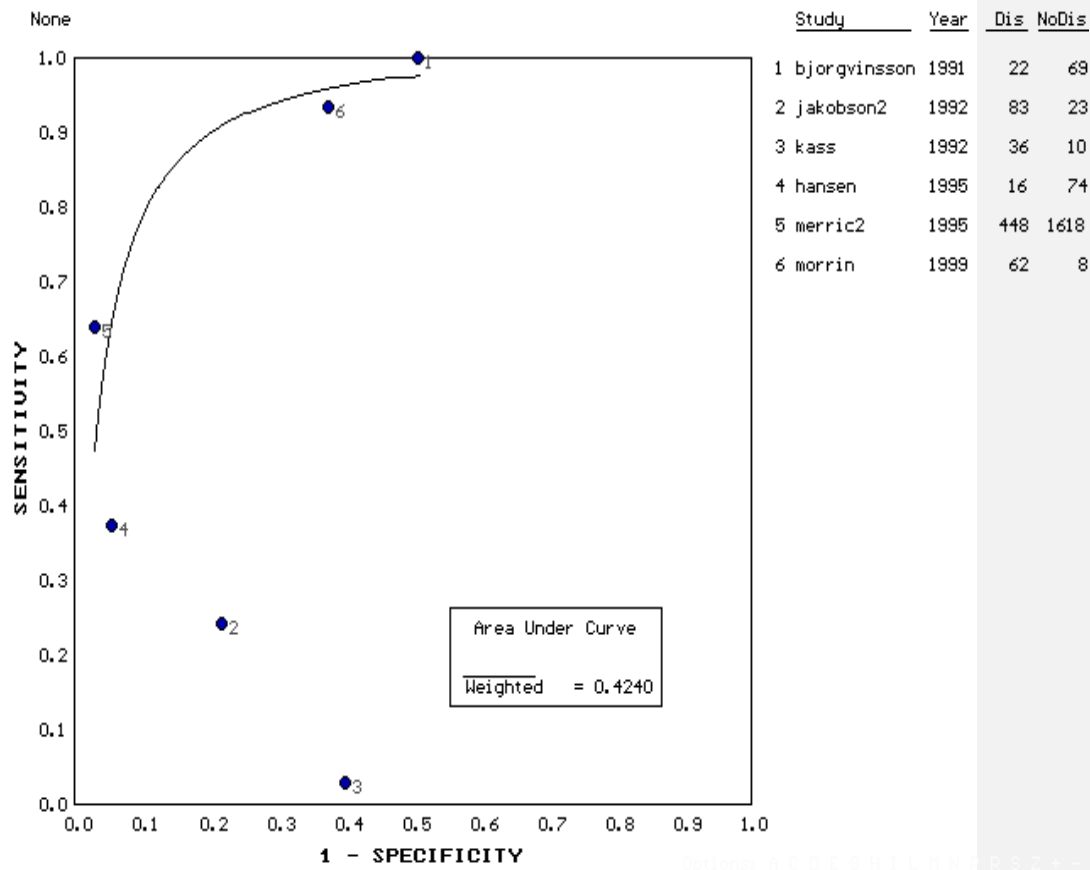


Figure 7.3 Forrest plot of **ultrasound** for the diagnosis of renal parenchymal abnormality (units kidneys)

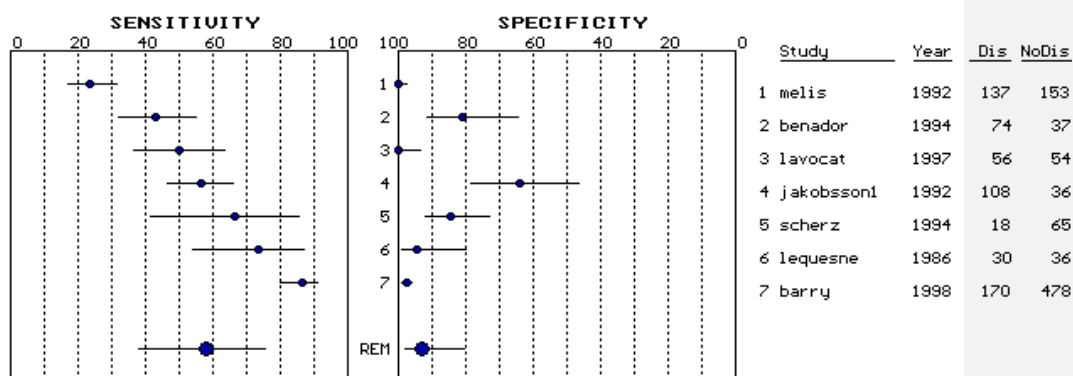
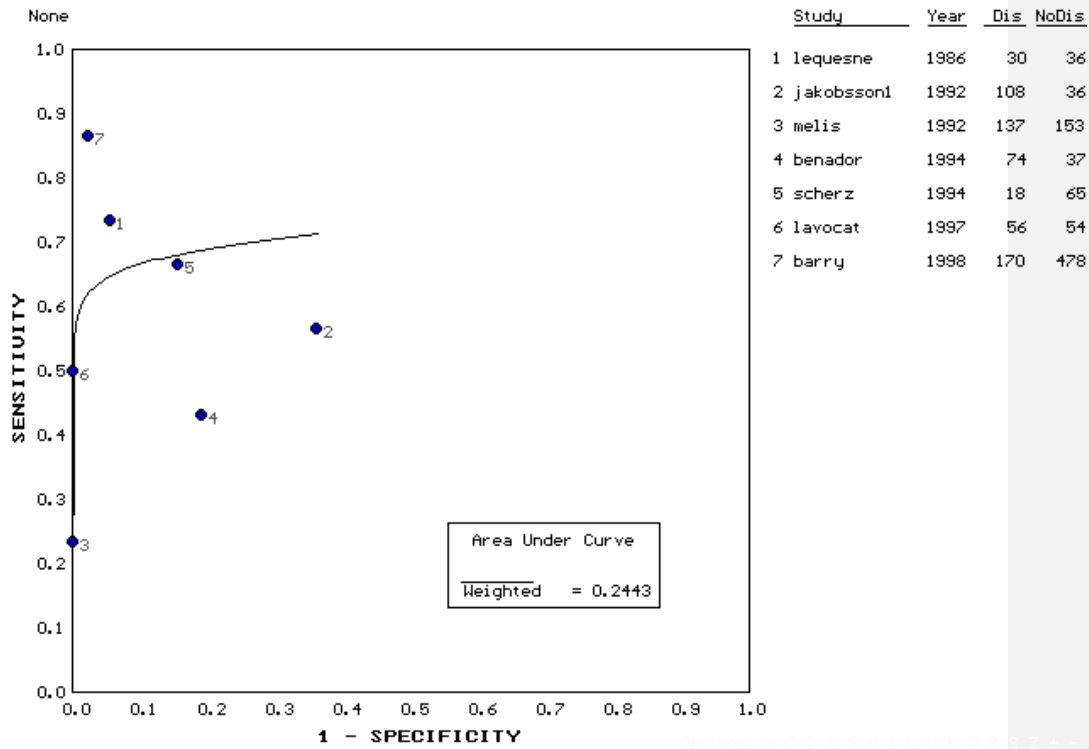


Figure 7.4 Summary receiver operator characteristic curve of ultrasound for the diagnosis of renal parenchymal abnormality (kidneys)



Comparisons that had less than 5 studies and were therefore not included in the

meta-analysis

Outcome renal parenchymal abnormality

Table 3.11 Sensitivity, specificity and diagnostic odds ratio for 2 test comparisons.

Author	Year	Ref Test vs Comparison test	Units	TP	TN	FP	FN	Sens.	Spec.	LnD OR	Prev. %
Hedman	1978	MCUr vs DTPA	Kidney	13	77	4	8	0.619	0.951	3.309	20.6
Leonidas	1983	IVP vs 5 min IVP	Child	9	42	4	7	0.563	0.913	11.96	25.8
Joseph	1990	Planar DMSA vs spect DMSA,	Kidney	14,	43,	3,	10,	0.583	0.935	2.843,	34.3
		Planar DMSA vs Pinhole DMSA,	Kidney	14,	29,	10,	3,	0.824	0.744	2.454	30.4
		Spect DMSA vs Pinhole	Kidney	21	43	3	3	0.875	0.935	4.335	34.3
Jakobsson2	1992	DMSA vs C reactive protein	Child	77	13	10	6	0.928	0.565	2.798	78.3
Pickworth	1992	DMSA vs Mag3,	Kidney	14,	36,	3,	2,	0.875	0.923	4.102	29.1
		IVP vs Mag3,	Kidney	13,	62,	0,	9,	0.591	1.000	15.180	26.2
		DMSA vs P.fim E.coli	Child	30	20	3	17	0.638	0.870	2.232	67.1
Sreenarasim~	1995	GHS vs US	Kidney	36	21	7	32	0.529	0.750	1.169	70.8
Lavocat	1997	DMSA vs CT	Kidney	14	11	0	11	0.560	1.000	3.367	69.4

Summary

No.	Units	Total	Range of Sensitivity	Range of Specificity
3	child	238	0.563-0.928	0.565-1.000
5	kidney	359	0.529-0.875	0.744-1.000

Outcome Obstruction

Table 3.12 Sensitivity, specificity and diagnostic odds ratio for tests used for examining renal obstruction with the unit of measure as kidneys.

Author	Year	Ref Test vs Comparison test	Units	TP	TN	FP	FN	Sens	Spec.	lnDOR	Prev.%
Obstruction Kass	1985	Renography vs diuretic reogram	Kidney	29,	6,	2,	1,	0.967	0.75	3.934	0.789
		Renog. Vs Whitaker pressure perfusion		26	10	1	2	0.929	0.909	4.307	0.178

Table 4.

Average false positive and associated true positive rates, plus abnormality detection rates at 10%, 30% and 60% prevalence levels.

Test	Ref Stand.	Units	Abn	Av. FPR	TPR at av. FPR	100 cases abn /1000 cases UTI		300 cases abn /1000 cases UTI		600 cases abn /1000 cases UTI	
						TP	FP	TP	FP	TP	FP
IVP	MCU	Kidney	VUR	0.22	0.49	49	198	147	154	294	88
IVP	MCU	Child	VUR	0.10	0.40	40	90	120	70	240	40
US	MCU	Kidney	VUR	0.10	0.34	34	90	102	70	204	40
US	MCU	Child	VUR	0.28	0.28	28	252	84	196	168	112
DMSA	MCU	Kidney	VUR	0.25	0.51	51	225	153	175	306	100
DMSA	MCU	Child	VUR	0.58	0.84	84	522	252	406	504	232
US	IVP	Child	RPA	0.06	0.59	59	54	177	42	354	24
IVP	DMSA	Kidney	RPA	0.05	0.54	54	45	162	35	324	20
MCU	DMSA	Kidney	RPA	0.27	0.56	56	243	168	189	336	108
MCU	DMSA	Child	RPA	0.18	0.38	38	162	114	126	228	72
US	DMSA	Kidney	RPA	0.07	0.67	67	63	201	49	402	28
US	DMSA	Child	RPA	0.19	0.89	89	171	267	133	534	76

Results of the economic evaluation

Table 5.1a Detection and missed diagnosis rates for vesicoureteric reflux, assuming 1000 children with urinary tract infection. Data from papers using **children** as the units of measure

Preval	Cases	Detected VUR			Missed VUR		
		US	DMSA	MCU*	US	DMSA	MCU
1%	10	3.6	8.3	10	6.4	1.7	0
10%	100	36	83	100	64	17	0
25%	250	90	207.5	250	160	42.5	0

* assumes MCU is a perfect reference standard

Table 5.1b Detection and missed diagnosis rates for vesicoureteric reflux, assuming 1000 children with urinary tract infection. Data from papers using kidneys as the units of measure

Preval	Cases	Detected VUR			Missed VUR		
		US	DMSA	MCU*	US	DMSA	MCU
1%	10	4.7	5.4	10	5.3	4.6	0
10%	100	47	54	100	53	46	0
25%	250	117.5	135	400	132.5	115	0

* assumes MCU is a perfect reference standard

Table 5.2a Detection and missed diagnosis rates for renal parenchymal abnormality, assuming 1000 children with urinary tract infection. Data from papers using children as the units of measure

Preval	Cases	Detected Abnormalities			Missed Abnormalities		
		US	MCU	DMSA*	US	MCU	DMSA
1%	10	5.5	3.5	10	4.5	6.5	0
10%	100	55	35	100	45	65	0
40%	400	220	140	400	180	260	0

* assumes DMSA is a perfect reference standard

Table 5.2b Detection and missed diagnosis rates for renal parenchymal abnormality, assuming 1000 children with urinary tract infection. Data from papers using kidneys as the units of measure

Preval	Cases	Detected Abnormalities			Missed Abnormalities		
		US	MCU	DMSA*	US	MCU	DMSA
1%	10	5.8	5.7	10	4.2	4.3	0
10%	100	58	57	100	42	43	0
40%	400	232	228	400	168	172	0

* assumes DMSA is a perfect reference standard

Table 6.1a Total costs, costs per detected case of vesicoureteric reflux and false positive detection of vesicoureteric reflux for renal tract imaging tests, assuming 100 children with urinary tract infection and 25% prevalence of vesicoureteric reflux. Data from papers using children as the units of measure

Test	Total cost (1000 tests)	Detected VUR	Cost per VUR case	False positive case
US	\$99 900	90	\$1 110.00	210
DMSA	\$375 450	207.5	\$1 809.40	580
MCU	\$130 300	250	\$ 521.20	0

Table 6.1b Total costs, costs per detected case of vesicoureteric reflux and false positive detection of vesicoureteric reflux for renal tract imaging tests, assuming 100 children with urinary tract infection and 25% prevalence of vesicoureteric reflux. Data from papers using kidneys as the units of measure

Test	Total cost (1000 tests)	Detected VUR	Cost per VUR case	False positive case
US	\$99 900	117.5	\$ 850.21	70
DMSA	\$375 450	135	\$2 781.11	250
MCU	\$130 300	250	\$ 521.20	0

Table 6.2c Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests, assuming 1000 children with Urinary tract infection and 40% prevalence of renal parenchymal abnormality (as would be expected with early testing practices). Data from papers using children as the units of measure

Test	Total cost (1000 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$99 900	220	\$ 454.73	190
MCU	\$130 300	140	\$ 930.71	180
DMSA	\$375 450	400	\$ 938.63	0

Table 6.2d Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests, assuming 1000 children with urinary tract infection and 40% prevalence of renal parenchymal abnormality (as would be expected with early testing practices). Data from papers using kidneys as the units of measure

Test	Total cost (1000 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$99 900	232	\$ 430.60	70
MCU	\$130 300	228	\$ 571.49	270
DMSA	\$375 450	400	\$ 521.20	0

Table 6.2e Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests. Assuming 1000 children with UTI, 400 had renal parenchymal abnormality detected early and so are re-tested and 20% prevalence of renal parenchymal abnormality (as would be expected with mid testing practices, 6-12 months post UTI). Data from papers using children as the units of measure.

Test	Total cost (400 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$ 39 960	110	\$363.27	76
MCU	\$ 52 120	70	\$744.57	72
DMSA	\$150 180	200	\$ 750.90	0

Table 6.2f Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests. Assuming 1000 children with UTI, 400 had renal parenchymal abnormality detected early and so are re-tested and 20% prevalence of renal parenchymal abnormality (as would be expected with mid testing practices, 6-12 months post UTI). Data from papers using kidneys as the units of measure

Test	Total cost (400 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$ 39 960	116	\$ 344.48	28
MCU	\$ 52 120	114	\$ 457.19	108
DMSA	\$150 180	200	\$ 750.90	0

Table 6.2g Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests. Assuming 1000 children with UTI, 400 had renal parenchymal abnormality detected early and so are re-tested and 5% prevalence of renal parenchymal abnormality (as would be expected with late testing practices, 12-24 months post UTI). Data from papers using children as the units of measure.

Test	Total cost (400 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$ 39 960	27.5	\$ 1453.09	76
MCU	\$ 52 120	14	\$ 3722.86	72
DMSA	\$150 180	50	\$ 3003.60	0

Table 6.2h Total costs, costs per detected case of renal parenchymal abnormality and false positive detection of renal parenchymal abnormality for renal tract imaging tests. Assuming 1000 children with UTI, 400 had renal parenchymal abnormality detected early and so are re-tested and 5% prevalence of renal parenchymal abnormality (as would be expected with late testing practices, 12-24 months post UTI). Data from papers using kidneys as the units of measure

Test	Total cost (400 tests)	Detected RPA	Cost per RPA case	False positive case
US	\$ 39 960	29	\$ 1377.93	28
MCU	\$ 52 120	28.5	\$1 828.77	108
DMSA	\$150 180	50	\$ 3003.60	0

Table 7. Summary table of rates of detection and missed cases of vesicoureteric reflux, rates of detection and missed diagnosis of renal parenchymal abnormality, costs per detected case of Vesicoureteric reflux and renal parenchymal abnormality for each commonly used renal tract imaging test. Data uses children as the units of measure.

Test	% VUR detected	% Missed VUR	%Renal parench. abnormality	% missed RPA	Cost per VUR case detected	Cost per RPA case detected
US	36	64	55	45	\$277.50	\$181.64
MCU	100	0	35	65	\$130.30	\$372.29
DMSA	83	17	100	0	\$452.35	\$375.45

References

1. Craig JC, Irwig LM, Knight JF, Sureshkumar P, Roy LP. Symptomatic urinary tract infection in preschool Australian children. *Journal of Paediatrics and Child Health* 1998, 34(2):154-159.
2. Downs SM. Technical report: urinary tract infections in febrile infants and young children. *American Academy of Pediatrics* 1999, 103:1-60
3. Haldane JBS. The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics* 1955; 20:309-314
4. Hellstrom A, Hanson E, Hansson S, Hjalmas K, Jodal U. Association between urinary symptoms at 7 years old and previous urinary tract infection. *Archives of Diseases in Childhood* 1991, 66(2):232-234.
5. Lijmer JG, Mol BW, Heisterkamp A, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests *JAMA* 1999;282 (11): 1061-1066.
6. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. *Medical Decision Making* 1993;13:313-321

Appendix 1. Literature search strategy

#	Search History
1	exp kidney failure, chronic
2	Exp dialysis/ or exp peritoneal dialysis/ or exp peritoneal dialysis, continuous ambulatory
3	“END STAGE RENAL DISEASE”.mp
4	exp Ureteral obstruction/
5	exp Urethral obstruction/
6	“URETEROVESICAL JUNCTION OBSTRUCTION”.mp
7	“URETEROPELVIC JUNCTION OBSTRUCTION”.mp
8	“OBSTRUCTIVE UROPATHY”.mp
9	“POSTERIOR URETHRAL VALVE OBSTRUCTION”.mp
10	Posterior urethral valves.tw
12	exp bladder neck obstruction/ or “bladder outlet obstruction”.mp
12	“RENAL SCARRING”.mp
13	“REFLUX SCARRING”.mp
14	exp Vesico-ureteral reflux/ and scarring.tw
15	Vesico-ureteral reflux.tw
16	“RENAL PARENCHYMAL ABNORMALITY”.mp
17	(kidney adj25 scarring).mp
18	Or /1-17
19	exp urinary tract infections/ or “urinary tract infection”.mp
20	18 and 19
21	Exp diagnosis/
22	Exp radiography/
23	Exp radionuclide imaging/

24	Exp ultrasonography/
25	Sensitivity.tw
26	Or /21-25
27	20 and 26
28	Limit to english
29	Limit 27 to (adult<19 to 44 years> or middle age <45 to 64 years or “aged <65 and over>” or “aged <80 and over>”)
30	27 not 28