US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis[1]

Andrea S. Doria, MD, MSc, PhD
Rahim Moineddin, PhD
Christian J. Kellenberger, MD
Monica Epelman, MD
Joseph Beyene, PhD
Suzanne Schuh, MD
Paul S. Babyn, MD
Paul T. Dick, MD, MSc

Purpose: To perform a meta-analysis to evaluate the diagnostic performance of ultrasonography (US) and computed tomography (CT) for the diagnosis of appendicitis in pediatric and adult populations.

Materials and Methods: Medical literature (from 1986 to 2004) was searched for articles on studies that used US, CT, or both as diagnostic tests for appendicitis in children (26 studies, 9356 patients) or adults (31 studies, 4341 patients). Prospective and retrospective studies were included if they separately reported the rate of true-positive, true-negative, false-positive, and false-negative diagnoses of appendicitis from US and CT findings compared with the positive and negative rates of appendicitis at surgery or follow-up. Clinical variables, technical factors, and test performance were extracted. Three readers assessed the quality of studies.

Results: Pooled sensitivity and specificity for diagnosis of appendicitis in children were 88% (95% confidence interval [CI]: 86%, 90%) and 94% (95% CI: 92%, 95%), respectively, for US studies and 94% (95% CI: 92%, 97%) and 95% (95% CI: 94%, 97%), respectively, for CT studies. Pooled sensitivity and specificity for diagnosis in adults were 83% (95% CI: 78%, 87%) and 93% (95% CI: 90%, 96%), respectively, for US studies and 94% (95% CI: 92%, 95%) and 94% (95% CI: 94%, 96%), respectively, for CT studies.

Conclusion: From the diagnostic performance perspective, CT had a significantly higher sensitivity than did US in studies of children and adults; from the safety perspective, however, one should consider the radiation associated with CT, especially in children.

Supplemental material:
radiology.rsnajnls.org/cgi/content/full/2411050913/DC1

[1] From the Departments of Diagnostic Imaging (A.S.D., C.J.K., M.E., P.S.B.), Population Health Sciences (J.B.), and Paediatrics (S.S., P.T.D.), the Hospital for Sick Children; Department of Public Health, Family and Community Medicine (R.M.), and Department of Health Policy Management and Evaluation (P.T.D.), Faculty of Medicine, University of Toronto, 555 University Ave, Toronto, ON, Canada M5G 1X8. Received June 1, 2005; revision requested July 21; revision received September 22; accepted October 18; final version accepted January 2, 2006. A.S.D. supported in part by a RESTRACOMP award from the Research Institute, the Hospital for Sick Children, Toronto, Ontario, Canada. **Address correspondence to** A.S.D. (e-mail: *andrea.doria@sickkids.ca*).

Radiology

Appendicitis is the most common acute condition of the abdomen requiring surgery in both adults (1,2) and children (3–5). The overall frequency of appendicitis for symptomatic patients younger than 20 years is 41%; the frequency for those older than 20 years is 59% (6,7). Imaging methods, such as ultrasonography (US) and computed tomography (CT), aimed at avoiding a misdiagnosis and facilitating earlier surgery, when necessary, have become increasingly important for decreasing the morbidity of the disease.

Advantages of US include low cost, the lack of ionizing radiation or need for patient preparation, and the ability to provide dynamic information through graded compression (8). Advantages of CT include less operator dependency than US; enhanced delineation of the extent of the disease in the case of perforated appendicitis (9,10); easier visualization of a retrocecal appendix; unchanged quality of imaging, regardless of the presence of bowel gas, obesity, or severe abdominal pain (11); and the possibility of multiplanar retrospective data reconstruction (12). Patients with abundant adipose tissue, who more frequently are adults rather than children, are prime candidates for CT; their bowels are more difficult to probe with a US transducer because of the thicker physical barrier and lesser compressibility of their abdomens.

The choice between US and CT for the diagnosis of appendicitis in different age groups should rest on the comparison of their diagnostic accuracy based on summary estimates of available experience. To our knowledge, no previous meta-analysis has been performed to evaluate the choice between CT and US for the diagnosis of appendicitis in both children and adults. The purpose of our study, therefore, was to perform a meta-analysis to evaluate the diagnostic performance of US and CT for the diagnosis of appendicitis in pediatric and adult populations.

Advances in Knowledge

- In contrast to specificity, the difference in pooled sensitivities and odds ratios between US and CT is significant in studies of children ($P = .001$) and adults ($P = .001$).
- CT has a significantly higher sensitivity than does US for the diagnosis of appendicitis in adults and children; we should note, however, that the sensitivity of US is reasonably high in children and that radiation issues are of special concern in this age group.

Materials and Methods

Data Extraction

Two reviewers (A.S.D., C.J.K.) independently conducted a literature review to identify articles published between January 1986 and December 2004 on studies that used US, CT, or both as diagnostic tests for appendicitis in children and/or adults. Articles were obtained through a search of MEDLINE, EMBASE, CINAHL, the Cochrane Controlled Trials Register, the Cochrane Database of Systematic Reviews, and the American College of Physicians Journal Club databases and through a manual search of reference lists. Pertinent medical subject headings ("appendicitis," "appendix," "sonography," "ultrasonography," "computed tomography," and "computed tomography scan") were used. To check for eligibility criteria, we retrieved the study articles for more detailed investigation.

Study Selection

Criteria for retaining articles were as follows: *(a)* prospective or retrospective studies evaluating the performance of abdominal US and/or CT that used either surgery findings plus follow-up (longer than 1 week) or histologic evaluation plus follow-up as reference standards; *(b)* availability of data for the absolute number of true-positive, true-negative, false-positive, and false-negative findings either reported, derivable from the results, or communicated by the authors in response to our request; *(c)* segmentation of results according to age groups, with a maximum age of 20 years for children and young adults and a minimum age of 13 years for adults, and if this criterion was not fully met, the proportion of patients with outlying ages could not exceed 5% of the total sample size; *(d)* imaging criteria for positivity for appendicitis that included visualization of an inflamed appendix (diameter, >6 mm) (13–15), noncompressible appendix at US (16), or, in the case of nonvisualization of the appendix, presence of inflammatory signs of appendicitis, such as an appendicolith, cecal thickening, arrowhead sign, or cecal bar (as seen on CT images) (17–21); *(e)* in studies evaluating the performance of CT scanning, a description of the technique used—namely, the use of oral, rectal, and intravenous contrast material with a limited or complete scan; *(f)* inclusion of both female and male patients (ie, ratio of one sex to the other, <3:1); and *(g)* studies with a 15%–75% sample prevalence of appendicitis derivable from the reported results (ie, true-positive plus false-negative results, divided by the total of true-positive, true-negative, false-positive, and false-negative results), as arbitrarily determined and checked by means of sensitivity analysis. No language restriction was applied.

Exclusion criteria were as follows: *(a)* case reports, case series, reviews, pictorial essays, unpublished data, abstracts, and letters to editor; *(b)* data for pregnant women; *(c)* focus on topics other than diagnostic test assessment, such as management decision issues or cost-effectiveness analyses; and *(d)* performance of more than one US or CT examination per patient in subsequent

Published online before print
10.1148/radiol.2411050913

Radiology 2006; 241:83–94

Abbreviations:
CI = confidence interval
DOR = diagnostic OR
OR = odds ratio
ROC = receiver operating characteristic

Author contributions:
Guarantors of integrity of entire study, A.S.D., R.M.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, A.S.D., C.J.K., M.E.; statistical analysis, A.S.D., R.M., J.B.; and manuscript editing, all authors

Authors stated no financial relationship to disclose.

days. We avoided duplication of data in the studies included in this meta-analysis. When more than one study used the same data or when the durations of studies overlapped, the study with the larger sample size was selected.

If the results of diagnostic test performance (sensitivity, specificity, accuracy, positive predictive value, and negative predictive value) did not match the reported values of true-positive, true-negative, false-positive, and false-negative results, we contacted the authors of the articles. If we received a reply, we reported the revised results; otherwise, we reported the results published in the original articles. We also contacted authors of articles that fulfilled the inclusion criteria for the meta-analysis except for the segmentation of the results by age group (written communication from David Mauger [22], Johan Styrud [23], Sang Hoon Cha [24], Miguel V. Ronco [25], and Blayne Standage [26]). The articles by those authors who provided the required information within our time frame were incorporated into our database. Only results from unenhanced gray-scale US were included in the meta-analysis.

Critical Appraisal of the Quality of Studies

The quality of methods was independently assessed by three radiologists by using a specific checklist (27–29) that yielded a total score for each study evaluated. Two radiologists (M.E., C.J.K.) were blinded to the journal names, author names, and year of publication; the third radiologist (A.S.D.) was unblinded. Detailed guidelines for interpretation of the items were discussed by the reviewers in an earlier meeting.

Statistical Analysis

Before defining the final inclusion criteria for our meta-analysis, we grouped positivity diagnostic criteria that resulted in a diagnosis of appendicitis according to three uncertainties for sensitivity analysis: *(a)* whether surgery alone or surgery plus clinical follow-up had been considered as the reference standards, *(b)* whether patients had undergone both US and CT or had undergone exclusively US or CT, and *(c)* whether the equivocal cases were handled as true-positive (best-case scenario) or as false-positive or false-negative (worst-case scenario).

Data from studies with these uncertainties were included in the pooled results of studies that used well-accepted criteria (ie, surgery and clinical follow-up as a reference standard, only one imaging modality per patient). If a significant difference was found between results obtained with and those obtained without inclusion of a particular factor of uncertainty in the pooled results, we considered that factor of uncertainty relevant and therefore incorporated it into our exclusion criteria; otherwise, it was not specified. The inclusion of data from studies that used true-positive, false-positive, or false-negative categorization to handle equivocal cases did not change our pooled results significantly; therefore, we decided to use the results as reported by the authors in our data analysis. Likewise, the exclusion of studies in which the prevalence of appendicitis was either less than 25% or greater than 75% did not change the results significantly, which allowed us to set a specific range of sample prevalence as an inclusion criterion.

We calculated the intraclass correlation coefficient (30) for scoring the quality of the studies of children and adults as a measure of the overall agreement among the three reviewers and between the two blinded raters. We also calculated the perforation rate in positive cases of appendicitis in studies of children and adults, weighted for the sample size. Information on the technique used for imaging acquisition was retrieved from the studies, if available.

Pooled point estimates and 95% confidence intervals (CIs) were obtained for the sensitivity and specificity of US or CT for the diagnosis of appendicitis in studies of children and adults after natural logarithmic transformation. Data were derived from the numbers of true-positive, true-negative, false-positive, and false-negative findings given in the publications. Pooled results were weighted according to the quality score for the study with a fixed-effects model (31,32). As an initial step, we applied the random-effects model to our meta-analysis to estimate the variance of τ^2, which is a statistical parameter that determines the study-to-study variation (33). If τ^2 is a value close to zero, as shown in our study, the results of a random-effects model and those of a fixed-effects model are similar. Pooled results weighted according to the quality score for the study are presented in this article with a fixed-effects model (31,32). We were able to calculate the pooled sensitivity and specificity statistics by using the Mantel-Haenszel fixed-effects model for US and CT studies of children and adults because no statistical evidence of heterogeneity was noted (Table E1, *radiology.rsnajnls.org/cgi/content/full/2411050913/DC1*). Results were evaluated by including outliers and by excluding outliers. We a priori defined that if, by the time outliers were included in the analysis, the results of the Mantel-Haenszel test of homogeneity (34) became significant (indicating heterogeneity), the outliers would be excluded from the pooled results.

We performed a one-way sensitivity analysis of three studies that seemed to be outliers for US on the basis of their recorded sensitivity in studies of children (35) and adults (22,36). When two or more CT techniques were performed for the same patient on the same examination day (22), we selected the most conservative approach available, such as CT scanning of the entire abdomen and pelvis with administration of intravenous and oral contrast material (37,38), for the analysis.

Although pooled point estimates are simpler summaries of diagnostic accuracy for clinicians (39,40) in comparison with diagnostic odds ratios (DORs) and summary receiver operating characteristic (ROC) curves, they are not recommended to summarize results when the diagnostic threshold varies between the studies, which is the case for appendicitis studies that used different thresholds to define positive and negative test results. Because of this limitation of the method, we also summarized the results of the studies by using best-fitting ROC curves. By doing so, we tried to ensure that apparent threshold effects, if present, were really due to variations in diagnostic threshold rather than alternative sources of heterogeneity.

Meta–regression analyses were performed to examine the effect of three confounding factors (covariates) on the sensitivity and specificity of US and CT in studies of children and adults: study design (retrospective vs prospective), year of publication (before 1999 vs 1999 or later), and continent of study origin (North American vs non–North American country). The weights assigned to individual studies were in inverse proportion to their variance. The odds ratio (OR) is the ratio of the odds of disease in those with the risk factor (in our case, clinical suspicion of appendicitis) to the odds of disease in those without the risk factor (41). The OR derived from logistic-regression modeling was used to determine the effect of covariates as a source of heterogeneity for the sensitivity and specificity of the studies as they were, one by one, included in the model. Both the level of significance and the magnitude of the effect were assessed.

We conducted a visual examination of the funnel plots of sensitivity and specificity of studies to investigate the potential for publication bias and overall heterogeneity (42–44). We plotted the scoring of the quality of the US and CT studies of children and adults as a function of the reported sensitivity and specificity of the corresponding diagnostic tests.

Reported P values for all statistical tests were two-sided. We used $P < .05$ to indicate a significant difference. Analyses and graphic illustration were performed with a statistical software package (SAS version 8.2; SAS Institute, Cary, NC).

Results

Overview of Studies

Of the 229 articles screened, 57 met the inclusion criteria for the meta-analysis. Summary results and selected characteristics of the studies (22–26,35,36,38,45–92) included in the meta-analysis are outlined in Table E2 (*radiology.rsnajnls.org/cgi/content/full/2411050913/DC1*). Pediatric and adults studies were analyzed separately. In 26 studies of children (mean age range, 7–12 years) published between 1990 and 2004, only US, only CT, or combined US and CT results were reported for 6850, 598, and 1908 patients, respectively. Thirty-one studies of adults (mean age range, 20–49 years) published between 1988 and 2003 yielded results for only US, only CT, or combined US and CT in 903, 2394, and 1044 patients, respectively. The mean sample prevalence of appendicitis calculated on the basis of the data provided by the articles was 0.31 for both US and CT in pediatric studies and was 0.48 for US and 0.40 for CT in studies of adults. The weighted perforation rate in positive cases of appendicitis was 26.5% in studies of children (n = 10) and 18.5% in studies in adults (n = 3). The relative risk of a false-negative US finding in perforated appendicitis rather than nonperforated appendicitis was 0.34 in pediatric studies. Insufficient data were available for calculation of relative risk in studies of adults.

With the exception of one study of children (46) and one study of adults (36), all other studies used US linear probes with frequencies of 5 MHz or greater with US units from different companies. Except for one study of children (58) and one study of adults (23), all remaining articles reported the use of third-generation helical CT units. CT section collimation parameters ranged from 3.0 to 7.0 mm in studies of children and from 2.5 to 10.0 mm in studies of adults. The pitch level ranged from 1.0 to 1.5 in most studies. The rate of injection ranged from 2 to 3 mL/sec, and the amperage ranged from 200 to 320 mA in studies of adults. Information about the radiation dose was unavailable in most studies of children.

Critical Appraisal of Quality of Studies

Results of assessment of study quality can be found in the Table. The median score for the studies of children was 34.4% (maximum score, 32 points or 100%), and the median score for the studies of adults was 42.2%. The interrater agreement between the one unblinded reviewer and the two blinded reviewers (intraclass correlation coefficient, 0.78; 95% CI: 0.64, 0.87) and between the two blinded readers (intraclass correlation coefficient, 0.70; 95% CI: 0.46, 0.84) was good.

Pooling Sensitivities and Specificities

The pooled sensitivity of CT (94%; 95% CI: 92%, 97%) was 6% higher than that of US (88%; 95% CI: 86%, 90%) among studies of children with one outlier removed (P = .001) (Fig 1). The difference in pooled sensitivities between CT (94%; 95% CI: 92%, 95%) and US (83%; 95% CI: 78%, 87%) was 11% in studies of adults with two outliers removed (P = .001). Weighted pooled data for specificity (1 − false-positive rate) yielded similar values for US (for children, 94% [95% CI: 92%, 95%]; for adults, 93% [95% CI: 90%, 96%]) and for CT (for children, 95% [95% CI: 94%, 97%]; for adults, 94% [95% CI: 94%, 96%]), regardless of the age group being investigated (P = .26 for either age group of studies). The omission of two studies of adults (22,36) and of one study of children (34) from the pool of US studies yielded a considerable improvement in the weighted pooled sensitivity of studies of adults and children (from 78% to 83% for studies of adults and from 86% to 88% for studies of children). Therefore, we removed from the analysis the outliers for US in the series of adults and children.

Assuming the prevalence of appendicitis as 0.15, the minimum prevalence in a study of this meta-analysis, the number of cases of appendicitis missed by using US rather than CT would be 10 cases per 1000 children imaged and 18 cases per 1000 adults imaged. Conversely, if we had an observed prevalence of 0.75, the maximum prevalence in a study of this meta-analysis, the number of cases of appendicitis missed by using US would be greater (48 per 1000 children imaged and 83 per 1000 adults imaged).

Diagnostic Performance Comparison with ROC Plots and Linear Regression Analysis

Diagnostic tests in which the DOR is constant, regardless of the diagnostic threshold, have ROC curves or point estimates that are symmetric around the line where sensitivity equals specificity. In our meta-analysis, most point estimates of the diagnostic tests of studies of children (Fig 2a) and adults (Fig 2b) tended to the symmetry around the "sensitivity = specificity" line, which enabled the use of standard meta-analysis methods to estimate the common DOR (93–95).

By computing a summary DOR for pooling ORs on the basis of the best-fitting ROC curve, we obtained estimates of 202 (95% CI: 159, 258), 239 (95% CI: 118, 487), and 46 (95% CI: 32, 67) for US only, CT only, and combined modalities, respectively, in studies of children; we obtained estimates of 15 (95% CI: 10, 21), 118 (95% CI: 85, 165), and 100 (95% CI: 57, 176) for US only, CT only, and combined modalities, respectively, in studies of adults.

In the regression models for sensitivity, CT had better diagnostic sensitivity than did US in studies of adults, as demonstrated by the significant OR for differences in sensitivity in these studies (OR = 3.1; 95% CI: 1.70, 5.69; $P <$ .001). The OR for differences in sensitivity in studies of children was 2.47 (95% CI: 1.16, 5.26; P = .02). No differences in the diagnostic performance of the imaging modalities were noted with regard to specificity of studies of any of the age groups (studies of children, OR = 0.77 [95% CI: 0.55, 1.09]; studies of adults, OR = 1.18 [95% CI: 0.61, 2.28]).

Sources of Heterogeneity

Meta–regression analysis failed to show any evidence of heterogeneity related to the potential confounders on comparison of studies by means of dichotomization of the covariates into prospective versus retrospective (study design), before 1999 versus 1999 or later (year of publication), and North America versus non–North America (continent of study origin). The overall ORs were shown not to be different between the covariate arms, regardless of the imaging modality and age group being evaluated.

Comparative Meta-Analysis

We evaluated 18 US and three CT studies of children and 10 US and 16 CT studies of adults that compared the performance of these imaging modalities for diagnosis of appendicitis when only a single modality was assessed in the original studies (Table E3, *radiology.rsnajnls.org/cgi/content/full/2411050913/DC1*). Except for specificity in US studies of adults, in which the variation of pooled estimates between overall studies and studies that evaluated individual imaging modalities

Assessment of the Quality of Studies of Children and Adults

Study Characteristic	No. of Studies Meeting Criterion: Children (n = 26)	Adults (n = 31)
Reference standard		
Valid reference standard (surgery and follow-up)	26 (100)	31 (100)
Well-defined threshold of reference standard (surgery)	3 (12)	2 (6)
Well-defined threshold of test (positive, negative, or equivocal findings)	20 (77)	26 (84)
Independence of assessments		
Surgeons blinded to clinical assessment	0 (0)	0 (0)
Assessment with reference standard independent of index test results (verification bias)	0 (0)	3 (10)
Acquisition or reading of US or CT images, blinded to clinical assessment	3 (12)	10 (32)
Two or more readers for US or CT	3 (12)	7 (23)
Reproducibility of US or CT	2 (8)	8 (26)
Study design		
Prospective	15 (58)	29 (94)
Retrospective	11 (42)	2 (6)
Technique described in detail	23 (88)	28 (90)
Selection of study subjects (random, consecutive, unknown)	12 (46)	21 (68)
CIs reported for point estimates of sensitivity, specificity	5 (19)	2 (6)
ROC curves used to estimate accuracy	2 (8)	0 (0)
Tables for sensitivity, specificity	17 (65)	23 (74)
Point estimates of predictive values or posttest likelihood	15 (58)	22 (71)
If both US and CT used, all cases verified with both	0 (0)	8 (26)
Selection for disease verification		
Random US or CT	0 (0)	0 (0)
Inclusion or exclusion criteria	9 (35)	15 (48)
Excluded sample percentage reported	7 (27)	8 (26)
Setting (community or tertiary-care hospital)	4 (15)	3 (10)
Demographic information		
Age	26 (100)	31 (100)
Sex	21 (81)	26 (84)
Weight or body mass index	0 (0)	0 (0)
Race	0 (0)	0 (0)
Comorbid condition	6 (23)	8 (26)
Duration of symptoms	6 (23)	4 (13)
Information about other diagnostic tests	9 (35)	9 (29)
Spectrum of nondisease	15 (58)	25 (81)
Wide spectrum of disease severity included	17 (65)	22 (71)
Sample-size estimation	1 (4)	0 (0)
Purpose clearly reported	22 (85)	26 (84)
Appropriateness of population for study	19 (73)	27 (87)
Pretest risk of appendicitis (clinical assessment)	5 (19)	1 (3)
Statistical analysis description	10 (38)	10 (32)
Randomization for US only or US and CT examinations	1 (4)	4 (13)

Note.—Data in parentheses are percentages.

Figure 1

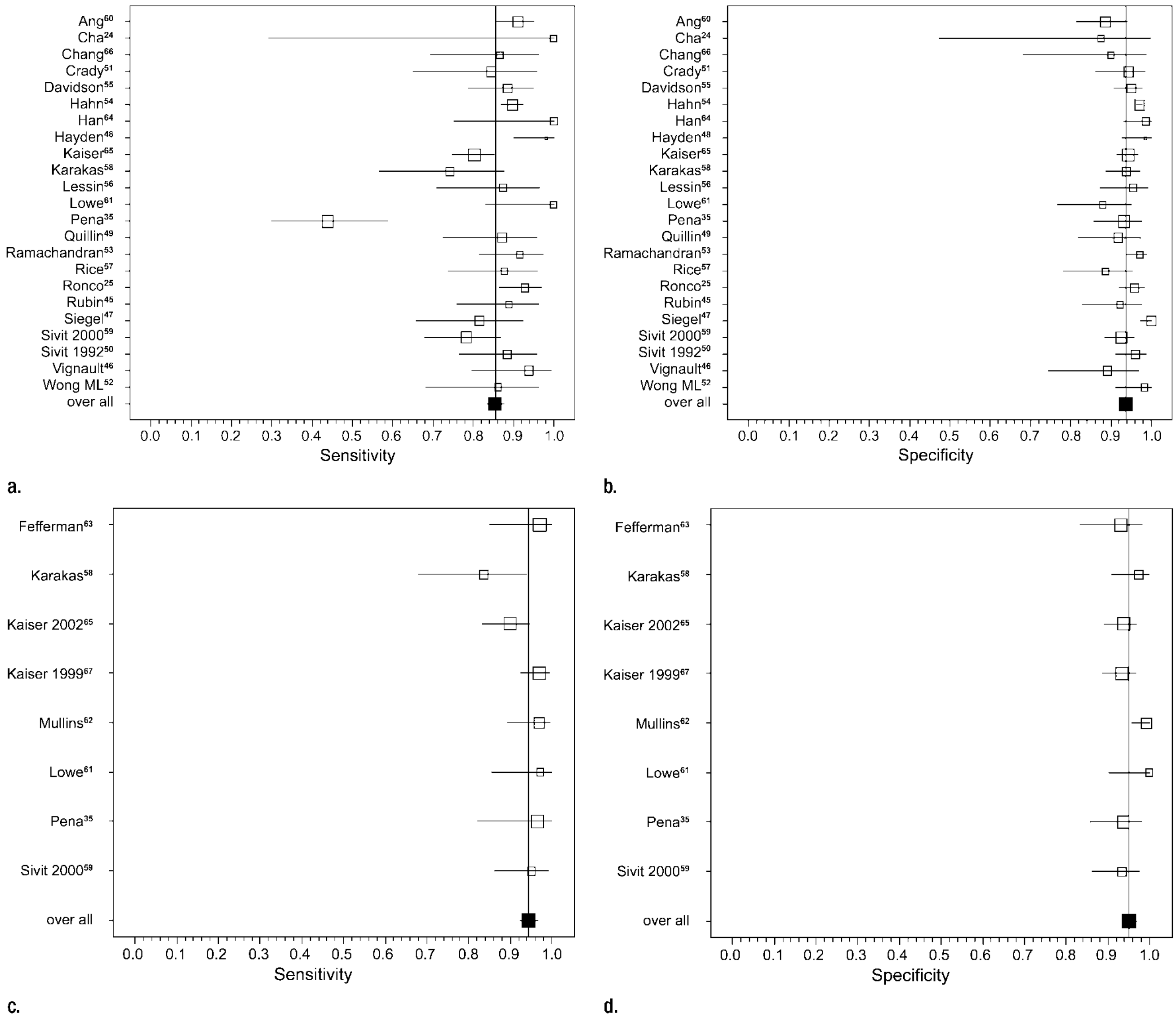

Figure 1: Graphs show sensitivity and specificity recorded in individual series of children and adults. **(a)** Sensitivity of US for children. **(b)** Specificity of US for children. **(c)** Sensitivity of CT for children. **(d)** Specificity of CT for children *(Fig 1 continues)*.

was 5%, all other summary estimates for US and CT were similar between overall studies and studies of individual modalities (maximum range of variation, 2%).

Funnel Plots

The overall distribution of data points associated with the sensitivity and specificity of CT and the specificity of US seemed fairly funnel-shaped and symmetric, which indicates that the presence of publication bias was unlikely in the studies of either children or adults. The data point distribution for the sensitivity of US in the studies of children and adults was also symmetric, except for the presence of one outlier value in each funnel plot.

Discussion

For patients in whom the clinical diagnosis is uncertain and in whom further evaluation with imaging is required, our analysis shows that the differences in weighted pooled sensitivities between CT and US and in ORs between CT and US were significant for adults and children, in spite of the lower sensitivity of US for the assessment of appendicitis in the adult group. US was moderately sensitive and highly specific for the diagnosis of appendicitis in children. No significant difference was noted in the specificity of the diagnostic performance between CT and US. This result

Figure 1 (continued)

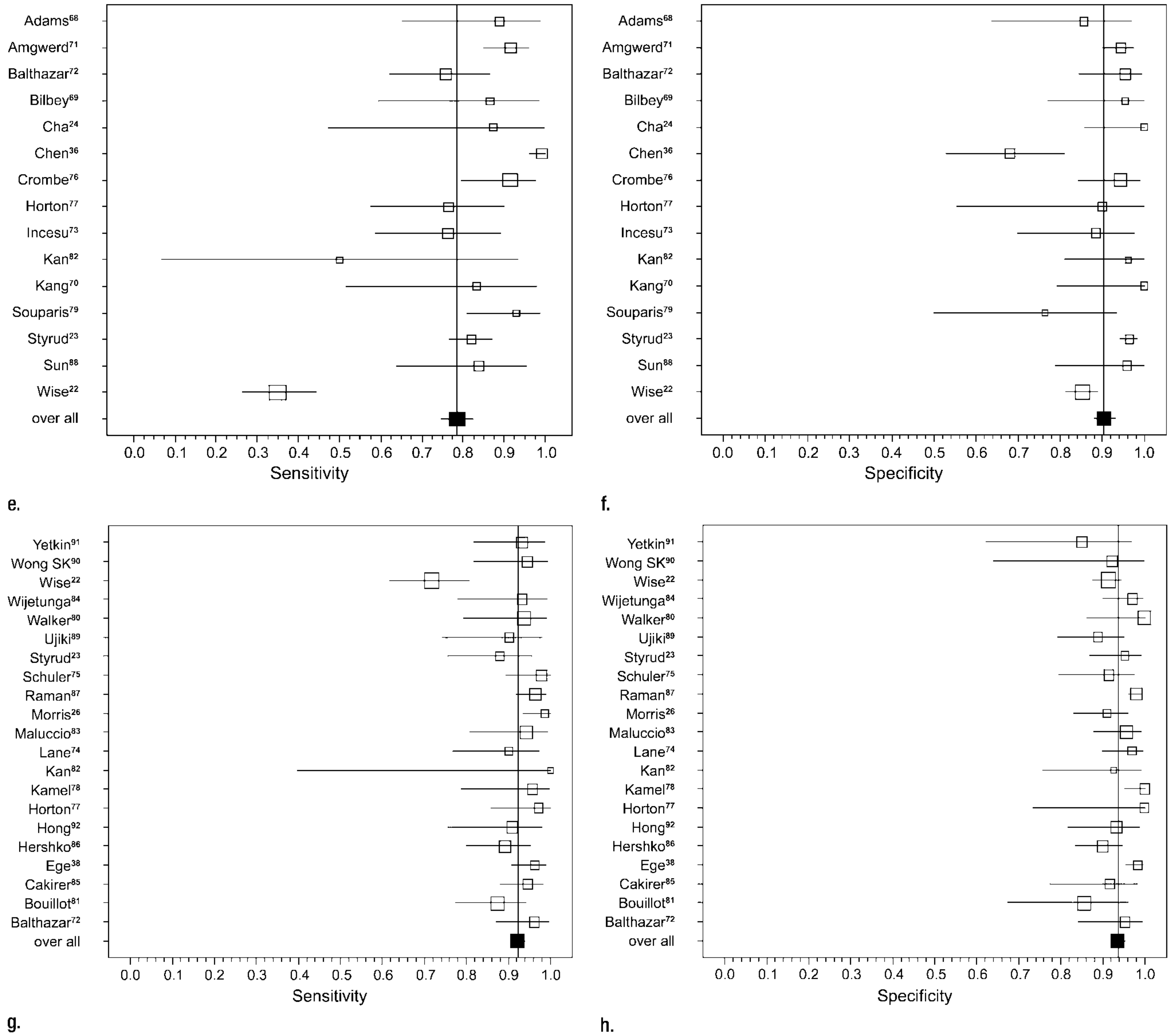

Figure 1: *(continued)* **(e)** Sensitivity of US for adults. **(f)** Specificity of US for adults. **(g)** Sensitivity of CT for adults. **(h)** Specificity of CT for adults. Point estimates (□) and 95% CIs (horizontal lines) are given for each series. The meta-analytic summary estimate is represented by the vertical line. Outliers have not been excluded on these graphs.

indicates that the major advantage of CT over US lies in the decrease in the false-negative rate for CT and the consequent improvement in sensitivity, as demonstrated elsewhere (95). The false-positive rate, however, was relatively unchanged with the use of either CT or US.

Although the sensitivity of US was higher for children than for adults, no significant differences were noted in the specificity of US. Prior meta-analyses that assessed the sensitivity of US for diagnosis of appendicitis in adults, which comprised studies from 1986 to 1994 (96) and studies from 1966 to 2003 (97), showed similar sensitivities (84.7% [96] and 86.0% [97]) in comparison with our results (83.0%). This minimal discrepancy in results may be explained by the different inclusion and exclusion criteria used for the retrieval of studies in the meta-analyses and by the different criteria used for weighted analysis. In addition, our results for CT were similar to those of a prior meta-analysis of appendicitis in adults (97), in which the pooled sensitivity and specificity of CT were 94% and 95%, respectively, compared with our results of 94% for both.

The greater radiosensitivity of the organs and tissues of children in comparison with adults (98) suggests the need for a trade-off between the future risk of cancer with the use of CT and the risk of missing positive cases with US. According to estimates of the lifetime risk of mortality attributable to cancer and the relative doses of radiation from CT, calculated as a function of age by using Monte Carlo methods (99,100), and assuming a hypothetical sample prevalence of 0.31 for diagnosis of appendicitis in children, for every 10 000 children 11 years of age scanned with US rather than with CT, 280 would have a missed diagnosis of appendicitis and 13 could be prevented from developing cancer in the future. On the other hand, if we considered 10 000 adults 35 years of age scanned with US rather than with CT in a center with a sample prevalence of 0.40, the diagnosis of appendicitis would be missed in 480 patients, but only two patients could be prevented from developing cancer in the future. Because of radiation concerns, one way of restricting the use of CT in the population would be to aim its use toward higher prevalence settings. The higher the prevalence of appendicitis in the population, the larger the difference in false-negative rates between US and CT and the lower the number of cases that need to be imaged at CT to detect one additional case missed at US. Because sensitivity and specificity are not confounded by prevalence (101), potential differences in the prevalence of appendicitis among studies should not affect the pooled sensitivity and specificity results of the meta-analysis.

Although the articles in our meta-analysis have reported a worse clinical outcome for appendicitis in pediatric (24,25,35,45–67) than in adult (22–24, 26,36,38,68–92) populations, our results indicate that the risk of a child with perforated appendicitis having a false-negative result with US is 34% of the risk that he or she would be exposed to if he or she had nonperforated appendicitis. Thus, the sensitivity of US was not reduced with the diagnosis of perforated appendicitis, but conversely it seemed to be improved with this condition compared with the sensitivity observed in cases of nonperforated appendicitis in studies of children.

The higher incremental improvement in the sensitivity of CT in relation to US for studies of adults compared with that for studies of children suggests that the within-study heterogeneity might have been higher in studies of adults. Nevertheless, our funnel-plot analysis suggests that publication bias (102) for quality of scoring in studies of children or adults either does not exist or is minimal.

In spite of the recent technologic advances in both US and CT, the ORs obtained in more recent (1999 or later) or older (before 1999) studies were not significantly different in either of the age groups. This may be related to the high threshold used as a cutoff point for studies according to publication year, which was based on the median year of publication of the pool of selected studies. The origin of the study, however, did not seem to play a significant role in the incremental discriminatory ability of CT compared with US in studies of either children or adults.

The critical appraisal of studies revealed important methodologic limitations in the retrieved articles that were related to difficulties in diagnostic test evaluation. The most important of these was the potential for verification and selection bias (102–104). These types of

Figure 2

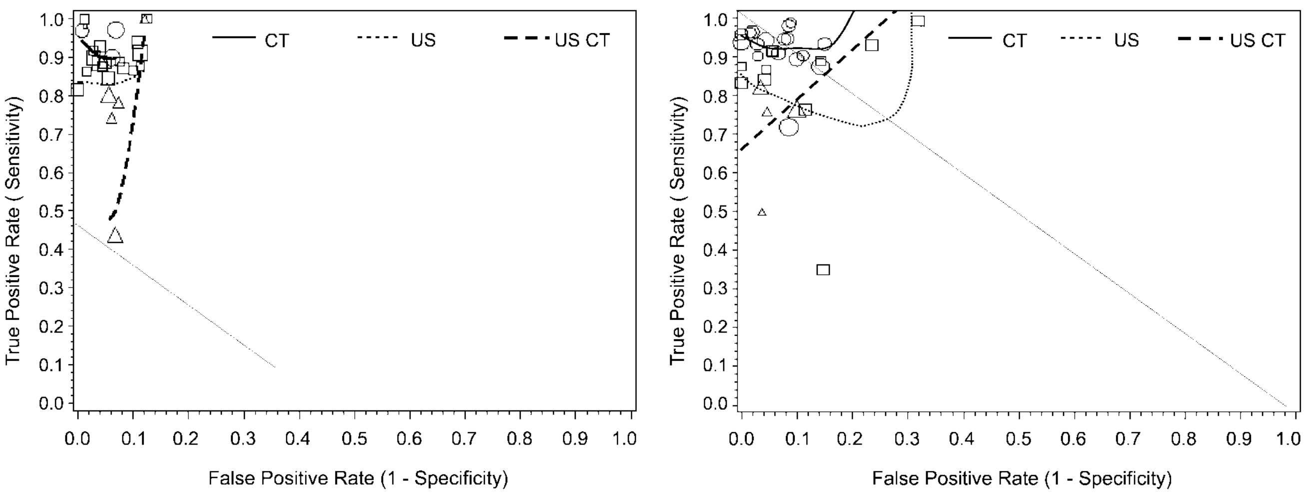

Figure 2: Graph shows individual study estimates of sensitivity and 1 − specificity for US, CT, and both imaging modalities in **(a)** children and **(b)** adults as a method for pooling DORs by assuming them constant regardless of diagnostic thresholds. Curved lines represent nonparametric curves fitted to data according to the summary sensitivity and specificity of studies. Horizontal axes represent false-positive ratio (1 − specificity) and vertical axes represent true-positive ratio (sensitivity). The size of the circles and squares is proportional to the quality of studies. ○ = CT-only studies, □ = US-only studies, △ = US and CT studies. Note that most point estimates of diagnostic tests of studies of children (**a**) and adults (**b**) tended to the symmetry around the "sensitivity = specificity" line (diagonal line).

biases may lead to an overestimation of the sensitivity and to an underestimation of the specificity (105). Verification bias, present in other meta-analyses (96,97), remains a concern in this meta-analysis since most surgeons were aware of the results of the diagnostic imaging tests before making their case management decisions. However, no differential bias is expected to occur between US and CT studies since the diagnostic assessment of both imaging modalities was subject to verification bias. Therefore, the differential performance between US and CT in children and adults should not be significantly affected by this bias.

In this meta-analysis, two of the readers who reviewed the articles and extracted the data for the appraisal of study quality also rated the study quality. This would have the potential for bias; however, our results demonstrated a good interrater agreement between the third reader, who did not participate in these methodologic phases of the study, and the other two readers, which minimizes a potential concern for selection bias. With regard to the number of references screened, this is directly dependent on the search strategy applied and thesaurus descriptors used, which may vary between different meta-analyses on diagnostic tests for appendicitis.

The lack of true randomization of patients to US or CT groups in most studies and a potential geographic bias (preferential use of CT in studies conducted in North America and of US in studies in non–North American countries) could not be controlled for in this meta-analysis. In some studies, the results of a negative CT scan were considered definitive, whereas those of a negative or nondiagnostic US scan indicated a need for further testing or observation. None of the studies in our meta-analysis included information about patient weight or body mass index, which are factors that are particularly important in the case of obese patients, because the results of US are not as definitive as those of CT for these patients (106). Confounding variables for enhanced CT techniques, such as the gauge and site of intravenous access, the amount and rate of intravenous contrast agent, and the scan delay, have not been reported in many of the studies. The lack of information about the severity of the symptoms or the pretest clinical likelihood of appendicitis in most studies impaired the assessment of spectrum bias (106).

Typically, the evaluation of a diagnostic test will include examination of thresholds for test positivity using ROC curves. However, with imaging diagnosis the data are often simplified into binary data. Because we analyzed data on the basis of reported values of individual studies, the cutoff approaches were already set. There were insufficient data in these studies to permit any analysis of the effect of different thresholds for test positivity on the results. We did conduct a sensitivity analysis based on information available for the incidence of indeterminate or equivocal cases from a few studies, but this may not adequately represent the way different centers handled equivocal cases. Although some authors considered nonvisualization of the appendix to indicate a negative result, others considered it to indicate a positive or equivocal result since nonvisualization of the appendix could represent perforation. We included positive and negative results in the meta-analysis as reported by the authors.

In summary, this meta-analysis demonstrated that, in contrast with specificity, the difference in pooled sensitivities and ORs between US and CT is significant in both studies of children and studies of adults. On the basis of the results of this meta-analysis, and taking into account only the diagnostic performance of imaging modalities, CT has a significantly higher sensitivity than does US for the diagnosis of appendicitis in adults and children. We should note, however, that the sensitivity of US is reasonably high in children and that radiation issues are of special concern in this age group. Thus, after taking into account both diagnostic performance and safety, one should face the trade-off between the future risk of cancer with the use of CT, which is particularly worrisome in the pediatric group, and the risk of missing positive cases with US.

Acknowledgments: We acknowledge Sharon Nancekivell, BA, BEd, MA, PhD, for her editorial work and Jill Bursey for her help with the submission of inquiry letters to authors of articles.

References

1. Storer EH. Appendix. In: Schwartz SI, ed. Principles of surgery. 3rd ed. New York, NY: McGraw-Hill International, 1979; 1257–1267.
2. Federle MP. Focused appendix CT technique: a commentary. Radiology 1997;202: 20–21.
3. Henderson J, Goldacre MJ, Fairweather JM. Conditions accounting for substantial time spent in hospital in children age 1–14 years of age. Arch Dis Child 1992;67:83–86.
4. Siegel MJ. Acute appendicitis in childhood: the role of US. Radiology 1992;185:341–342.
5. Lund DP, Kolkman J. Appendicitis. In: Walker WA, Durie PR, Hamilton JR, eds. Pediatric gastrointestinal disease. 2nd ed. Vol 1. St Louis, Mo: Mosby–Year Book, 1996; 907–914.
6. Berry J Jr, Malt RA. Appendicitis near its centenary. Ann Surg 1984;200:567–575.
7. Lewis FR, Holcroft JW, Boey J, et al. Appendicitis: critical review of 1000 patients. Arch Surg 1975;110:677–684.
8. Puylaert JB, van der Zant FM, Rijke AM. Sonography and the acute abdomen: practical considerations. AJR Am J Roentgenol 1997;168:179–186.
9. Sivit CJ. Imaging children with acute right lower quadrant pain. Pediatr Clin North Am 1997;44:575–589.
10. Sivit CJ, Siegel MJ, Applegate KE, Newman KD. When appendicitis is suspected in children. RadioGraphics 2001;21:247–262.
11. Malone AJ, Wolf CR, Malmed AS, Melliere BF. Diagnosis of acute appendicitis: value of unenhanced CT. AJR Am J Roentgenol 1993;160:763–766.
12. Frush DP, Donnelly LF. Helical CT in children: technical considerations and body applications. Radiology 1998;209:37–48.
13. Rao PM, Rhea JT, Novelline RA. Helical CT of appendicitis and diverticulitis. Radiol Clin North Am 1999;37:895–910.
14. Puylaert JB. Acute appendicitis: US evaluation using graded compression. Radiology 1986;158:355–360.
15. Larson JM, Peirce JC, Ellinger DM, et al. The validity and utility of sonography in the diagnosis of appendicitis in the community

setting. AJR Am J Roentgenol 1989;153: 687–691.

16. Perdigon PL. Imaging techniques for diagnosis of serious intraabdominal and pelvic infections. Primary Care Update Obstet Gynecol 1999;6:115–121.

17. Rao PM, Wittenberg J, McDowell RK, et al. Appendicitis: use of arrowhead sign for diagnosis at CT. Radiology 1997;202:363–366.

18. Rao PM, Rhea JT, Novelline RA. Helical computed tomographic incidence and characterization of appendoliths in 100 patients with appendicitis. Emerg Radiol 1997;4: 55–61.

19. Rao PM, Rhea JT, Rattner DW, Venus LG, Novelline RA. Introduction of appendiceal CT: impact on negative appendectomy and appendiceal perforation rates. Ann Surg 1999;229:344–349.

20. Birnbaum BA. CT evaluation of patients with acute lower abdominal pain. In: Balfe DM, Levine MS, eds. Syllabus: categorical course in diagnostic radiology—gastrointestinal 1997. Oak Brook, Ill: Radiological Society of North America, 1997; 199–209.

21. Birnbaum BA, Balthazar EJ. CT of appendicitis and diverticulitis. Radiol Clin North Am 1994;32:885–898.

22. Wise SW, Labuski MR, Kasales CJ, et al. Comparative assessment of CT and sonographic techniques for appendiceal imaging. AJR Am J Roentgenol 2001;176:933–941.

23. Styrud J, Josephson T, Eriksson S. Reducing negative appendectomy: evaluation of ultrasonography and computed tomography in acute appendicitis. Int J Qual Health Care 2000;12:65–68.

24. Cha SH, Kim SH, Lee ES, et al. Saline-filled appendiceal ultrasonography in clinically equivocal acute appendicitis. Abdom Imaging 1996;21:525–529.

25. Vazquez Ronco MA, Morteruel Arizkuren E, Garcia Ojeda E, Mintegui Raso S, Capape Zache S, Benito Fernandez J. Rendimiento de la ecografia abdominal en el diagnostico de appendicitis aguda. An Pediatr (Barc) 2003;58:556–561.

26. Morris KT, Kavanagh M, Hansen P, Whiteford MH, Deveny K, Standage B. The rational use of computed tomography scans in the diagnosis of appendicitis. Am J Surg 2002;183:547–550.

27. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995;48:119–130.

28. Mulrow CD, Linn WD, Gaul MK, et al. Assessing quality of a diagnostic test evaluation. J Gen Intern Med 1989;4:288–295.

29. Deville WL, van der Windt AW, Dzaferagic A, Bezemer PD, Bouter LM. The test of Lasegue: systematic review of the accuracy in diagnosing herniated discs. Spine 2000; 25:1140–1147.

30. Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. J Clin Epidemiol 1991;44:381–390.

31. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18:2693–2708.

32. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Smith GD, Altman DG, Egger M, eds. Systematic reviews in health care: meta-analysis in context. 2nd ed. London, England: BMJ, 2002; 248–482.

33. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7: 177–188.

34. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–1558.

35. Garcia Pena BM, Mandl KD, Kraus SJ, et al. Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children. JAMA 1999;282:1041–1046.

36. Chen SC, Chen KM, Wang SM, Chang KJ. Abdominal sonography screening of clinically diagnosed or suspected appendicitis before surgery. World J Surg 1998;22: 449–452.

37. Birnbaum BA, Wilson SR. Appendicitis at the millennium. Radiology 2000;215:337–348.

38. Ege G, Akman H, Sahin A, Bugra D, Kuzucu K. Diagnostic value of unenhanced helical CT in adult patients with suspected acute appendicitis. Br J Radiol 2002;75:721–725.

39. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston, Mass: Little, Brown, 1991.

40. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York, NY: Churchill Livingstone, 1997.

41. Newman TB, Browner WS, Cummings SR, Hulley SB. Designing an observational study: cross-sectional and case-control studies. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. Designing clinical research: an epidemiologic approach. 2nd ed. Philadelphia, Pa: Lippincott Williams & Wilkins, 2001.

42. Egger M, Davey SG. Misleading meta-analysis: lessons from "an effective, safe, simple" intervention that wasn't. BMJ 1995;310:752–754.

43. Copas JB, Shi JQ. A sensitivity analysis for publication bias in systematic reviews. Stat Methods Med Res 2001;10:251–265.

44. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Stat Soc 1988;151:419–463.

45. Rubin SZ, Martin DJ. Ultrasonography in the management of possible appendicitis in childhood. J Pediatr Surg 1990;25:737–740.

46. Vignault F, Filiatrault D, Brandt ML, Garel L, Grignon A, Ouimet A. Acute appendicitis in children: evaluation with US. Radiology 1990;176:501–504.

47. Siegel MJ, Carel C, Surrat S. Ultrasonography of acute abdominal pain in children. JAMA 1991;266:1987–1989.

48. Hayden CK, Kuchelmeister J, Lipscomb TS. Sonography of acute appendicitis in childhood: perforation versus nonperforation. J Ultrasound Med 1992;11:209–216.

49. Quillin SP, Siegel MJ, Coffin CM. Acute appendicitis in children: value of sonography in detecting perforation. AJR Am J Roentgenol 1992;159:1265–1268.

50. Sivit CJ, Newman KD, Boenning DA, et al. Appendicitis: usefulness of US in diagnosis in a pediatric population. Radiology 1992; 185:549–552.

51. Crady SK, Jones JS, Wyn T, Luttenton CR. Clinical validity of ultrasound in children with suspected appendicitis. Ann Emerg Med 1993;22:1125–1129.

52. Wong ML, Casey SO, Leonidas JC, Elkowitz SS, Becker J. Sonographic diagnosis of acute appendicitis. J Pediatr Surg 1994;29: 1356–1360.

53. Ramachandran P, Sivit CJ, Newman KD, Schwartz MZ. Ultrasonography as an adjunct in the diagnosis of acute appendicitis: a 4-year experience. J Pediatr Surg 1996; 31:164–167.

54. Hahn HB, Hoepner FU, Kalle T, et al. Sonography of acute appendicitis in children: 7 years experience. Pediatr Radiol 1998;28: 147–151.

55. Davidson PM, Douglas CD, Hosking CS. Graded compression ultrasonography in the assessment of the "tough decision"

acute abdomen in childhood. Pediatr Surg Int 1999;15:32–35.

56. Lessin MS, Chan M, Catallozzi M, et al. Selective use of ultrasonography for acute appendicitis in children. Am J Surg 1999; 177:193–196.

57. Rice HE, Arbesman M, Martin DJ, et al. Does early ultrasonography affect management of pediatric appendicitis? a prospective analysis. J Pediatr Surg 1999;34:754–759.

58. Karakas SP, Guelfguat M, Leonidas JC, Springer S, Singh SP. Acute appendicitis in children: comparison of clinical diagnosis with ultrasound and CT imaging. Pediatr Radiol 2000;30:94–98.

59. Sivit CJ, Applegate KE, Stallion A, et al. Imaging evaluation of suspected appendicitis in a pediatric population: effectiveness of sonography versus CT. AJR Am J Roentgenol 2000;175:977–980.

60. Ang A, Chong NK, Daneman A. Pediatric appendicitis in "real-time": the value of sonography in diagnosis and treatment. Pediatr Emerg Care 2001;17:334–340.

61. Lowe LH, Penney MW, Stein SM, et al. Unenhanced limited CT of the abdomen in the diagnosis of appendicitis in children: comparison with sonography. AJR Am J Roentgenol 2001;176:31–35.

62. Mullins ME, Kircher MF, Ryan DP, et al. Evaluation of suspected appendicitis in children using limited helical CT and colonic contrast material. AJR Am J Roentgenol 2001;176:37–41.

63. Fefferman NR, Roche KJ, Pinkney LP, Ambrosino MM, Genieser NB. Suspected appendicitis in children: focused CT technique for evaluation. Radiology 2001;220: 691–695.

64. Han TI. Improved sonographic visualization of the appendix with a saline enema in children with suspected appendicitis. J Ultrasound Med 2002;21:511–516.

65. Kaiser S, Frenckner B, Jorulf HK. Suspected appendicitis in children: US and CT—a prospective randomized study. Radiology 2002;223:633–638.

66. Chang CC, Tsai CY, Lin CC, Jeng LB, Lee CC, Kao CH. Comparison between technetium-99m hexamethylpropyleneamineoxide labeled white blood cell abdomen scan and abdominal sonography to detect appendicitis in children with an atypical clinical presentation. Hepatogastroenterology 2003; 50:426–429.

67. Kaiser S, Finnbogason T, Jorulf HK, Soderman B, Frenckner B. Suspected appendicitis in children: diagnosis with contrast-enhanced versus nonenhanced helical CT. Radiology 2004;231:427–433.

68. Adams DH, Fine C, Brooks DC. High-resolution real-time ultrasonography: a new tool in the diagnosis of acute appendicitis. Am J Surg 1988;155:93–97.

69. Bilbey JH, Gibney RG, Cooperberg PL. Ultrasonography in acute appendicitis. Can Assoc Radiol J 1989;40:22–24.

70. Kang WM, Lee CH, Chou YH, et al. A clinical evaluation of ultrasonography in the diagnosis of acute appendicitis. Surgery 1989;105:154–159.

71. Amgwerd M, Rothlin M, Candinas D, Schimmer R, Klotz HP, Largiader F. Ultrasound diagnosis of appendicitis by surgeons: a matter of experience? a prospective study. Langenbecks Arch Chir 1994;379:335–340.

72. Balthazar EJ, Birnbaum BA, Yee J, et al. Acute appendicitis: CT and US correlation in 100 patients. Radiology 1994;190:31–35.

73. Incesu L, Ciskun A, Selcuk MB, Akan H, Sozubit S, Bernay F. Acute appendicitis: MR imaging and sonographic correlation. AJR Am J Roentgenol 1997;168:669–674.

74. Lane MJ, Katz DS, Ross BA, Clautice-Engle TL, Mindelzun RE, Jeffrey RB Jr. Unenhanced helical CT for suspected acute appendicitis. AJR Am J Roentgenol 1997;168: 405–409.

75. Schuler JG, Shortsleeve MJ, Goldenson RS, Perez-Rossello JM, Perlmutter RA, Thorsen A. Is there a role for abdominal computed tomographic scans in appendicitis? Arch Surg 1998;133:373–377.

76. Crombe A, Weber F, Gruner L, Martins A, Fouque P, Barth X. Abdominopelvic ultrasonography in suspected acute appendicitis: prospective study in adults. Ann Chir 2000;125:57–61.

77. Horton MD, Counter SF, Florence MG, Hart MJ. A prospective trial of computed tomography and ultrasonography for diagnosing appendicitis in the atypical patient. Am J Surg 2000;179:379–381.

78. Kamel IR, Goldberg SN, Keogan MT, Rosen MP, Raptopoulos V. Right lower quadrant pain and suspected appendicitis: nonfocused appendiceal CT—review of 100 cases. Radiology 2000;217:159–163.

79. Souparis AC, Makris JG, Arvaniti M, et al. The role of ultrasound scanning of the lower abdomen in the differential diagnosis of acute appendicitis. Arch Gastroenterohepatol 2000;19:28–30.

80. Walker S, Haun W, Clark J, McMillin K, Zeren F, Gilliland T. The value of limited computed tomography with rectal contrast in the diagnosis of acute appendicitis. Am J Surg 2000;180:450–455.

81. Bouillot JL, Ruiz A, Alamowitch B, et al. Suspected acute appendicitis: role of enhanced helical computed tomography—prospective study of 100 patients. Ann Chir 2001;126:427–433.

82. Kan JH, Fines BP, Funaki B. Conventional and hydrocolonic US of the appendix with CT correlation performed by on-call radiology residents. Acad Radiol 2001;8:1208–1214.

83. Maluccio MA, Covey AM, Weyant MJ, Eachempati SR, Hydo LJ, Barie PS. A prospective evaluation of the use of emergency department computed tomography for suspected acute appendicitis. Surg Infect (Larchmt) 2001;2:205–214.

84. Wijetunga R, Tan BS, Rouse JC, Bigg-Wither GW, Doust BD. Diagnostic accuracy of focused appendiceal CT in clinically equivocal cases of acute appendicitis. Radiology 2001;221:747–753.

85. Cakirer S, Basak M, Colakoglu B, Bankaoglu M. Diagnosis of acute appendicitis with unenhanced helical CT: a study of 130 patients. Emerg Radiol 2002;9:155–161.

86. Hershko DD, Sroka G, Bahouth H, Ghersin E, Mahajna A, Krausz MM. The role of selective computed tomography in the diagnosis and management of suspected acute appendicitis. Am Surg 2002;68:1003–1007.

87. Raman SS, Lu DS, Kadell BM, Vodopich DJ, Sayre J, Cryer H. Accuracy of nonfocused helical CT for the diagnosis of acute appendicitis: a 5-year review. AJR Am J Roentgenol 2002;178:1319–1325.

88. Sun SS, Wu HS, Wang JJ, Ho ST, Kao A. Comparison between technetium 99m hexamethylpropyleneamine oxide labeled white blood cell abdominal scan and abdominal sonography to detect appendicitis in adult patients with atypical clinical presentation. Abdom Imaging 2002;27:734–738.

89. Ujiki MB, Murayama KM, Cribbins AJ, et al. CT scan in the management of acute appendicitis. J Surg Res 2002;105:119–122.

90. Wong SK, Chan LP, Yeo A. Helical CT imaging of clinically suspected appendicitis: correlation of CT and histological findings. Clin Radiol 2002;57:741–745.

91. Yetkin G, Basak M, Isgor A, Kebudi A, Akgun I. Can negative appendectomy rate be decreased by using spiral computed tomography without contrast material? Acta Chir Belg 2002;102:334–337.

92. Hong JJ, Cohn SM, Ekeh AP, et al. A prospective randomized study of clinical assessment versus computed tomography for the diagnosis of acute appendicitis. Surg Infect (Larchmt) 2003;4:231–239.

93. Moses LE, Shapiro D, Littenberg B. 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.

94. Shapiro DE. Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. Acad Radiol 1995; 2(suppl 1):S37–S47.

95. Vermeulen B, Morabia A, Unger PF, et al. Acute appendicitis: influence of early pain relief on the accuracy of clinical and US findings in the decision to operate—a randomized trial. Radiology 1999;210:639–643.

96. Orr RK, Porter D, Hartman D. Ultrasonography to evaluate adults for appendicitis: decision making based on meta-analysis and probabilistic reasoning. Acad Emerg Med 1995;2:644–650.

97. Terasawa T, Blackmore CC, Bent S, Kohlwes RJ. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med 2004;141:537–546.

98. Gray JE. Safety (risk) of diagnostic radiology exposures radiation risk: a primer. American College of Radiology Commission on Physics and Radiation Safety, Committee on Radiologic Units, Standards, and Protection. Reston, Va: American College of Radiology, 1996; 15–17.

99. Brenner D, Elliston C, Hall E, Berdon W. Estimated risks of radiation induced fatal cancer from pediatric CT. AJR Am J Roentgenol 2001;176:289–296.

100. Williams G, Zankl M, Abmayr W, Veit R, Drexler G. The calculation of dose from external photon exposures using reference and realistic human phantoms and Monte Carlo methods. Phys Med Biol 1986;31:449–452.

101. Redd JT. Sensitivity, specificity, prevalence, and disease stage [letter]. Ann Intern Med 1994;120(4):345.

102. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51–63.

103. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. Am J Epidemiol 1995;141:680–689.

104. Kwikkel HJ, Quaak MJ, de With C. Predictive value of the abnormal Pap smear: a retrospective analysis of error rates. Eur J Obstet Gynecol Reprod Biol 1986;21:101–112.

105. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061–1066.

106. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926–930.