US or CT for diagnosis of appendicitis in children and adults: a meta-analysis

CRD summary
The authors concluded that computed tomography is more sensitive than ultrasound, i.e. fewer cases of appendicitis are missed, but that specificity is similar, i.e. similar numbers of false positives are generated. These conclusions reflect the data presented, but should be viewed with caution given the methodological limitations of the review.

Authors' objectives
To evaluate the diagnostic performance of ultrasonography (US) and computed tomography (CT) for the detection of appendicitis in children and adults.

Searching
Studies published between 1986 and December 2004 were sought by searching MEDLINE, EMBASE, CINAHL, the Cochrane Controlled Trials Register, the Cochrane Database of Systematic Reviews and ACP Journal Club; the search terms were reported. The bibliographies of retrieved articles were screened for additional studies.

Study selection
Study designs of evaluations included in the review
Prospective and retrospective studies were eligible for inclusion. Cases series were excluded.

Specific interventions included in the review
Studies evaluating the diagnostic performance of abdominal US and/or CT were eligible for inclusion. The majority of included studies used linear US probes with a frequency of 5 MHz or greater and/or third-generation helical CT units; further technical details and requisite criteria for diagnosis were reported in the paper.

Reference standard test against which the new test was compared
The included studies were required to use either surgical findings plus follow-up of greater than 1 week, or histological evaluation plus follow-up as the reference standard.

Participants included in the review
Studies of adults and/or children, where data could be stratified (maximum age for children 20 years, minimum age for adults 13 years) such that outlying ages did not exceed 5% of the total sample size, were eligible for inclusion. The studies were required to include both males and females and have a calculable prevalence of appendicitis of between 15 and 75%. Data on pregnant women were excluded.

Outcomes assessed in the review
The included studies were required to report sufficient information to allow the derivation of absolute numbers of true positives, false negatives, false positives and true negatives. Calculated values for sensitivity and specificity, with 95% confidence intervals (CIs), were reported in the review.

How were decisions on the relevance of primary studies made?
Two reviewers independently assessed articles for inclusion.

Assessment of study quality
Three radiologists (two of whom were blinded to journal name, author and year of publication) independently assessed the methodological quality of the included studies using pre-specified guidelines. The quality criteria related to participant characteristics, study design (including selection of the participants), the method of reporting the results and details of the index test and reference standard, blind interpretation of the test and reference standard, and application
of the reference standard independently of the index test result. The assessment process was used to generate an overall score; the maximum was 32 points or 100%.

Data extraction
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction. The authors of the primary studies were contacted for further data where calculated values for diagnostic tests differed from reported values.

Methods of synthesis
How were the studies combined?
Pooled point estimates and 95% CIs were generated for the sensitivity and specificity of US and CT in children and adults using a Mantel-Haenszel fixed-effect model, with studies weighted by quality score. Because summary point estimates are not recommended where the diagnostic threshold varies between studies, as was the case in this article, the results were also summarised by fitting summary receiver operating characteristic curves. The potential for publication bias was assessed by visual examination of funnel plots of sensitivity and specificity.

How were differences between studies investigated?
Overall heterogeneity was assessed by visual examination of funnel plots of sensitivity and specificity. The chi-squared test was used to assess between-study heterogeneity prior to the decision to use a fixed-effect model. A one-way sensitivity analysis was used to assess the impact of three studies deemed to be outliers on the basis of their reported sensitivity. Multivariate regression analysis was used to assess the impact of three pre-specified covariates (retrospective or prospective study design, publication date and study location) on diagnostic performance, as indicated by the diagnostic odds ratio. The median quality score was 34.4% for studies of children and 42.2% for studies of adults.

Results of the review
A total of 57 studies were included in the review. Twenty-six studies of children (mean age 7 to 12 years) reported results for US alone (6,850 participants), CT alone (598 participants), or combined US and CT (1,908 participants). Thirty-one studies of adults (mean age 20 to 49 years) reported results for US alone (903 participants), CT alone (2,394 participants), or combined US and CT (1,044 participants).

The pooled sensitivity of CT in children (94%, 95% CI: 92, 97) was 6% higher than that of US (88%, 95% CI: 86, 90) after the removal of one outlier (p=0.001). The pooled sensitivity of CT in adults (94%, 95% CI: 92, 95) was 11% higher than that of US (83%, 95% CI: 78, 87) after the removal of two outliers (p=0.001). The pooled estimates of specificity were similar for CT in children (95%, 95% CI: 94, 97) and in adults (94%, 95% CI: 94, 96), and for US in children (94%, 95% CI: 92, 95) and in adults (93%, 95% CI: 90, 96). Regression models did not indicate any evidence of heterogeneity related to the covariates tested.

Examination of the funnel plots indicated that publication bias was unlikely.

Authors' conclusions
CT had a significantly higher sensitivity than US in both children and adults. However, the radiation associated with CT is also a consideration, especially in children.

CRD commentary
The review addressed a clearly stated research question, which was appropriately defined by the inclusion criteria. A thorough literature search was conducted, no language restrictions were specified, and attempts were made to assess publication bias. It should be noted, however, that standard funnel plots are generally considered to be of limited value in assessing publication bias with respect to diagnostic accuracy studies. The review methods included measures to avoid the introduction of error and bias in the study selection and quality assessment processes, though it was unclear
whether these measures were applied to the data extraction. An assessment of the quality of the included studies was reported in full and used to weight studies in the meta-analyses. The criteria used addressed some aspects of methodological quality which are important to diagnostic accuracy studies, but also included study characteristics not directly relevant to methodological quality. The use of a summary quality score to weight studies in meta-analyses is unlikely to be informative since it may mask potentially important differences in individual items. The authors indicated that the generation of point estimates of sensitivity and specificity where the diagnostic threshold varies between studies is problematic and can lead to the underestimation of diagnostic performance. The conclusions of the review reflect the data presented, but should be interpreted with caution given the methodological limitations outlined.

**Implications of the review for practice and research**

Practice: The authors stated that CT had a significantly higher sensitivity than US in both children and adults. However, the radiation associated with CT should also be considered, especially in children.

Research: The authors did not state any implications for further research.

**Bibliographic details**


**PubMedID**

16928974

**DOI**

10.1148/radiol.2411050913

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Adult; Appendicitis /radiography /ultrasonography; Child; False Negative Reactions; False Positive Reactions; Female; Humans; Male; Prospective Studies; Retrospective Studies; Safety; Sensitivity and Specificity; Tomography, X-Ray Computed

**AccessionNumber**

12006007429

**Date bibliographic record published**

31/12/2007

**Date abstract record published**

31/12/2007

**Record Status**

This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.