Diagnosis of appendicitis in adults by ultrasonography or computed tomography: a systematic review and meta-analysis
Weston A R, Jackson T J, Blamey S

CRD summary
This review assessed the accuracy of ultrasonography and computed tomography (CT) in diagnosing appendicitis in adults. The authors concluded that evidence suggests that CT is a sensitive and specific diagnostic test, but its routine use cannot be supported. The lack of reporting of review methods and the inclusion of generally poor-quality studies mean that the conclusions may be overstated.

Authors' objectives
To assess the accuracy of ultrasonography and computed tomography (CT) in diagnosing acute appendicitis in adults.

Searching
MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews, DARE, NHS EED and HTA were searched for reports published in English from January 1985 to February 2003; the search terms were reported. The reference lists of key publications were also checked.

Study selection
Study designs of evaluations included in the review
Diagnostic studies with at least 50 patients were eligible for inclusion. The studies had to adequately describe methods used to recruit patients and previous clinical and diagnostic investigations. Most of the studies were prospective studies; some studies did not provide details of the study design.

Specific interventions included in the review
Studies of ultrasonography or helical CT were eligible for inclusion. Studies were excluded if the diagnostic test was conducted in a non-standard or an outdated manner, or by non-routine operators (trainees or surgeons only).

Reference standard test against which the new test was compared
The inclusion criteria were not defined in terms of the reference standard test. The included studies generally used histopathology or follow-up as the reference standard tests, but many studies provided no details of the follow-up.

Participants included in the review
The review focused on studies of adults. Excluded studies were those that only included patients who had undergone appendicectomy; studies that recruited pregnant women only; studies conducted in children only or those in which more than 50% were children; and studies conducted in the elderly only. Studies were also excluded if they used additional patient inclusion criteria that made their population irrelevant to the review (no further details of this exclusion criterion were reported). Studies reporting duplicate samples of patients or subsets of patients in other studies were also excluded.

The review focused on two groups of patients: patients with suspected appendicitis after routine clinical investigations (all presentations); and only patients with an equivocal diagnosis after routine clinical investigations (equivocal only). Most of the included studies selected consecutive patients; other studies used non-consecutive recruitment or gave no details of the recruitment methods.

Outcomes assessed in the review
Studies that evaluated diagnostic performance and reported the sensitivity, specificity, diagnostic accuracy, positive predictive value (PPV) or negative predictive value (NPV) were eligible for inclusion.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
The studies were assessed and scored using a modification of the diagnostic-specific checklist published by the Cochrane Screening and Diagnostic Tests Methods group. The authors did not state who performed the validity assessment.

Data extraction
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction.

Where studies reported more than one value for diagnostic performance tests of ultrasonography, the authors selected values where non-visualisations and non-diagnostic ultrasonography were classified as disease-negative.

Methods of synthesis
How were the studies combined?
The mean prevalence of appendicitis was calculated for all studies combined, for all presentations and for equivocal presentations only. Pooled diagnostic performance measures were calculated for studies with similar populations, test techniques, interpretation and study quality. Since the studies were considered similar, all studies were included in the meta-analyses. The association between sensitivity and specificity was assessed before pooling; no association was found. Pooling was undertaken on the assumption that there was no underlying cut-point effect, and with weighting according to the sample size. The pooled sensitivity, specificity, PPV and NPV for ultrasonography and CT were calculated, along with 95% confidence intervals (CIs), for all studies, for all presentations and for equivocal presentations only.

How were differences between studies investigated?
Statistical heterogeneity was not investigated, though some potential causes of differences in the prevalence of appendicitis among studies were mentioned in the narrative synthesis.

Results of the review
Thirty-two studies were included. Ultrasonography was examined in 22 studies (n=5,060) and CT was examined in 12 studies (n=1,745).

The studies were generally of a poor quality, with quality scores ranging from 2 to 9 out of a possible 13.

Ultrasonography.

All presentations (17 studies): the overall prevalence of appendicitis was 41% (range: 23 to 78). The sensitivity was 88.3 (95% CI: 86.7, 89.8) and the specificity was 92.3 (95% CI: 91.2, 93.3).

Equivocal presentation only (5 studies): the overall prevalence of appendicitis was 28% (range: 19 to 54). The sensitivity was 76.4 (95% CI: 70.1, 81.7) and the specificity was 95.2 (95% CI: 93.0, 96.7).

CT.

All presentations (7 studies): the overall prevalence of appendicitis was 46% (range: 24 to 72). The sensitivity was 96.6 (95% CI: 94.3, 98.0) and the specificity was 94.0 (95% CI: 91.6, 95.8).

Equivocal presentation only (5 studies): the overall prevalence of appendicitis was 46% (range: 30 to 78). The sensitivity was 96.4 (95% CI: 93.4, 98.0) and the specificity was 95.7 (95% CI: 92.9, 97.4).

The pooled sensitivity, specificity, PPV and NPV were also reported for all studies combined, while the pooled PPV
and NPV were reported for all presentations and equivocal presentations only.

**Authors’ conclusions**

Ultrasound has only modest ability to detect patients with acute appendicitis. Evidence suggests that CT is a sensitive and specific diagnostic test for appendicitis, but the cost and availability of CT need to be taken into account before the routine use of CT can be supported. Further research is required.

**CRD commentary**

The review addressed a clear question in terms of the participants, intervention, outcomes and study design. Three databases were searched, but no attempts were made to minimise publication and language bias. The methods used to select studies, assess validity and extract the data were not described, so it is not known whether any efforts were made to reduce errors and bias. Validity was assessed using a modification of a specified checklist but the modifications made were not described. Quality scores were reported, but few details of methodological flaws were discussed; this made it difficult to assess the quality and hence the strength of evidence provided by the included studies. No details were provided of prior investigations or of inclusion criteria for individual studies.

The authors pooled data from studies considered to be in similar populations and of similar quality. However, the studies generally appeared to be of a poor quality, with several studies not describing methods used to exclude appendicitis in test-negative patients or reporting whether the test diagnosis was made independently of clinical information. In addition, as the authors stated, the wide range of appendicitis rates in the studies suggests dissimilar populations. The sensitivity, specificity, PPV and NPV results for individual studies were not reported, so it was not possible to determine if pooling was appropriate and, therefore, if the pooled estimate of sensitivity and specificity applies in all situations. Confining the meta-analysis to higher quality studies with an examination of different results among studies would have provided stronger evidence. An interpretation of the one study comparing ultrasonography and CT was only mentioned briefly in the discussion, without any accompanying diagnostic performance test data. Given the limitations highlighted, the authors’ conclusions about the high sensitivity and specificity of CT may be overstated.

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

Practice: The authors stated that the review did not support the routine use of CT, and that cost and time delays need to be taken into account when considering the use of CT.

Research: The authors stated that future studies should be conducted in populations that are well-defined with respect to prior investigations, and should evaluate the sequelae of false-negative and false-positive diagnoses.

**Bibliographic details**


**PubMedID**

16110717

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Adolescent; Adult; Aged; Aged, 80 and over; Appendicitis /diagnosis /radiography /ultrasonography; Child; Child, Preschool; Female; Humans; Male; Middle Aged; Sensitivity and Specificity; Tomography, Spiral Computed

**AccessionNumber**
12005008397

Date bibliographic record published
31/03/2006

Date abstract record published
31/03/2006

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.