Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents

Terasawa T, Blackmore C, Bent S, Kohlwes R J

CRD summary
This review assessed the diagnostic accuracy of computed tomography (CT) and ultrasonography in adults and adolescents with suspected appendicitis. The authors concluded that CT is probably more accurate than ultrasonography in diagnosing appendicitis. The conclusions of this well-conducted review are appropriately cautious given the methodological weaknesses of the included studies.

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

Searching
MEDLINE and EMBASE were searched from 1966 to December 2003. The reference lists of eligible studies, review articles and relevant textbooks were also checked, and experts in this field were consulted.

Study selection
Study designs of evaluations included in the review
Prospective diagnostic accuracy studies were eligible for inclusion. Cohort studies and randomised controlled trials (RCTs) were included in the review.

Specific interventions included in the review
Studies of CT or graded compression ultrasonography were eligible for inclusion.

Reference standard test against which the new test was compared
Eligible studies had to confirm appendicitis using either clinical follow-up or surgical and pathologic confirmation.

Participants included in the review
Studies of patients aged 14 years or older with suspected appendicitis were eligible for inclusion. The original inclusion criteria specified that patients had to be 18 years or older, but this was changed after the pilot search since many studies included adolescents as well as adults.

Outcomes assessed in the review
Studies presenting true-positive, false-positive, true-negative and false-positive results that allowed the construction of 2x2 tables were eligible for inclusion. The authors used these tables to calculate the sensitivity, specificity, and positive and negative likelihood ratios (LRs).

How were decisions on the relevance of primary studies made?
Two reviewers assessed study eligibility.

Assessment of study quality
Study quality was assessed using the checklist of the Cochrane Methods Group on systematic reviews of screening and diagnostic tests. The authors did not state how the papers were assessed for quality, or how many reviewers performed the quality assessment.

Data extraction
Two reviewers, who were not blinded to journal details, independently extracted the data. One reviewer extracted non-English language studies, assisted by a clinician who was a native speaker of the relevant language. Any discrepancies were resolved by discussion or referral to a third reviewer. Study authors were contacted for clarification if necessary.

**Methods of synthesis**

How were the studies combined?
Summary sensitivities and specificities were calculated using the standard method for pooling proportions, regardless of the presence of heterogeneity. Summary LRs were calculated using either the Mantel-Haenszel fixed-effect model where there was no evidence of statistical heterogeneity, or the DerSimonian-Laird random-effects model if there was evidence of statistical heterogeneity.

How were differences between studies investigated?
Heterogeneity was assessed using a chi-squared test and by calculating the I-squared statistic. Pre-planned subgroup analyses of adult participants (18 years or over), different patient presentations (suspected versus atypical), and studies with a high proportion of women (more than 67%) were conducted. Sensitivity analyses were also conducted to assess the impact of uninterpretable results, by undertaking a 'worst case' analysis in which any results that could not be interpreted were assumed to be false-positive or false-negative.

**Results of the review**

Twelve studies (n=1,172) of CT were included, of which three were RCTs (n=192) and the rest were prospective cohort studies (n=980). Fourteen studies of ultrasonography (all prospective cohort studies; n=1,516) were included. Four of the included studies (including one RCT) directly compared CT and ultrasonography.

All of the included studies were at risk of differential reference standard bias (differing reference standards for patients with positive or negative test results). Most studies did not provide a full description of the reference standard and whether its interpretation was conducted blind to the imaging test results and other clinical information. Only a few studies reported the disease spectrum.

There was no evidence of statistical heterogeneity amongst the positive or negative LRs (P>0.2, I-squared 0%) for the studies of CT. However, there was significant evidence of statistical heterogeneity for the studies of ultrasonography studies (positive LR, P<0.001, I-squared 85%; negative LR, P<0.001, I-squared 75%).

For the detection of appendicitis, CT had pooled estimates of 0.94 (95% confidence interval, CI: 0.91, 0.95) for sensitivity, 0.95 (95% CI: 0.93, 0.96) for specificity, 13.3 (95% CI: 9.9, 17.9) for the positive LR and 0.09 (95% CI: 0.07, 0.12) for the negative LR. Ultrasonography had pooled estimates of 0.86 (95% CI: 0.83, 0.88) for sensitivity, 0.81 (95% CI: 0.78, 0.84) for specificity, 5.8 (95% CI: 3.5, 9.5) for the positive LR and 0.19 (95% CI: 0.13, 0.27) for the negative LR.

For the four studies that directly compared CT against ultrasonography, CT demonstrated better diagnostic performance. The results of the other subgroup analyses were similar to the main analyses. The sensitivity analyses of the treatment of uninterpretable results demonstrated pooled sensitivities and specificities of, respectively, 0.90 (95% CI: 0.86, 0.92) and 0.92 (95% CI: 0.89, 0.93) for CT, and 0.84 (95% CI: 0.80, 0.86) and 0.79 (95% CI: 0.76, 0.82) for ultrasonography.

**Authors’ conclusions**

CT is probably more accurate than ultrasonography for diagnosing appendicitis in adults and adolescents.

**CRD commentary**

This review had a clear research question and specified inclusion criteria with respect to the participants, interventions and reference standard. The inclusion criteria did not specify which study designs were eligible, other than they had to be prospective studies. The search for published studies seemed appropriate, no language restrictions were applied, and
some efforts were made to find additional studies. However, no efforts were made to locate unpublished data, thereby raising the possibility of publication bias (the authors themselves stated that weaknesses in the current literature have led to an overestimation of accuracy for both techniques). Two reviewers selected studies and extracted the data, which helps minimise the risk of errors and bias. Study quality was assessed using an appropriate checklist, although the authors did not report if this assessment was also performed by two reviewers.

The authors presented pooled values for all summary statistics, even in the presence of statistically significant heterogeneity. Therefore, the pooled estimates of sensitivity, specificity, and positive and negative LRs for ultrasonography may not be reliable. The authors discussed their results in light of the limitations of the evidence, especially with respect to the poor quality and reporting of the included studies. This was a well-conducted review and the authors’ conclusions are suitably cautious given the methodological weaknesses of the available evidence.

Implications of the review for practice and research
Practice: The authors stated that the results of this review should not change the clinicians’ approach to patients with a typical clinical presentation of appendicitis, who should receive immediate surgical evaluation.

Research: The authors stated that future studies should investigate the effect of CT on patient outcomes, including negative results on laparotomy, surgical complications and perforation rates. They added that the cost-effectiveness of appendiceal imaging should also be investigated.

Bibliographic details

PubMedID
15466771

Original Paper URL
http://www.annals.org/cgi/content/full/141/7/537

Indexing Status
Subject indexing assigned by NLM

MeSH
Acute Disease; Adolescent; Adult; Appendicitis /radiography /ultrasonography; Bias (Epidemiology); Humans; Likelihood Functions; Predictive Value of Tests; Tomography, X-Ray Computed /standards; Ultrasonography /standards

AccessionNumber
12004008769

Date bibliographic record published
30/09/2006

Date abstract record published
30/09/2006

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.