CRD summary
The review assessed the performance of clinical examination and basic laboratory tests in the diagnostic evaluation of children with suspected appendicitis. The authors concluded that clinical examination does not establish diagnosis, but may be useful to select children for immediate surgical assessment or further testing. These conclusions are likely to be reliable.

Authors’ objectives
To assess the precision and accuracy of symptoms, signs and laboratory test results in diagnosing children with suspected appendicitis.

Searching
MEDLINE was searched from 1966 to March 2007; the search terms were reported. The bibliographies of retrieved articles and reviews, medical textbooks and the Cochrane Library were searched for additional articles.

Study selection
Specific interventions included in the review
Studies of medical history, physical examination or basic laboratory tests were eligible for inclusion. Studies assessing combinations of signs and symptoms (diagnostic algorithms) were included if these were assessed in a different population to the one in which the algorithm was derived. The included studies evaluated a range of pain and gastrointestinal signs and symptoms, scoring systems and laboratory inflammatory markers; full details were reported.

Reference standard test against which the new test was compared
Studies in which the diagnosis was confirmed by surgical pathological findings, clinical observation or follow-up were eligible for inclusion.

Participants included in the review
Studies of children (18 years or younger), in whom the diagnosis of appendicitis was considered, were eligible for inclusion. The study participants covered the full age range (0 to 18 years).

Outcomes assessed in the review
Inclusion criteria were not specified for the study outcomes. The outcomes reported in the review were the sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR-, respectively).

How were decisions on the relevance of primary studies made?
Three authors independently reviewed the titles and abstracts of retrieved articles. Articles selected by at least two authors were retrieved for full-text analysis, while articles selected by one reviewer were re-examined and selected by consensus. The authors did not state how the full-text review was undertaken.

Assessment of study quality
Three authors independently assessed the methodological quality of the included studies. Any disagreements were resolved by discussion. Studies were assigned an evidence level (1 to 5): level 1 articles were independent, blind comparisons of the index test with the specified reference standard, in 200 or more consecutive children with abdominal pain; level 2 articles were as level 1, but in less than 200 children; level 3 articles were as level 1 and 2, but in non-consecutive patients: level 4 articles were non-independent comparisons in cases and healthy controls; level 5 articles were as level 4, but using a reference standard of uncertain validity (level 4 and 5 studies were excluded from the analysis).
Data extraction
Two authors independently extracted the data from all selected articles. Details of the index test and associated 2x2 data, to calculate the sensitivity, specificity and LRs, were extracted.

Methods of synthesis
How were the studies combined?
Summary LRs were estimated using a random-effects model. Where there were only two studies in a group, ranges of LRs were reported.

How were differences between studies investigated?
The studies were stratified by index test and methodological quality level. Statistical heterogeneity was assessed (method not reported), but was not used to rule out the generation of pooled estimates.

Results of the review
A total of 42 studies met the initial inclusion criteria, of which 25 met the quality criteria and were included in the analysis. One additional study provided precision data only. The total number of participants was unclear.

In terms of study quality, the majority of the included studies (24 out of 25) were level 3; the remaining study was level 1.

Precision (one study).
Inter-examiner precision was poor; when assessed over seven clinical findings, only rebound tenderness was found to produce a kappa statistic of greater than 0.5 (k=0.54).

Accuracy of symptoms.
The presence of right lower quadrant pain had minimal impact on the likelihood of appendicitis (LR+ 1.2, 95% confidence interval, CI: 1.0, 1.5) while its absence decreased the likelihood (LR- 0.56, 95% CI: 0.43, 0.73), based on three level 3 studies.

The presence of fever increased the likelihood of appendicitis (LR+ 3.4, 95% CI: 2.4, 4.8) while its absence decreased the likelihood (LR- 0.32, 95% CI: 0.16, 0.64), based on data from the level 1 study; this was less useful when data from the four level 3 studies were considered (LR+ 1.2, 95% CI: 1.1, 1.4; LR- 0.53, 95% CI: 0.29, 0.97).

The other symptoms considered were less useful.

Accuracy of signs.
Rebound tenderness increased the likelihood of appendicitis (LR+ 3.0, 95% CI: 2.3, 3.9) while its absence decreased the likelihood (LR- 0.28, 95% CI: 0.14, 0.55), based on data from three level 3 studies.

The other signs considered were less useful.

Accuracy of symptom-sign combinations.
The Alvarado, or MANTRELS, score was the most commonly evaluated (three level 3 studies). A score of seven or higher increased the likelihood of appendicitis (LR+ 4.0, 95% CI: 3.2, 4.9) while a score of less than seven decreased the likelihood (LR- 0.20, 95% CI: 0.09, 0.41).

Accuracy of laboratory tests.
A white blood cell count of less than 10,000 per microlitre, or an absolute neutrophil count of less than 6,750 per microlitre, decreased the likelihood of appendicitis: LR- 0.22 (95% CI: 0.17, 0.30; based on four level 3 studies) and
0.06 (95% CI: 0.03, 0.16; based on one level 1 study), respectively.

The other laboratory tests considered were less useful.

**Authors’ conclusions**
Clinical examination does not establish a diagnosis of appendicitis, but can be useful in determining which children require immediate surgical assessment and which require further diagnostic assessment. Further age-stratified data are needed.

**CRD commentary**
The review addressed a clearly stated research question using appropriate inclusion criteria. The search strategy was somewhat limited and, as such, might not have retrieved all relevant data. The authors reported appropriate measures to minimise the potential for error and bias in the review process, and the methodological quality of the included studies was assessed and incorporated in the analysis. The methods used to calculate the summary statistics appear reasonable given the available data; however, the description of these methods was limited. The authors' conclusions are broadly supported by the data presented.

**Implications of the review for practice and research**
Practice: The authors stated that clinical examination does not establish a diagnosis of appendicitis, but can be useful in determining which children require immediate surgical assessment and which require further diagnostic assessment.

Research: The authors stated that future research, using prospective age-specific data from large cohorts of children with undifferentiated abdominal pain, could increase the usefulness of clinical examination.

**Funding**
Robert Wood Johnson, Clinical Scholars Program; National Institutes of Health K23 career development award, grant 1K23HD051817-01A1.

**Bibliographic details**

**PubMedID**
17652298

**DOI**
10.1001/jama.298.4.438

**Original Paper URL**
http://jama.ama-assn.org/

**Other publications of related interest**
This additional published commentary may also be of interest.


**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Abdomen, Acute /etiology; Appendicitis /diagnosis; Blood Cell Count; Blood Sedimentation; C-Reactive Protein
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.