Reviewer's report

Title: Systematic quantitative overviews of the literature to determine the value of diagnostic tests for predicting acute appendicitis

Authors:

Lucas M Bachmann (lucas.bachmann@evimed.ch)
Dominique Bischof (dbischof@gmx.ch)
Stefan A Bischofberger (knut@access.unizh.ch)
Marco Bonani (marcobonani@access.unizh.ch)
Franziska Osann (fransl@gmx.ch)
Johann Steurer (polstf@usz.unizh.ch)

Version: 1 Date: 18 Mar 2002

Reviewer: Dr Stefan Sauerland

Level of interest: A paper of limited interest

Advice on publication: Reject because scientifically unsound

This manuscript describes the methods for a systematic review on methods in the diagnosis of appendicitis. Within the last decades, the introduction of ultrasonography, computer tomography, laparoscopy and new laboratory parameters have broadened the diagnostic options for patients with suspected appendicitis. Therefore, such reviews as proposed in this manuscript are of large clinical importance. Nevertheless, this review protocol unfortunately contains several flaws with regard to clinical background, clear aim, and review methods. I am unable to value the statistical methods, but this part looks quite good.

In the introduction, the authors state that quantitative reviews on the given topic are currently lacking. This is simply not correct, as there are several diagnostic metaanalyses, which fulfill high quality standards:

- With regard to signs and symptoms of appendicitis, Wagner et al. have reported a superb meta-analysis (JAMA 1996; 276: 1589-1594). In their article, likelihood ratios were combined for all major clinical signs, such as rebound tenderness or migrating pain.
- The potential of laboratory parameters in identifying patients with or without appendicitis has been addressed by Hallan and Asberg (Scand J Clin Lab Invest 1997; 57: 373-380). In their analysis, the summary ROC curve indicated that CRP is not as helpful as previously thought.
- To evaluate the importance of ultrasonography, Orr et al. pooled 17 studies (Acad Emerg Med 1995; 2: 644-650). Orr et al. recommended ultrasound only for patients with unclear symptoms.
- Together with two colleagues, I recently published a Cochrane review on the value of diagnostic and/or therapeutic laparoscopy in patients with suspected appendicitis. Most probably, this review was not yet published at the time when this manuscript was written, but data on diagnostic laparoscopy now indicate that
the rate of unnecessary appendectomies is reduced to about one fifth in fertile women.

I know of another meta-analysis on computer tomography which was begun about a year ago and will be finished soon.

Having these important meta-analyses in mind, I fear that the authors of the submitted manuscript are not familiar with the current knowledge base in their field. Furthermore, their stated aim to review "the accuracy of all procedures in the diagnosis of appendicitis" gives me reason to believe that they do not know what enourmos work this would mean. Such an endeavour would result in not one, but a large number of meta-analyses, and therefore would be extremely time-consuming.

This urges me to recommend the authors to restrict their review to only one or two diagnostic possibilities in acute abdominal pain. For example, the available reviews on laboratory parameters and sonography need to be updated.

To publish a review protocol before the review is begun, is an idea that was introduced by the Cochrane Collaboration in order to safeguard against manipulations of meta-analyses from a posteriori. Thus, a diagnostic review protocol is worth publishing, but only if the methods are described in enough detail. However, the protocol currently is very vague in many aspects:

. The literature search performed by the authors yielded 800 citations. This shows that only a minor proportion of relevant articles has been identified. In contrast, a simple PubMed search performed by myself with the MeSH-heading "appendicitis/diagnosis" identifies 4,567 articles.

A review protocol must describe, how exactly the literature will be searched. This includes a list of search terms (MeSH's). Furthermore, the hand search of conference proceedings must be described in more detail. What happens if only a congress abstract is available? Will all languages be searched?

. The authors speak of searching DARE and the Cochrane Library. In truth, DARE (Database of Abstracts of Reviews of Effectiveness) is a part of the Cochrane Library.

Bearing this and the previous comments in mind, I fear again that the authors are not very familiar with expert literature searches. This surprises me, since the Horten centre is a well-known authority in the field of evidence-based medicine.

The study selection criteria are also unclear:

. First, a gold standard is required against which the test is compared. This reference standard will be most likely the pathologist's report concerning the resected appendix specimen, but the review protocol does not say so explicitly. What is the reference standard in those patients, who are discharged home, because appendicitis is unlikely? Are primary studies required to have a minimum follow-up to make sure that none of the patients later underwent appendectomy in another hospital? Otherwise, the rate of false-negatives will be biased. It is insufficient to mention only that the reviewers will "look for evidence of verification bias".

. Since the aim of the review is the comparison of diagnostic options, it is important to discern between studies which compared two tests against the gold standard or only one.

. In these studies with two tests (eg leukocytes and CRP), the presence of work-up bias should be assessed. Studies that measured leukocytes in all patients, but CRP only in those with borderline leukocytosis will be more likely to underestimate the comparative accuracy of CRP.

. Another foreseeable key problem of the review will be that many diagnostic tests
are correlated among each other. If we assume that CRP and leukocytes are highly correlated, both will probably have excellent statistical accuracy. Clinically, however, the additional determination of CRP will then still be worthless. Therefore, primary studies are of special value, if they compared a test "A" against a combination of "A" and "B" instead of simply comparing "A" versus "B". The inclusion of studies on prediction rules will cause similar problems, because these studies will most likely report on a combination of more than a few tests. The patient collectives are not defined. What exactly is the "appropriate spectrum of eligible patients"? Will studies be included that examined patients with acute abdomen, too? Or is the review restricted to patients with more specific pain, e.g. in the right lower quadrant?

In my opinion, it is important whether a study recruited patients consecutively or not. This should be assessed, since it may be a sign of work-up bias. Some sensitivity analyses should be defined a priori to compare the included studies in groups among each other. Due to the higher prevalence of gynaecologic diseases in young women, it seems reasonable to keep studies on women separately. Paediatric studies should be treated similarly, and shouldn't be mixed with studies on unselected patient groups. Other sensitivity analyses should address differences in study design and methods (eg blinding).

How will randomised intervention studies and observational cohort studies be combined?

Will costs be considered in the review?

Some comments on language and writing style.

It is not good style to simply repeat sentences from the abstract in the introduction.

Replace "if" with "whether" at the end of the paragraph on "study identification".

The word "manipulation" in the paragraph on "analysis" should be replaced.

Leukocytes are cells and shouldn't be summarized under the heading "Inflammatory mediators" in Table 1.

Laparoscopy is misspelled in Table 1.

The appendix mentioned on page 5 is missing.

Summary

The current manuscript needs major revisions to become acceptable. Perhaps, the methods of the review should be reported only together with the completed review. It is clear, that the review protocol requires definite changes in the following points

Focus of the review.

Literature searches.

Assessment of primary studies.

Sensitivity analyses.

Details of literature appraisal.

Level of interest:

A paper of limited interest, (but the completed review will probably contain important conclusions.)

Competing interests:

None declared.