Comparison of ultrasound, radiography, and clinical examination in the diagnosis of acute maxillary sinusitis: a systematic review

Varonen H, Makela M, Savolainen S, Laura E, Hilden J

Authors' objectives
To evaluate the efficacy of clinical examination, ultrasound and radiography as diagnostic measures for acute maxillary sinusitis (AMS) in unselected populations.

Searching
MEDLINE was searched from 1966 to April 1999, and the Finnish medical database 'Medic' from 1977. The reference lists of relevant articles were examined and four major journals in the field were handsearched from 1980. Experts in the field were also contacted. Unpublished studies were included. Articles written in English, German, French, a Scandinavian language or Finnish were accepted.

Study selection
Study designs of evaluations included in the review
Any prospective study that included a comparison between one of the diagnostic techniques being evaluated with an eligible reference test was included, provided it met certain validity criteria.

Specific interventions included in the review
The eligible diagnostic procedures for evaluation were clinical examination, ultrasound and radiography. Ultrasound findings were considered positive according to the criteria established by Revonta (see Other Publications of Related Interest no.1): scans with the presence of a back wall echo of at least 3.5 cm from the initial echo were considered positive; others were considered negative. In sinus radiographs, findings with air-fluid level, complete opacity, or mucosal thickening of at least 6 mm were considered positive. In studies of clinical examination, only data on the physician's overall impression were collected.

Reference standard test against which the new test was compared
The eligible reference standard tests were sinus puncture and computed tomography. Some studies also employed X-rays as part of their reference standard. Sinus puncture is considered to be the 'gold' standard.

Participants included in the review
The participants had to be adults with suspected AMS (symptoms for less than 3 months) in primary care or a comparable setting. Where a setting was unclear, acute patients without previous diagnostic procedures were included. Patients with chronic sinusitis or other selected populations, e.g. solely allergic or surgical patients, were excluded. The participants in the included studies were aged 10 to 75 years (two studies included children as well as adults). Where reported, the mean age was 37 or 38 years. The studies were set in general practice, ear, nose and throat clinics, military clinics and an emergency ward. Most studies included men and women, with the proportion of men varying from 30 to 81%. One study was confined to men.

Outcomes assessed in the review
Studies presenting 2x2 tables, or where the relevant data for their construction could be obtained from the authors, were included. The outcome measures used were prevalence, sensitivity, specificity, and the positive and negative likelihood ratios (LRs). In most studies, sinuses were counted rather than patients, disregarding right-left cross-dependence.

How were decisions on the relevance of primary studies made?
Two investigators independently assessed the searches to find potential study articles. The authors do not state how the papers were finally selected for the review prior to exclusion on the grounds of the critical assessment, or how many of the reviewers performed the selection.
Assessment of study quality
The methodological standards of the Cochrane Screening and Diagnostic Tests Methods Group were used to assess the validity of comparisons between a test and a reference standard. (See Other Publications of Related Interest no.2). The criteria applied were: (1) use of a valid reference standard; (2) assessment of the test and reference results blind to each other; (3) patients chosen for reference standard independently of the test results; (4) test measured without other clinical information; (5) reference tests carried out before treatment was started on the basis of test results; (6) tests compared in a valid design. The studies were excluded if they did not meet criteria 3, 5 and 6. Two investigators independently rated the methodological quality of the studies, and any disagreements resolved by consensus.

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

The authors extracted 2x2 tables or the data with which to construct them. From the original numbers, the authors calculated the sensitivity, specificity and the confidence intervals (CIs) for the diagnostic method under study. The prevalence of sinusitis and the positive and negative LR s were also calculated. The tabulated data included the setting, sample size, patients’ age, study duration, diagnostic method and reference standard.

Methods of synthesis
How were the studies combined?
Summary sensitivities, specificities and LR s were calculated, weighted by inverse variance (see Other Publications of Related Interest no.3), and were grouped according to the type of reference standard used. Summary receiver operating curves were formed by modifying the method of Moses et al., using zero as the slope parameter (See Other Publications of Related Interest nos.4-6). The Q* value, the point on the curve where sensitivity and specificity are equal, was used as a summary estimate, with coordinates (false positive, true positive fraction) of (1-Q*, Q*).

How were differences between studies investigated?
Each reference standard grouping of the studies was further categorised by the test used, and the results were summarised by test and reference standard. No formal method for testing heterogeneity was reported, although comments were made where heterogeneity was observed within the groups.

Results of the review
Nine studies (n=1,144) were included, two of which were unpublished. The number of patients in the studies ranged from 39 to 215.

Four studies employed clinical examination, with two of these also employing ultrasound or X-ray. Six studies employed ultrasound, five of these in conjunction with X-ray and the sixth with clinical examination.

Of the nine studies included, none met all the methodological criteria. No study reported that the test was assessed blind of all clinical information (this was infeasible in two studies). Six studies met the criterion of independent measurement of the test and reference standard.

The prevalences varied from 0.26 to 0.80, with the majority in the range 0.5 to 0.6. The lowest prevalences were found in general practice, where the lowest sensitivities were also found (0.54 to 0.63). The sensitivities ranged from 0.61 (specificity 0.98) to 0.93 (specificity 0.62) for radiography, from 0.54 (specificity 0.94) to 0.94 (specificity 0.72) for ultrasound, and from 0.44 (specificity 0.66) to 0.48 (specificity 0.75) for clinical examination.

The summary weighted mean sensitivities and specificities were, respectively:

for radiography (7 studies, 996 sinuses), 0.87 (95% CI: 0.85, 0.88) and 0.89 (95% CI: 0.88, 0.91); for ultrasound (7 studies, 940 sinuses), 0.85 (95% CI: 0.84, 0.87) and 0.82 (95% CI: 0.80, 0.83); and for clinical examination (2 studies, 245 sinuses), 0.69 (95% CI: 0.65, 0.73) and 0.79 (95% CI: 0.75, 0.82). It was observed that there were differences in the sensitivity of radiography between hospital-based studies and those in primary care; the sensitivity was lower in
primary care.

The summary weighted positive and negative LRs were, respectively:

for radiography, 3.36 (range: 2.43 to 29.83) and 0.26 (range: 0.11 to 0.52);

for ultrasound, 2.78 (range: 1.30 to 9.94) and 0.30 (range: 0.08 to 0.74); and

for clinical examination, 3.25 (range: 3.13 to 3.45) and 0.40 (range: 0.32 to 0.43).

The Q* point from the summary receiver operating curve where the sensitivity and specificity were equal was 0.82 (95% CI: 0.78, 0.85) for radiography, 0.80 (95% CI: 0.76, 0.83) for ultrasound, and 0.75 (95% CI: 0.58, 0.86) for clinical examination.

Only one study used computed tomography as the reference standard. This was a general practice study using clinical examination with a sensitivity of 0.85 (95% CI: 0.82, 0.89), specificity of 0.23 (95% CI: 0.18, 0.27), positive LR of 1.10, and negative LR of 0.64.

Authors' conclusions

Compared with sinus puncture, ultrasound was slightly less accurate than radiography and its results were more heterogeneous. Clinical examination was a rather less reliable method for diagnosing AMS, even in the hands of experienced specialists. Ultrasound or radiography improves the accuracy of diagnosis and should be used if a correct diagnosis is considered important, i.e. if antibiotic treatment is effective. The diagnosis of AMS is rarely studied in primary care, and results from secondary care are not safely generalisable.

CRD commentary

The review question was clear with well-defined inclusion criteria. The search strategy was adequate, although we are unfamiliar with the Finnish database (Medic) searched in addition to MEDLINE; no comment can, therefore, be made on whether it was likely to be an adequate source of relevant references. Good attempts were made to locate additional published and unpublished studies following the database searches. Two reviewers were involved in the screening and validity assessments, but it was not stated whether the final selection and data extraction were carried out by more than one reviewer. Thus, it is impossible to comment on the adequacy of the whole review process. The validity criteria applied were appropriate and rigorous. Details of the studies were reported sufficiently well. The authors commented on the lack of detail on important factors, such as symptoms, in the individual study reports.

It is arguable whether the sensitivities and specificities should have been pooled in the presence of heterogeneity. Formal testing of heterogeneity should have been carried out before pooling. It would appear that the primary care studies should have been pooled separately to the clinic- and hospital-based studies. However, the analysis was careful and the discussion of its limitations was thorough. The authors noted the possibility of publication bias, and the possible effects of different settings and prevalences. They also noted the problem of interdependent observations when a sinus is the unit of observation.

The authors’ conclusion that radiography is slightly superior to ultrasound did not appear to be based on a large difference. It would appear that there might be no real difference between them, particularly when different settings are taken into account. While the conclusion that clinical examination is less reliable appears justified, the estimate of its Q* statistic (sensitivity and specificity) is based on relatively little evidence, and so has a wide confidence interval that includes the other Q* estimates. Further studies of clinical examination are perhaps needed to justify the authors' stated implication for practice.

Implications of the review for practice and research

Practice: The authors state that, if a correct diagnosis is considered important, i.e. if antibiotic treatment is found to be effective despite claims to the contrary, ultrasound or radiography should be used as diagnostic aids.
Research: The authors state that future comparative trials should preferably combine diagnosis and treatment, evaluating the two aspects of clinical management as a unit.

Bibliographic details

PubMedID
11004420

Other publications of related interest

This additional published commentary may also be of interest. Diagnosing acute sinusitis. Bandolier 2001;83:5.

Indexing Status
Subject indexing assigned by NLM

MeSH
Acute Disease; Adult; Humans; Maxillary Sinusitis /diagnosis /radiography /ultrasonography; Punctures; ROC Curve; Sensitivity and Specificity; Tomography, X-Ray Computed

AccessionNumber
12000001917

Date bibliographic record published
28/02/2003

Date abstract record published
28/02/2003

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.