Abdominal symptoms: do they predict gallstones? A systematic review

Berger M Y, van der Velden J J, Lijmer J G, de Kort H, Prins A, Bohnen A M

Authors' objectives
To evaluate the diagnostic accuracy of abdominal symptoms for gallstones in studies using ultrasonography or oral cholecystography as the reference standard, and to assess the extent to which variability in diagnostic accuracy is explained by patient selection and other characteristics of the study design.

Searching
MEDLINE was searched from January 1966 to May 1998 for papers published in English, French, Dutch or German. Combinations of the MeSH terms 'cholelithiasis' and 'abdominal pain' were used with 'symptom', 'sign', and 'complaint' as textwords. Additional references were obtained from the bibliographies of review articles and original papers.

Study selection

Study designs of evaluations included in the review
No a priori inclusion criteria relating to the study design were reported. Studies that did not use original data were excluded.

Specific interventions included in the review
The following symptoms and signs were evaluated for their diagnostic ability: upper abdominal pain; biliary colic; pain radiating to the back or shoulder; the use of analgesics to ease the pain; tenderness of the upper right quadrant of the abdomen; food intolerance; and fat intolerance.

Reference standard test against which the new test was compared
Ultrasonography or oral cholecystography. The reader of the reference standard had to be unaware of the symptoms of the patient. If independent reading of the reference standard was not described, the reviewers assumed no independence.

Participants included in the review
Studies that included patients who were at extraordinary risk of gallstones, or patients aged less than 18 years, were excluded. Studies that examined patients living in non-comparable circumstances (e.g. Pima Indians) or patients with extreme conditions selected as a study group (e.g. only patients with diabetes mellitus) were also excluded. The included studies were divided into two groups.

1. Studies based in the general population in which a random sample of the population was invited for gallstone screening. This was defined as 'no disease'.

2. Hospital-based studies in which patients were referred for gallbladder investigation of abdominal symptoms. This was defined as either 'mild disease when it included elective referral, or 'serious disease' when it included emergency referrals or hospitalised patients.

Only patients with intact gallbladders were included in the estimates of the prevalence of gallstones. Patients with a history of cholecystectomy were excluded. The age range for the included participants was 14 to 78 years in the general population screening studies, and 10 to 90 years (where reported) in the hospital-based studies.

Outcomes assessed in the review
The prevalence of gallstones and the diagnostic accuracy of abdominal symptoms in gallstones were assessed. Studies that did not report the true positive and false positive rates were excluded if the rates could not be calculated from the data presented.

How were decisions on the relevance of primary studies made?

Database of Abstracts of Reviews of Effects (DARE)
Produced by the Centre for Reviews and Dissemination
Copyright © 2017 University of York
Two independent readers reviewed the abstracts of all publications, blinded to the source and authors of the publication. Consensus was reached in cases of disagreement.

**Assessment of study quality**
The authors do not report the method used to assess quality, or how the quality assessment was performed. Information relating to quality issues was presented in the text.

**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 following data were extracted from each study: reference details; patient selection method; disease spectrum (serious or mild); the number of men and women included; the age range of the participants; blinding of the radiologist; reference standard used; and the prevalence of gallstones. For each study, the true positive and false positive rates, positive and negative predictive values, and diagnostic odds ratios (DORs) were calculated from the published data.

**Methods of synthesis**

*How were the studies combined?*
Receiver operating characteristic curves were constructed by plotting the estimates of the true positive rate against the estimates of the false positive rate. The outliers were identified in these plots and excluded from further analyses. In case of statistical heterogeneity (P>0.05), a logistic regression model was used to evaluate whether the setting, spectrum of the disease, gender of the study population, blinding of the radiologist, or the reference standard, statistically significantly influenced the discriminative capacity of the abdominal symptoms (P<0.05). The models with one of these variables added were compared with a model without the variable, using generalised likelihood ratio test statistics (see Other Publications of Related Interest). Logistic regression was used to estimate a pooled DOR. Asymptotic 95% confidence intervals (CIs) around the DOR were calculated.

*How were differences between studies investigated?*
Homogeneity was tested using the chi-squared statistic. The tests were performed separately for the true positive rates and false positive rates.

**Results of the review**

Twenty-four studies with a total of 36,302 patients: 14 studies based in the general population (n=31,993), and 10 hospital-based studies (n=4,309). Of the hospital-based studies, 4 selected patients with 'mild disease' and 6 selected patients with 'serious disease'.

Over 13% (4,891 out of 36,302) of the patients had gallstones. The prevalence of gallstones depended on the setting of the study: the prevalence was 12.4% (3,967 out of 31,993) in the general population, and 21% (924 out of 4,309) in a referred (to hospital) population. In both settings, the prevalence varied widely between studies (5 to 35% and 11 to 50%, respectively).

The positive and negative predictive values, respectively, for each symptom in the general population were: 23 and 90 for upper abdominal pain; 17 and 92 for biliary colic; 10 and 93 for radiating pain; 24 and 89 for the use of analgesics; 9 and 92 for food intolerance; and 16 and 86 for fat intolerance.

The positive and negative predictive values, respectively, for each symptom in the referred population (hospital-based) were: 30 and 79 for upper abdominal pain; 50 and 81 for biliary colic; 38 and 81 for radiating pain; 50 and 81 for the use of analgesics; 40 and 73 for tenderness of the upper abdomen; 23 and 82 for food intolerance; and 30 and 79 for fat intolerance.

The unadjusted pooled DORs for all the symptoms studied were low. The DOR was 1.7 (95% CI: 1.5, 2.0) for upper abdominal pain (13 studies); 2.6 (95% CI: 2.4, 2.9) for biliary colic (9 studies); 2.8 (95% CI: 2.2, 3.7) for radiating pain.
(5 studies); 2 (95% CI: 1.6, 2.5) for the use of analgesics (4 studies); 1.8 (95% CI: 1.3, 2.4) for tenderness of the upper abdomen (3 studies); 1.3 (95% CI: 1.1, 1.6) for food intolerance (6 studies); and 1.3 (95% CI: 1.1, 1.5) for fat intolerance (11 studies).

The chi-squared test statistic showed that both the true positive rates and false positive rates were heterogeneous for all the symptoms studied (P<0.001). For the symptoms abdominal pain, biliary colic and the use of analgesics, logistic regression showed a statistically significantly different diagnostic accuracy for the different clinical settings. Different spectra of the disease showed a statistically significantly different discriminative capacity in the symptoms abdominal pain, biliary colic, and food intolerance.

Neither the gender of the studied population, the blinding of the reader to the reference standard, or the reference standard used, could explain the heterogeneity in the results. The lack of control for confounding age in most studies precluded the evaluation of age as an explanation for the discrepancies in the results.

The symptoms biliary colic, radiating pain and analgesics used were consistently related to gallstones. The DOR of biliary colic increased with the extent of gallstone disease: the DOR was 3.5 (95% CI: 2.3, 5.3) for ‘mild disease’ and 13.3 (95% CI: 4.2, 42) for ‘serious disease’.

The methodological quality of most of the included studies was low. Only eight studies blinded the reader to the reference standard for the abdominal symptoms of the patients; only eight studies controlled for both age and gender; and in the hospital-based studies, the extent of gallstone disease in the included patients was barely described.

**Authors' conclusions**

Although biliary colic was specific for gallstones, 80% of the referred patients with gallstones presented with other abdominal symptoms. There is no current evidence that justifies the use of single abdominal symptoms, other than biliary colic, in the diagnosis of symptomatic gallstones.

**CRD commentary**

This was a reasonably well-conducted review. The objectives of the review were clearly stated. The literature search was not extensive, with only one electronic database being searched for papers in English, French, Dutch or German. This means that some important information might have been missed. No attempt was made to look for unpublished data. The studies were selected for inclusion (by more than one reviewer) in a systematic way, but it was not stated how many of the reviewers were involved in the data extraction. Relevant details of the included studies were presented in tabular format and in the text. It was not reported whether the studies had been quality assessed in a systematic way, but information relating to the quality of the included studies was presented in the text.

The pooled DOR for each symptom was presented and, in view of both the statistical and clinical heterogeneity between the studies (especially in terms of the setting), this may not have been appropriate. The effects of various sources of heterogeneity (setting, spectrum, blinding and reference standard) on the pooled results were investigated in a logistic regression analysis and clearly presented in a tabular format.

The authors' conclusions regarding the diagnostic use of biliary colic alone seem optimistic in the light of their discussion of its discriminatory capacity and the results presented.

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

Practice: The authors did not state any implications for further practice.

Research: The authors state that further research should focus on the prognosis of patients with non-specific abdominal symptoms in the clinically relevant setting of primary care. They also state that the evaluation of the complexity of the symptoms of presenting patients should be emphasised and evaluated.

**Bibliographic details**

PubMedID
10672838

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Abdominal Pain /etiology; Analgesics /therapeutic use; Cholelithiasis /complications /diagnosis; Colic /etiology; Food /adverse effects; Humans; Odds Ratio; Sensitivity and Specificity

AccessionNumber
12000000299

Date bibliographic record published
31/10/2002

Date abstract record published
31/10/2002

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.