Diagnostic accuracy of interleukin-6 and interleukin-8 in predicting severe acute pancreatitis: a meta-analysis

Aoun E, Chen J, Reighard D, Gleeson FC, Whitcomb DC, Papachristou GI

CRD summary
This review concluded that interleukin (IL)-6 and IL-8 seemed to perform at an acceptable level in predicting severe acute pancreatitis. The review process was generally well conducted, but there was potential for missed studies, limitations of the analyses employed and the included studies were generally small and unblinded. Therefore, the conclusions should be treated with some caution.

Authors' objectives
To assess the accuracy of interleukin (IL)-6 and IL-8 in predicting severe acute pancreatitis.

Searching
MEDLINE and EMBASE were searched to December 2007 for peer-reviewed articles of studies reported in English; search terms were reported. Reference lists of identified publications were searched.

Study selection
Studies that enrolled at least 10 patients to assess the accuracy of IL-6 or IL-8 compared to the Atlanta 1992 symposium criteria for predicting the severity of acute pancreatitis were eligible for inclusion. Any method for establishing IL-6 or IL-8 were eligible. No population characteristics were reported. The tests used to establish IL-6 and IL-8 were enzyme-linked immunosorbent assay (ELISA), chemiluminescence immunoassays and 7TD1 cell line assays.

Two independent reviewers selected studies for the review.

Assessment of study quality
Study quality was assessed using the QUADAS 14-criteria tool. It was unclear how many reviewers performed quality assessment.

Data extraction
Two independent reviewers extracted data to construct 2x2 tables of test performance; disagreements were resolved by discussion with a third reviewer. Sensitivity, specificity and positive and negative predictive values (PPV and NPV), likelihood ratios (LR+ and LR-) and diagnostic odds ratio (DOR) with 95% confidence intervals (CI) were calculated. For studies with a zero cell, 0.5 was added to all cells. Study authors were contacted where clarification of data was necessary or missing data was sought.

Methods of synthesis
Summary estimates of each of the diagnostic outcomes, with 95% CI, were calculated using a random-effects model. IL-6 and IL-8 were analysed separately; analyses were stratified by day post-hospitalisation on which the IL-test was performed. Heterogeneity was assessed using Cochran Q and the $I^2$ tests; $p<0.05$ indicated statistically significant heterogeneity, $I^2>75\%$ indicated substantial heterogeneity that precluded pooling. Summary receiver operating characteristic (SROC) curves were produced using the Moses-Littenberg model. The area under the curve was calculated.

Results of the review
Ten studies met the inclusion criteria. Eight studies evaluated IL-6 (n=415, range 24 to 117). Five studies evaluated IL-8 (n=299, range 38 to 117). All included studies reported selection criteria, recruited an appropriate patient spectrum and avoided verification and progression biases. None of the studies was blinded. No intermediate results or withdrawals were reported. Results for the test for heterogeneity were presented only for sensitivity and specificity; where reported, $I^2$ was over 50% on day one and reduced to 20% or less on days two and three.
IL-6:

Pooled estimates for day one (seven studies) were: sensitivity 84% (95% CI 77% to 89%), specificity 76% (95% CI 70% to 81%), positive predictive value 68% (95% CI 61% to 75%), negative predictive value 88% (95% CI 83% to 92%), LR+ 3.43 (95% CI 2.51 to 4.60) and DOR 13.5 (95% CI 7.0 to 25.8).

Pooled estimates for day two (four studies) were: sensitivity 72% (95% CI 60% to 81%), specificity 85% (95% CI 77% to 91%), positive predictive value 77% (95% CI 65% to 85%), negative predictive value 82% (95% CI 74% to 88%), LR+ 4.90 (95% CI 2.63 to 8.93) and DOR 23.8 (95% CI 7.5 to 75.8).

Pooled estimates for day three (four studies) were: sensitivity 81% (95% CI 70% to 89%), specificity 82% (95% CI 72% to 88%), positive predictive value 76% (95% CI 65% to 85%), negative predictive value 86% (95% CI 76% to 92%), LR+ 4.40 (95% CI 2.50 to 7.65) and DOR 20.3 (95% CI 7.3 to 56.3).

The area under the curve was 0.75 (95% CI 0.69 to 0.92) on day one, 0.88 (95% CI 0.57 to 0.98) on day two and 0.85 (95% CI 0.83 to 0.92) on day three.

IL-8:

Pooled estimates for day one (five studies) were: sensitivity 66% (95% CI 57% to 74%), specificity 67% (95% CI 59% to 73%), positive predictive value 55% (95% CI 46% to 63%), negative predictive value 76% (95% CI 69% to 82%), LR+ 1.96 (95% CI 1.40 to 2.72) and DOR 6.1 (95% CI 2.8 to 13.2).

Pooled estimates for day two (four studies) were: sensitivity 71% (95% CI 60% to 80%), specificity 91% (95% CI 84% to 95%), positive predictive value 86% (95% CI 76% to 93%), negative predictive value 80% (95% CI 72% to 87%), LR+ 8.15 (95% CI: typographical error in paper) and DOR 23.4 (95% CI 8.1 to 63.7).

The area under the curve was 0.73 (95% CI 0.21 to 0.97) on day one and 0.91 (95% CI 0.85 to 0.92) on day two.

Authors' conclusions

IL-6 and IL-8 seemed to perform at an acceptable level in predicting severe acute pancreatitis.

CRD commentary

The authors addressed a clear review question supported by appropriate inclusion criteria. Several relevant sources were searched. Only published, peer-reviewed articles published in English were included, so relevant studies could have been missed and publication and language biases could not be ruled out. Study selection and data extraction were conducted in duplicate; it was unclear whether similar methods to reduce error and bias were employed during assessment of study quality. Study quality was assessed with appropriate criteria. The included studies passed most of the quality criteria, but interpreters of tests were not blinded in any of the studies. No population details beyond the inclusion criteria and number recruited were reported. The authors used p<0.05 to indicate statistically significant heterogeneity, so the presence of heterogeneity may have been underestimated. Results for the tests for heterogeneity were inconsistently reported. The model used to produce SROC curves did not account for between-study variation. Positive and negative predictive values were pooled; however, there was no indication as to the prevalence of acute pancreatitis in the included studies. Visual inspection of forest plots and SROC curves showed that the studies were heterogeneous and a threshold effect may be present; therefore, the reliability and generalisability of the pooled estimates of sensitivity and specificity are uncertain.

The review process was generally well conducted, but there was potential for missed studies, limitations of the analyses employed and the included studies were generally small and unblinded. Therefore, the conclusions should be treated with some caution.

Implications of the review for practice and research

Practice: The authors stated that they were unable to recommend optimal cutoff values for IL-6 or IL-8.
Research: The authors stated that larger confirmatory studies that compared interleukins with more commonly used serum markers were required.

Funding

Bibliographic details

PubMedID
20110745

DOI
10.1159/000214191

Original Paper URL

Indexing Status
Subject indexing assigned by NLM

MeSH
Biomarkers /blood; Humans; Interleukin-6 /blood; Interleukin-8 /blood; Pancreatitis, Acute Necrotizing /blood /diagnosis; Sensitivity and Specificity

AccessionNumber
12010003826

Date bibliographic record published
21/07/2010

Date abstract record published
01/12/2010

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.