Cochrane Library

Cochrane Database of Systematic Reviews

Prognostic models for predicting the severity and mortality in people with acute pancreatitis (Protocol)

Gurusamy KS, Debray TPA, Rompianesi G

Gurusamy KS, Debray TPA, Rompianesi G.
Prognostic models for predicting the severity and mortality in people with acute pancreatitis.
Cochrane Database of Systematic Reviews 2018, Issue 5. Art. No.: CD013026.
DOI: 10.1002/14651858.CD013026.

www.cochranelibrary.com

TABLE OF CONTENTS

[Prognosis Protocol]

Prognostic models for predicting the severity and mortality in people with acute pancreatitis

Kurinchi Selvan Gurusamy[1], Thomas P A Debray[2], Gianluca Rompianesi[3]

[1]Department of Surgery, Royal Free Campus, UCL Medical School, London, UK. [2]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, Utrecht, Netherlands. [3]International Doctorate School in Clinical and Experimental Medicine, University of Modena and Reggio Emilia, Modena, Italy

Contact address: Kurinchi Selvan Gurusamy, Department of Surgery, Royal Free Campus, UCL Medical School, Royal Free Hospital, Rowland Hill Street, London, NW3 2PF, UK. k.gurusamy@ucl.ac.uk.

Editorial group: Cochrane Upper GI and Pancreatic Diseases Group.
Publication status and date: New, published in Issue 5, 2018.

Citation: Gurusamy KS, Debray TPA, Rompianesi G. Prognostic models for predicting the severity and mortality in people with acute pancreatitis. *Cochrane Database of Systematic Reviews* 2018, Issue 5. Art. No.: CD013026. DOI: 10.1002/14651858.CD013026.

ABSTRACT

This is a protocol for a Cochrane Review (Prognosis). The objectives are as follows:

The primary objective is to synthesise available evidence from external validation studies evaluating the predictive accuracy of clinical scoring systems (measured on admission and up to 48 hours following admission) for severity and mortality within six months in people with acute pancreatitis.

The secondary objective is to compare different risk thresholds of available scoring systems (i.e. the level at which the risk of severe acute pancreatitis or mortality is considered to be high) to predict severity and mortality within six months in people with acute pancreatitis.

For both objectives, we will explore differences in patient populations, length of follow-up, and study design as potential sources of between-study heterogeneity.

BACKGROUND

See Appendix 1 for a glossary of terms.

Description of the condition

The pancreas is an abdominal organ that secretes several digestive enzymes into the pancreatic ductal system before it empties into the small bowel. The pancreas also lodges the Islets of Langerhans, which secrete several hormones including insulin (NCBI 2017a). Acute pancreatitis is a sudden inflammatory process in the pancreas with variable involvement of nearby organs or other organ systems (Bradley 1993). The annual incidence of acute pancreatitis ranges from 5 to 30 per 100,000 population (Roberts 2013; Yadav 2006). The incidence of acute pancreatitis has increased over the past 10 to 20 years in both the UK and the USA (Roberts 2013; Yang 2008). Acute pancreatitis is the most common gastrointestinal (digestive tract) reason for hospital admission in the USA (Peery 2012), and gallstones and alcohol are the two main causes. Approximately 50% to 70% of cases of acute pancreatitis are caused by gallstones (Roberts 2013; Yadav 2006); these slip into the common bile duct and obstruct the ampulla of Vater (a common channel formed by the union of common bile duct and

pancreatic duct), resulting in obstruction to the flow of pancreatic enzymes and leading to activation of trypsinogen within the pancreas and acute pancreatitis (Sah 2013). Advanced age, male sex, and lower socioeconomic class are associated with higher incidence of acute pancreatitis (Roberts 2013).

Clinicians generally diagnose acute pancreatitis when at least two of the following three features are present (Banks 2013).

1. Acute onset of a persistent, severe, epigastric pain, often radiating to the back.
2. Serum lipase activity (or amylase activity) at least three times greater than the upper limit of normal.
3. Characteristic findings of acute pancreatitis on contrast-enhanced computed tomography (CECT) and, less commonly, magnetic resonance imaging (MRI) or transabdominal ultrasonography.

Depending upon the type of inflammation, acute pancreatitis can be classified as interstitial oedematous pancreatitis (diffuse (widespread) or occasionally localised enlargement of the pancreas due to inflammatory oedema, as seen on CECT) or necrotising pancreatitis (necrosis involving the pancreas, peripancreatic tissues, or both) (Banks 2013). Approximately 90% to 95% of people with acute pancreatitis have interstitial oedematous pancreatitis, and the remainder have necrotising pancreatitis (Banks 2013). Necrotising pancreatitis may be sterile or infected (Banks 2013). Various theories exist as to how pancreatic and peripancreatic tissues become infected. These include spread from blood circulation, lymphatics, bile, and the small bowel (duodenum) through the pancreatic duct, as well as movement (translocation) through the large bowel wall (Schmid 1999).

Local complications of acute pancreatitis include acute peripancreatic fluid collection, pancreatic pseudocyst, acute necrotic collection, and walled-off necrosis (Banks 2013). Systemic complications of acute pancreatitis include worsening of preexisting illnesses such as heart or chronic lung disease (Banks 2013). Rates of mortality following an attack of acute pancreatitis are between 2% and 20% (Munigala 2016; Roberts 2013; Yadav 2006), according to severity. Acute pancreatitis can be classified as mild, moderate, or severe, depending on the presence of local or systemic complications; transient organ failure involving one or more lungs, kidneys, and the cardiovascular system (heart and blood vessels) lasting up to 48 hours; or persistent failure of these organs lasting beyond 48 hours. Mild pancreatitis has the best prognosis and is associated with no local or systemic complications nor with organ failure. Moderately severe acute pancreatitis may include local or systemic complications or transient organ failure. Severe acute pancreatitis carries the worst prognosis in terms of mortality (Munigala 2016), along with persistent organ failure (Banks 2013).

According to the revised Atlanta classification, severe pancreatitis is associated with the presence of persistent organ failure (organ failure for longer than 48 hours) (Banks 2013), but according to the original Atlanta classification, severe pancreatitis is associated with the presence of either organ failure or local complications such as necrosis, pseudocyst, or abscess (Bradley 1993).

The clinical manifestation of acute pancreatitis is believed to be caused by activation of inflammatory pathways either directly by the pathologic insult or indirectly by activation of trypsinogen (an enzyme that digests protein or a protease), resulting in formation of trypsin - a protease that can break down the pancreas (Sah 2013). This activation of inflammatory pathways manifests clinically as systemic inflammatory response syndrome (SIRS) (Banks 2013; Sah 2013; Tenner 2013). Systemic inflammatory response syndrome is characterised by two or more of the following criteria (Bone 1992).

1. Temperature less than 36°C or greater than 38°C.
2. Heart rate less than 90 beats/min.
3. Respiratory rate greater than 20/min or PCO_2 less than 32 mmHg.
4. White blood cell count greater than 12,000/mm^3 or less than 4000/mm^3, or more than 10% immature (band) forms.

SIRS can cause multiple organ failure (Barie 2009; Bhatia 2004; Cuesta 2012), which in turn increases mortality rates (Guo 2014; Lytras 2008; Thandassery 2013).

Clinical scoring systems for predicting severity and mortality of acute pancreatitis

Clinical examination alone has low value in determining the severity of pancreatitis (BSG 2005). For this reason, several clinical scoring systems have been developed to assess the severity of pancreatitis at presentation or at 48 hours. These scoring systems combine multiple patient-level characteristics (so-called predictors) to assess their individual risk of producing unfavourable outcomes. For instance, the Bedside Index of Severity in Acute Pancreatitis (BISAP) score (Wu 2008), which is used to predict severe acute pancreatitis, is based on blood urea nitrogen, impaired mental status, SIRS, age, and pleural effusion.

We provide below a brief overview of available scoring systems for predicting the severity and mortality of acute pancreatitis.

1. Ranson criteria (Ranson 1974; Ranson 1977).
2. Glasgow-Imrie score (Imrie 1978).
3. Acute Physiology and Chronic Health Evaluation II (APACHE II) (Knaus 1985; Larvin 1989).
4. Simplified Acute Physiology Score (SAPS II) (Larvin 1989; Le Gall 1984).
5. Medical Research Council Sepsis Score (MRC Sepsis Score) (Elebute 1983; Larvin 1989).
6. Multiple Organ System Score (MOSS) (Taylor 2005).
7. BISAP score (Wu 2008).

Typically, these scoring systems are based on patient demographics, clinical features, laboratory parameters, or imaging modalities, as assessed on admission or within 48 hours. Below, we discuss these so-called (potential) predictors in greater detail.

Patient demographics

The main patient demographic factor included in all of the scoring systems mentioned above is age (Imrie 1978; Knaus 1985; Larvin 1989; Le Gall 1984; Ranson 1974; Ranson 1977; Wu 2008).

Clinical examination

Simple clinical examination typically involves previous history of organ failure or immunocompromise, previous history of chronic disease, temperature, blood pressure, pulse rate, respiratory rate, body mass index, conscious level, presence of peritonitis, presence of acute renal failure, presence of bleeding diathesis, and assessment of anticipated third space (interstitial space) loss (i.e. loss of intravascular fluid into interstitial space or peritoneal space) (Elebute 1983; Knaus 1985; Larvin 1989; Le Gall 1984; Ranson 1974; Ranson 1977; Taylor 2005; Wu 2008).

Blood and serum markers

Common blood and serum parameters of interest consist of blood white cell count, blood haematocrit, blood platelet count, blood glucose, blood urea nitrogen, serum creatinine, serum aspartate transaminase, serum lactate dehydrogenase, serum calcium, serum electrolytes, serum bilirubin, plasma albumin, oxygen saturation, pH, and base deficit (Elebute 1983; Imrie 1978; Knaus 1985; Larvin 1989; Le Gall 1984; Ranson 1974; Ranson 1977; Taylor 2005). These blood and serum markers are used mainly to assess the inflammatory response, adequacy of tissue perfusion, and organ failure.

Imaging modalities

Several imaging techniques can be used to assess the severity of inflammation of the pancreas and its surrounding structures, along with systemic inflammation (Balthazar 1985; Bollen 2012; Viremouneix 2007). Computed tomography (CT) scan and magnetic resonance imaging (MRI) are the modalities commonly used for assessment of severity of acute pancreatitis. Fluid collections, oedema, and necrosis (which manifest as altered signal intensity) of the pancreas, disruption of main pancreatic duct continuity, or an increase in colonic wall thickness adjacent to the pancreas are features on CT scan or MRI that are suggestive of severe pancreatitis (Balthazar 1985; Viremouneix 2007). However, it must be noted that such morphological changes may not be evident immediately (Banks 2013; BSG 2005).

Various radiological scoring rules combine multiple imaging modalities into a single (severity) score, which, in turn, can be used as a predictor of a clinical scoring system. Examples include Balthazar grade (Balthazar 1985), CT Severity Index (CTSI) (Balthazar 1990), modified CTSI (Mortele 2004), pancreatic size index (PSI) (London 1989), mesenteric oedema and peritoneal fluid score (MOP score) (King 2003), extrapancreatic (EP score) (Schroder 1985), and extrapancreatic inflammation on CT (EPIC score) (De Waele 2007).

Why it is important to do this review (importance of evidence about scoring systems)?

Timely and accurate identification of patients who are at high risk of developing severe acute pancreatitis may help to improve their individual prognosis while reducing their risk of complications. In particular, high-risk individuals can be admitted into the high-dependency care unit, can receive better monitoring, and can be given treatments such as appropriate fluid and nutritional therapy. In addition, timing of definitive treatment of individuals with gallstones - one of the important aetiological factors for acute pancreatitis - depends upon the severity of acute pancreatitis (Gurusamy 2013).

Although several scoring systems have been developed to predict severe acute pancreatitis, evidence on their predictive performance is variable and inconsistent (Bollen 2012; Gao 2015; Papachristou 2010).

Currently, no systematic review has included studies assessing the accuracy of different clinical scoring systems used to predict severity and mortality in people with acute pancreatitis. Existing systematic reviews compare selected prognostic scoring methods or do not include prediction of the severity of acute pancreatitis (Aoun 2009; Di 2016; Gao 2015; Gravante 2009; Yang 2016).

A formal synthesis of studies providing evidence on the predictive ability of the various available scoring systems will show which systems are more reliable in routine care, thereby facilitating evidence-based decision making for treatment of patients with acute pancreatitis.

OBJECTIVES

The primary objective is to synthesise available evidence from external validation studies evaluating the predictive accuracy of clinical scoring systems (measured on admission and up to 48 hours following admission) for severity and mortality within six months in people with acute pancreatitis.

The secondary objective is to compare different risk thresholds of available scoring systems (i.e. the level at which the risk of severe acute pancreatitis or mortality is considered to be high) to predict severity and mortality within six months in people with acute pancreatitis.

For both objectives, we will explore differences in patient populations, length of follow-up, and study design as potential sources of between-study heterogeneity.

METHODS

Criteria for considering studies for this review

Types of studies

We will include all prospective and retrospective longitudinal studies that evaluate the predictive accuracy of established prognostic scoring systems for severity or mortality in people with acute pancreatitis. We will consider validation studies as eligible if they apply the scoring system as originally developed to patients on admission or up to 48 hours following admission, with minimum follow-up until discharge or for 30 days. We will include studies reported as full text, those published as abstract only, and unpublished data. We will exclude studies in which researchers developed the prognostic scoring system as we generally consider these studies to provide less reliable estimates of predictive accuracy (Altman 1998; Hayden 2008), except if they contain validation of a sample independent from development data: In such cases, we will include only validation data on the independent sample or data adjusted for over-optimism via cross-validation or bootstrapping. We will also exclude studies in which investigators measured the level of the prognostic system or factors beyond 48 hours of admission in people with acute pancreatitis.

Types of participants

We will include adults with acute pancreatitis irrespective of the method used for diagnosis of acute pancreatitis or the type of acute pancreatitis diagnosed (acute interstitial oedematous pancreatitis or necrotising pancreatitis).

Types of prognostic models

We will include all types of clinical, laboratory, and radiological prognostic scoring systems mentioned above and those that we identify through our search, provided trialists have measured scores within 48 hours of admission in people with acute pancreatitis.

Outcomes to be predicted

1. Severity of acute pancreatitis (based on final diagnosis within six months) based on original Atlanta classification (Bradley 1993): presence of organ failure or local complications such as necrosis, pseudocyst, or abscess.
2. Severity of acute pancreatitis (based on final diagnosis within six months) based on revised Atlanta classification (Banks 2013): presence of persistent organ failure (organ failure for longer than 48 hours).
3. Mortality (in-hospital mortality or mortality within six months).

Search methods for identification of studies

Electronic searches

We will conduct a literature search to identify all published and unpublished studies. Through this search, we will identify potential studies for inclusion published in all languages. We will translate non-English language papers and will fully assess them for potential inclusion in the review as necessary.

We will search the following electronic databases for potential included studies.

1. Cochrane Central Register of Controlled Trials (date of search; Appendix 2).
2. MEDLINE (1946 to date of search; Appendix 3).
3. Embase (1947 to date of search; Appendix 4).
4. Science Citation Index (from inception to date of search; Appendix 5).

We will use the 'best sensitivity' prognosis filter of Wilczynski 2004, combined with updated search strings identified by Geersing 2012 for MEDLINE, and the 'Best optimization of sensitivity and specificity' prognosis filter of Wilczynski 2005 for EMBASE, to keep reference numbers to manageable levels (approximately 10,000 to 15,000 references).

We will also conduct a search of ClinicalTrials.gov (Appendix 6) and the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (Appendix 7) on the search date.

Searching other resources

We will check the reference lists of all primary studies and review articles for additional references. We will also contact authors of identified studies and will ask them to identify any other published and unpublished studies.

We will search for errata or retractions from eligible studies on Pubmed before performing analysis.

Data collection and analysis

Selection of studies

Two review authors (KG and GR) will independently screen titles and abstracts of all potential studies for inclusion identified through the searches and will code them as 'retrieve' (eligible or potentially eligible/unclear) or 'do not retrieve'. We will retrieve the full-text study reports, and two review authors (KG and GR) will independently screen them to identify studies for inclusion; we will identify and record reasons for exclusion of ineligible studies. We will resolve any disagreements through discussion. We will identify and exclude duplicates, and will collate multiple reports of the same study, so that each study rather than each report is

the unit of interest in the review. We will contact investigators of studies with unclear eligibility. We will record the selection process in sufficient detail to complete a PRISMA flow diagram and a 'Characteristics of excluded studies' table.

Data extraction and management

Data extraction was guided by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) checklist (Moons 2014). We will use a standard MS Excel-based data collection form for study characteristics and outcome data, which we will pilot on three studies eligible for inclusion. Two review authors (KG and GR) will independently extract the following study data for each validation of a scoring system.

1. **Scoring system characteristics:** description of the model based on the development study in which the model was first described.
2. **Study characteristics:** study design, total study duration, number of study centres and locations, study setting, date of study.
3. **Participants:** inclusion criteria, exclusion criteria, withdrawals, total sample size, presence of missing data, summary details about common predictors (e.g. mean age and its standard deviation, proportion of males).
4. **Predictors:** predictors used by the scoring system, ways that each predictor is defined and measured, threshold(s) of the prognostic scoring system to predict the severity of acute pancreatitis or mortality.
5. Method of diagnosis and type of acute pancreatitis (related to gallstone, alcohol, or other causes).
6. Discrimination performance of the scoring systems (e.g. quantified by the concordance statistic (c-statistic: equivalent to the area under the receiver operating characteristic curve and a measure of the predictive accuracy of a logistical regression model).
7. Calibration performance of the scoring system (how close predicted risks are to actual observed risks) (e.g. quantified by the observed:expected ratio (O:E ratio) in overall or across different risk strata).
8. Clinical utility of specific risk thresholds, quantified by their respective odds ratio, sensitivity, and specificity.
9. **Analysis:** methods used to deal with missing data; numbers of participants at low and high levels of prognostic scoring systems; methods used in assessing the C-statistic, calibration performance, and other quantities of interest. We will calculate the logit of the C-statistic (logit C) (other terms for C-statistic include area under curve (AUC) or area under receiver operating characteristics curve (AUROC)) and the standard error (SE) of the prognostic scoring system reported in the study for each of the primary outcomes of interest. If the C-statistic is not reported, we will approximate it from the distribution (Debray 2017), or we will examine association of the linear predictor (or a categorised version thereof). If the standard error of the (logit) C-statistic is unavailable, we will approximate it from reported information. Debray 2017 and Walter 2007 have recently described methods applied for this purpose. To assess calibration, we will extract or calculate the natural logarithm of the total observed versus expected events ratio Ln(O:E) and its SE, using methods described by Debray 2017. We will also obtain event counts at each specific risk score (threshold) whenever available. We will record the natural logarithm of the odds ratio (lnOR) and its SE at each different threshold of the prognostic scoring system reported in the study for each of the primary outcomes of interest. We will obtain this information from the odds ratio (OR) and 95% confidence intervals (CIs) reported in the study. If this is not available, we will calculate the unadjusted lnOR and its SE from the number of people with and without the specific primary outcome in the high-level group versus the low-level group of the prognostic score. If reported details were insufficient to allow calculation of the lnOR and its SE, but the study reports the risk ratio (RR) or the hazard ratio (HR), we will convert these to OR using methods described by Symons 2002 based on the proportion of people with the outcome among people with low-level prognostic score from the same study as the baseline risk of the outcome. In the absence of information on proportions of people with the outcome among people in the same study with low-level prognostic score, we will extract the RR or the HR with 95% CIs but will convert these to OR and 95% CIs before analysis, using the median proportion of people with the outcome from remaining studies included in the meta-analysis. If outcomes are reported at multiple time points, we will extract data for the longest period of follow-up (within six months). We will also extract true positive, true negative, false positive, and false negative data from studies, if available.
10. **Notes:** funding for study, notable conflicts of interest of study authors.

We will contact the study authors to request additional information as appropriate. We will resolve disagreements by consensus. One review author (GR) will copy across the data for 'Characteristics of included studies' and 'Characteristics of excluded studies' from the data collection form into the Review Manager 5 (RevMan 5) file (RevMan 2014). One review author (KG) will copy across the data for 'Data and analyses' from the data collection form into the RevMan 5 file. We will double-check that the data were entered correctly by comparing study reports against data presented in the systematic review.

Assessment of risk of bias in included studies

Two review authors (KG and GR) will independently assess the risk of bias for each study (without being blinded to study authors) using the 'PROBAST - A risk-of-bias tool for prediction-

modelling studies' according to the following domains relevant for validation studies (Wollf 2017).

1. Source of data.
2. Participants.
3. Outcome to be predicted.
4. Candidate predictors.
5. Sample size
6. Missing data.
7. Model performance.

We will include external validation studies only; therefore, we will not assess remaining risk of bias domains in the CHARMS (i.e. model development), results, and 'interpretation and discussion' sections.

We will grade each potential source of bias as high, low, or moderate as presented in Table 1, and we will provide a quote from the study report together with a justification for our judgement in the 'Risk of bias' tables. We will resolve disagreements by discussion and will summarise risk of bias judgements across different studies for each listed domain. We will consider studies at low risk of bias in all domains as having overall low risk of bias. When considering risk prediction ability, we will take into account risk of bias for all studies that contributed to that analysis.

Assessment of bias in conducting the systematic review

We will conduct the systematic review according to details provided in the published protocol and will report any deviations from it in the 'Differences between protocol and review' section of this review.

Dealing with missing data

We will attempt to contact investigators or study sponsors to verify key study characteristics and obtain missing numerical outcome data when possible (e.g. when a study was identified as abstract only). If we do not obtain the required information, we will use data available from the study report to estimate the LnOR, logit (C), and Ln (O:E) and their respective SEs using methods described in 'Data extraction and management' section.

Assessment of reporting biases

If we are able to pool more than 10 studies for a specific meta-analysis, we will create and examine a funnel plot to explore the presence of small-study effects, and thus the potential for selective reporting. We will use a funnel inverse variance test (with multiplicative dispersion factor) as suggested by Debray 2017a to determine the statistical significance of funnel plot asymmetry. We will consider a P value less than 0.10 to indicate statistical significance.

Measures of effect and data synthesis

We will calculate the following measures of effect to determine the ability of each scoring system to predict the severity of acute pancreatitis or mortality.

For primary outcomes of the review, we will calculate the summary C-statistic with 95% confidence interval (CI) and prediction interval (PI), summary observed versus expected events ratio with 95% CI and PI, and risk of developing the outcome at each specific threshold of the prognostic scoring system.

For secondary outcomes, we will calculate the summary odds ratio (OR) with 95% CI and PI of the outcome at high versus low levels of the prognostic risk model.

We will perform a meta-analysis only when this is meaningful (e.g. we will combine only studies with identical thresholds used to define high and low levels of the prognostic score).

We will perform a meta-analysis of the C-statistic, the O:E ratio, the OR, and the risk at each specific risk score (threshold) using the packages METAFOR in R (METAFOR 2017, R 2016). For meta-analysis of the O:E ratio and the OR, we will use log transformation. For meta-analysis of the C-statistic, we will use the logit transformation.

We will use the random-effects model as default because of anticipated clinical heterogeneity among studies. Hereby, we will adopt restricted maximum likelihood (REML) estimation and apply the Sidik-Jonkman correction when calculating confidence intervals.

We will try to estimate the calibration slope if we have information on observed versus predicted events for multiple risk strata in the validation studies.

Comparison of thresholds and different prognostic scoring systems

We will use the test for interaction presented in Altman 2003 to compare different thresholds of a specific prognostic scoring system against one another to identify the best threshold for the specific prognostic scoring system used. This involves calculating the difference and the standard error of the difference in the lnOR, logit (C), and Ln (O:E) at each threshold compared (Altman 2003). To calculate the LnOR, logit (C), and Ln (O:E) of the two thresholds compared, we will include only studies that report the LnOR, logit (C), and Ln (O:E) of the outcome at each of the two thresholds compared (i.e. the two thresholds are used in the same population). Therefore, we will compare differences in dependent samples and will calculate the pooled standard error using the formula suggested by Borenstein 2009. However, for this formula, we will need the correlation coefficient for the two thresholds, but we do not anticipate finding this in any of the identified studies. Therefore, we will use three correlation coefficients: a moderate positive correlation of +0.5, a moderate negative correlation of -0.5, and no correlation. We will interpret findings with extreme caution if study results are different for different values of the correlation coefficient. To avoid excessive reliance on a single study,

we will compare the different thresholds of a specific prognostic scoring system only when a minimum of two studies can be meta-analysed to calculate the lnOR for each of the two thresholds compared. We will also construct a calibration slope when we have found sufficient information. This will help us determine the best threshold for use in clinical practice.

We will use similar methods to compare different prognostic scoring systems (except for the calibration slope). For this purpose, we will choose the threshold of the prognostic scoring system with the highest logit (C), if different thresholds have been used for the prognostic scoring system.

Subgroup analysis and investigation of heterogeneity

We plan to carry out the following subgroup analyses regardless of heterogeneity.

1. Study design (prospective vs retrospective studies).
2. Definitions used for diagnosis of acute pancreatitis (based on revised Atlanta criteria vs other standards) (i.e. differences in population).
3. Follow-up length (until discharge or until 30 days vs beyond 30 days of follow-up).

We will perform subgroup analyses for primary outcomes. We will assess the presence of statistical heterogeneity by conducting random-effects meta-analyses and constructing the 95% prediction intervals (Riley 2011).

Sensitivity analysis

We plan to perform a sensitivity analysis by excluding studies at moderate or high risk of bias (one or more of the 'Risk of bias' domains classified as moderate or high).

Presentation of results

We will present results in the following order.

1. Severity of acute pancreatitis based on original Atlanta classification (Bradley 1993) and severity of acute pancreatitis (with final diagnosis within six months) based on revised Atlanta classification (Banks 2013) and mortality (in-hospital mortality or mortality within six months).
2. Prognostic scoring systems.
3. Thresholds.

'Summary of findings' table and interpretation of results

We will broadly use the GRADE approach and will present our results for each of the primary outcomes in a 'Summary of findings' table (i.e. a total of three 'Summary of findings' tables) (Guyatt 2011); we will adapt data from prognostic studies by following the principles suggested by Huguet 2013 and Iorio 2015. We will assess the quality of evidence in the same way as for intervention reviews (i.e. using the five GRADE considerations - study limitations (risk of bias), inconsistency of effect, imprecision, indirectness, and publication bias). We will consider an absolute increase of 20% in the proportion of people with the outcome between high and low levels of the prognostic score as clinically significant and will use these data to judge imprecision. We will justify all decisions to downgrade or upgrade the quality rating of studies by using footnotes and making comments to aid the reader's understanding of the review when necessary. The one major revision to the standard 'Summary of findings' table will be seen in the presentation of absolute values: Instead of numbers of people with the outcome in intervention and control groups, we will present numbers of people with the outcome at the high level of the prognostic score and at the low level of the prognostic score. We will use the weighted median proportion of people with the outcome in the low level of the prognostic score and the odds ratios with its 95% CI to estimate the number of people with the outcome among people with a high-level prognostic score. If sufficient data are available (outcome counts at each threshold), we will also present summary estimates of absolute risk of the outcome with 95% CI and PI for that specific threshold.

ACKNOWLEDGEMENTS

We thank Karin Dearness and Kieran Elliot, Managing Editors, Cochrane Upper Gastrointestinal and Pancreatic Diseases (UGPD) Group, for providing administrative and logistical support for the conduct of the current review. We thank the peer reviewers for their comments.

REFERENCES

Additional references

Altman 1998
Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Research and Treatment* 1998;**52**(1-3):289–303.

Altman 2003
Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;**326**(7382):219.

Aoun 2009
Aoun E, Chen J, Reighard D, Gleeson FC, Whitcomb DC, Papachristou GI. Diagnostic accuracy of interleukin-6 and interleukin-8 in predicting severe acute pancreatitis: a meta-analysis. *Pancreatology* 2009;**9**(6):777–85.

Balthazar 1985
Balthazar EJ, Ranson JH, Naidich DP, Megibow AJ, Caccavale R, Cooper MM. Acute pancreatitis: prognostic value of CT. *Radiology* 1985;**156**(3):767–72.

Balthazar 1990
Balthazar EJ, Robinson DL, Megibow AJ, Ranson JH. Acute pancreatitis: value of CT in establishing prognosis. *Radiology* 1990;**174**(2):331–6.

Banks 2013
Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus. *Gut* 2013;**62**(1):102–11.

Barie 2009
Barie PS, Hydo LJ, Pieracci FM, Shou J, Eachempati SR. Multiple organ dysfunction syndrome in critical surgical illness. *Surgical Infections* 2009;**10**(5):369–77.

Bhatia 2004
Bhatia M, Moochhala S. Role of inflammatory mediators in the pathophysiology of acute respiratory distress syndrome. *Journal of Pathology* 2004;**202**(2):145–56.

Bollen 2012
Bollen TL, Singh VK, Maurer R, Repas K, van Es HW, Banks PA, et al. A comparative evaluation of radiologic and clinical scoring systems in the early prediction of severity in acute pancreatitis. *American Journal of Gastroenterology* 2012;**107**(4):612–9.

Bone 1992
Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992;**101**(6):1644–55.

Borenstein 2009
Borenstein M, Hedges L, Higgins J, Rothstein HR. Chapter 4: Effect sizes based on means. *Introduction to Meta-analysis*. New York: John Wiley & Sons Inc, 2009:21–32.

Bradley 1993
Bradley EL 3rd. A clinically based classification system for acute pancreatitis. Summary of the International Symposium on Acute Pancreatitis, Atlanta, GA, September 11-13, 1992. *Archives of Surgery* 1993;**128**(5):586–90.

BSG 2005
Working Party of British Society of Gastroenterology, Association of Surgeons of Great Britain and Ireland, Pancreatic Society of Great Britain and Ireland, Association of Upper GI Surgeons of Great Britain and Ireland. UK guidelines for the management of acute pancreatitis. *Gut* 2005;**54 Suppl 3**:iii1–9.

Cuesta 2012
Cuesta JM, Singer M. The stress response and critical illness: a review. *Critical Care Medicine* 2012;**40**(12):3283–9.

De Waele 2007
De Waele JJ, Delrue L, Hoste EA, De Vos M, Duyck P, Colardyn FA. Extrapancreatic inflammation on abdominal computed tomography as an early predictor of disease severity in acute pancreatitis: evaluation of a new scoring system. *Pancreas* 2007;**34**(2):185–90.

Debray 2017
Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2017a
Debray TPA, Moons KGM, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: a comparison of new and existing tests. Research Synthesis Methods 2017:http://doi.wiley.com/10.1002/jrsm.1266.

Di 2016
Di MY, Liu H, Yang ZY, Bonis PA, Tang JL, Lau J. Prediction models of mortality in acute pancreatitis in adults: a systematic review. *Annals of Internal Medicine* 2016;**165**(7):482–90.

Elebute 1983
Elebute EA, Stoner HB. The grading of sepsis. *British Journal of Surgery* 1983;**70**(1):29–31.

Gao 2015
Gao W, Yang HX, Ma CE. The value of BISAP score for predicting mortality and severity in acute pancreatitis: a systematic review and meta-analysis. *PloS One* 2015;**10**(6):e0130412.

Geersing 2012
Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PloS One* 2012;**7**(2):e32844.

Gravante 2009
Gravante G, Garcea G, Ong SL, Metcalfe MS, Berry DP, Lloyd DM, et al. Prediction of mortality in acute

pancreatitis: a systematic review of the published evidence. *Pancreatology* 2009;**9**(5):601–14.

Guo 2014
Guo Q, Li A, Xia Q, Liu X, Tian B, Mai G, et al. The role of organ failure and infection in necrotizing pancreatitis: a prospective study. *Annals of Surgery* 2014;**259**(6):1201–7.

Gurusamy 2013
Gurusamy KS, Nagendran M, Davidson BR. Early versus delayed laparoscopic cholecystectomy for acute gallstone pancreatitis. *Cochrane Database of Systematic Reviews* 2013, Issue 9. DOI: 10.1002/14651858.CD010326.pub2

Guyatt 2011
Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. Grade guidelines: 1. Introduction-grade evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2011;**64**(4):383–94.

Hayden 2008
Hayden JA, Cote P, Steenstra IA, Bombardier C, Group Q-LW. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *Journal of Clinical Epidemiology* 2008;**61**(6): 552–60.

Huguet 2013
Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic Reviews* 2013;**2**:71.

Imrie 1978
Imrie CW, Benjamin IS, Ferguson JC, McKay AJ, Mackenzie I, O'Neill J, et al. A single-centre double-blind trial of trasylol therapy in primary acute pancreatitis. *British Journal of Surgery* 1978;**65**(5):337–41.

Iorio 2015
Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:h870.

King 2003
King NK, Powell JJ, Redhead D, Siriwardena AK. A simplified method for computed tomographic estimation of prognosis in acute pancreatitis. *Scandinavian Journal of Gastroenterology* 2003;**38**(4):433–6.

Knaus 1985
Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Apache ii: A severity of disease classification system. *Critical Care Medicine* 1985;**13**(10):818–29.

Larvin 1989
Larvin M, McMahon MJ. Apache-ii score for assessment and monitoring of acute pancreatitis. *Lancet* 1989;**2**(8656): 201–5.

Le Gall 1984
Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Critical Care Medicine* 1984;**12**(11):975–7.

London 1989
London NJ, Neoptolemos JP, Lavelle J, Bailey I, James D. Contrast-enhanced abdominal computed tomography scanning and prediction of severity of acute pancreatitis: a prospective study. *British Journal of Surgery* 1989;**76**(3): 268–72.

Lytras 2008
Lytras D, Manes K, Triantopoulou C, Paraskeva C, Delis S, Avgerinos C, et al. Persistent early organ failure: defining the high-risk group of patients with severe acute pancreatitis?. *Pancreas* 2008;**36**(3):249–54.

METAFOR 2017 [Computer program]
Wolfgang Viechtbauer. METAFOR: meta-analysis package for R. Version 2.0–0. Wolfgang Viechtbauer, 2017.

Moons 2014
Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine* 2014;**11** (10):e1001744.

Mortele 2004
Mortele KJ, Wiesner W, Intriere L, Shankar S, Zou KH, Kalantari BN, et al. A modified CT severity index for evaluating acute pancreatitis: improved correlation with patient outcome. *AJR: American Journal of Roentgenology* 2004;**183**(5):1261–5.

Munigala 2016
Munigala S, Yadav D. Case-fatality from acute pancreatitis is decreasing but its population mortality shows little change. *Pancreatology* 2016;**16**(4):542–50.

NCBI 2017a
NCBI. MeSH. NLM Controlled Vocabulary. Pancreas. www.ncbi.nlm.nih.gov/mesh/68010179 (accessed 9 July 2017).

Papachristou 2010
Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, et al. Comparison of BISAP, Ranson's, APACHE-II, and CTSI scores in predicting organ failure, complications, and mortality in acute pancreatitis. *American Journal of Gastroenterology* 2010;**105**(2):435–41.

Peery 2012
Peery AF, Dellon ES, Lund J, Crockett SD, McGowan CE, Bulsiewicz WJ, et al. Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology* 2012;**143** (5):1179–87.

R 2016 [Computer program]
R Foundation for Statistical Computing. R: A language and environment for statistical computing. Version 3.3.2. Vienna, Austria: R Foundation for Statistical Computing, 2016.

Ranson 1974
Ranson JH, Rifkind KM, Roses DF, Fink SD, Eng K, Spencer FC. Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology and Obstetrics* 1974;**139**(1):69–81.

Ranson 1977
Ranson JH, Pasternack BS. Statistical methods for quantifying the severity of clinical acute pancreatitis. *Journal of Surgical Research* 1977;**22**(2):79–91.

RevMan 2014 [Computer program]
Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2011
Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d:549.

Roberts 2013
Roberts SE, Akbari A, Thorne K, Atkinson M, Evans PA. The incidence of acute pancreatitis: impact of social deprivation, alcohol consumption, seasonal and demographic factors. *Alimentary Pharmacology and Therapeutics* 2013;**38**(5):539–48.

Sah 2013
Sah RP, Dawra RK, Saluja AK. New insights into the pathogenesis of pancreatitis. *Current Opinion in Gastroenterology* 2013;**29**(5):523–30.

Schmid 1999
Schmid SW, Uhl W, Friess H, Malfertheiner P, Buchler MW. The role of infection in acute pancreatitis. *Gut* 1999; **45**(2):311–6.

Schroder 1985
Schroder T, Kivisaari L, Somer K, Standertskjold-Nordenstam CG, Kivilaakso E, Lempinen M. Significance of extrapancreatic findings in computed tomography (CT) of acute pancreatitis. *European Journal of Radiology* 1985;**5** (4):273–5.

Symons 2002
Symons MJ, Moore DT. Hazard rate ratio and prospective epidemiological studies. *Journal of Clinical Epidemiology* 2002;**55**(9):893–9.

Taylor 2005
Taylor SL, Morgan DL, Denson KD, Lane MM, Pennington LR. A comparison of the Ranson, Glasgow, and APACHE II scoring systems to a multiple organ system score in predicting patient outcome in pancreatitis. *American Journal of Surgery* 2005;**189**(2):219–22.

Tenner 2013
Tenner S, Baillie J, DeWitt J, Vege SS, American College of Gastroenterology. American College of Gastroenterology guideline: management of acute pancreatitis. *American Journal of Gastroenterology* 2013;**108**(9):1400–15.

Thandassery 2013
Thandassery RB, Yadav TD, Dutta U, Appasani S, Singh K, Kochhar R. Dynamic nature of organ failure in severe acute pancreatitis: the impact of persistent and deteriorating organ failure. *HPB* 2013;**15**(7):523–8.

Viremouneix 2007
Viremouneix L, Monneuse O, Gautier G, Gruner L, Giorgi R, Allaouchiche B, et al. Prospective evaluation of nonenhanced MR imaging in acute pancreatitis. *Journal of Magnetic Resonance Imaging* 2007;**26**(2):331–8.

Walter 2007
Walter SD, Sinuff T. Studies reporting ROC curves of diagnostic and prediction data can be incorporated into meta-analyses using corresponding odds ratios. *Journal of Clinical Epidemiology* 2007;**60**(5):530–4.

Wilczynski 2004
Wilczynski NL, Haynes RB, Hedges T. Developing optimal search strategies for detecting clinically sound prognostic studies in Medline: an analytic survey. *BMC Medicine* 2004;**2**:23.

Wilczynski 2005
Wilczynski NL, Haynes RB. Optimal search strategies for detecting clinically sound prognostic studies in Embase: an analytic survey. *Journal of the American Medical Informatics Association* 2005;**12**(4):481–5.

Wollf 2017
Wolff R, Moons K, Riley R, Whiting P, Westwood M, Collins G, et al. PROBAST - a risk-of-bias tool for prediction-modelling studies. https://www.globalevidencesummit.org/abstracts/probast-%E2%80%93-risk-bias-tool-prediction-modelling-studies 2017 (accessed 4 March 2018).

Wu 2008
Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. *Gut* 2008;**57**(12):1698–703.

Yadav 2006
Yadav D, Lowenfels AB. Trends in the epidemiology of the first attack of acute pancreatitis: a systematic review. *Pancreas* 2006;**33**(4):323–30.

Yang 2008
Yang AL, Vadhavkar S, Singh G, Omary MB. Epidemiology of alcohol-related liver and pancreatic disease in the United States. *Archives of Internal Medicine* 2008;**168**(6):649–56.

Yang 2016
Yang YX, Li L. Evaluating the ability of the bedside index for severity of acute pancreatitis score to predict severe acute pancreatitis: a meta-analysis. *Medical Principles and Practice* 2016;**25**(2):137–42.

* *Indicates the major publication for the study*

ADDITIONAL TABLES

Table 1. Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS)

Domain	Risk of bias classification
Source of data	Low risk of bias • The study is a prospective cohort study High risk of bias • The study is a retrospective cohort study, registry data, case-control studies Moderate risk of bias • The study design was not clear
Participants	Low risk of bias (both criteria should be met) • The study includes only adult participants with symptoms lasting less than 24 hours and meeting the revised Atlanta classification of diagnosis of acute pancreatitis • No inappropriate exclusion of participants with organ failure or recurrent acute pancreatitis High risk of bias (at least one of the following criteria are met) • The study includes children or participant symptoms lasting longer than 24 hours or does not use revised Atlanta classification for diagnosis of acute pancreatitis • Participants with organ failure or recurrent acute pancreatitis were excluded Moderate risk of bias (at least one of the following criteria are met but the criteria for high risk of bias are not met) • If it was not clear whether the study includes only adult participants with symptoms lasting less than 24 hours and meeting the revised Atlanta classification of diagnosis of acute pancreatitis • If it was not clear whether participants with organ failure or recurrent acute pancreatitis were excluded
Outcome to be predicted	Low risk of bias (all criteria are met) • The outcome is defined appropriately in all participants (i.e. Atlanta classification or revised Atlanta classification is used for identifying severe pancreatitis and all-cause mortality is used for identifying mortality) • The same definition of the outcome is used for all participants • Outcome assessment was blinded to the level of the prognostic scoring system (this is not applicable for 'all-cause mortality') High risk of bias (at least one of the following criteria are met) • We had concerns about the way the outcome was defined (e.g. deaths from treatment of complications of acute pancreatitis were excluded from all-cause mortality) • Different definitions of the outcome are used in different participants (e.g. severity of pancreatitis was based on radiological findings in people with severe symptoms and based on clinical features in people with mild symptoms) • Outcome assessors were aware of the level of the prognostic scoring system (this is not applicable for 'all-cause mortality') Moderate risk of bias (at least one of the following criteria are met but the criteria for high risk of bias are not met) • The definition of the outcome is not reported • It is not clear whether the outcome is measured in the same way in all participants • It was not clear whether outcome assessors were blinded to the level of the prognostic scoring system (this is not applicable for 'all-cause mortality')

Table 1. Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) (*Continued)*

Candidate predictors	Low risk of bias (all criteria are met) • The components of the prognostic scoring system are measured using appropriate methods in all participants • The components of the prognostic scoring system are measured in the same way in all the participants • The components of the prognostic scoring system are measured blinded to the outcome • The predictors were measured within 6 hours of admission (so that prediction and treatment can be performed immediately) High risk of bias (at least one of the following criteria are met) • We had concerns about the way components of the prognostic scoring system are measured (e.g. presence of peripancreatic fluid collection was measured using ultrasound rather than computed tomography (CT scan) or magnetic resonance imaging (MRI)) • The components of the prognostic scoring system are not measured in the same way in all participants (e.g. only participants with high likelihood of severe pancreatitis underwent CT scan) • The components of the prognostic scoring system are measured without blinding to the outcome • The predictors were measured later than 6 hours of admission Moderate risk of bias (at least one of the following criteria are met but the criteria for high risk of bias are not met) • The method of measurement of components of the prognostic scoring system is not reported. • It is not clear whether components of the prognostic scoring system are measured in the same way in all participants • It is not clear whether components of the prognostic scoring system are measured with blinding to the outcome • It was clear that predictors were measured within 48 hours, but the timing of measurement was highly variable (e.g. some patients had the measurement on admission, and others had the measurement after 6 hours of admission) or if the exact timing of measurement within the 48 hours was not reported clearly.
Sample size	Low risk of bias • Minimum of 100 people with events and 100 people without events High risk of bias • Less than 100 people with events or 100 people without events
Missing data	Low risk of bias (at least one of the following criteria are met) • The study includes all participants who meet the eligibility criteria (i.e. there was no loss to follow-up or excluded because of lack of measurement of components of the prognostic scoring system) • There were no important differences in the people in whom the prognostic scoring system was measured and were followed-up adequately versus those in whom the prognostic scoring system was not measured or those were lost to follow-up High risk of bias • There were important differences in the people in whom the prognostic scoring system was measured and were followed-up adequately versus those in whom the prognostic scoring system was not measured or those were lost to follow-up Moderate risk of bias • If it was not clear whether the prognostic scoring system was measured in all eligible people or whether people were lost to follow-up

Table 1. Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) (*Continued)*

Model performance	Low risk of bias (both criteria should be met) • Calibration measures (observed vs expected or predicted events), discrimination measures (C-statistic), and classification measures (diagnostic odds ratio) are reported or can be calculated from the data available in the report • The threshold used for calculating diagnostic odds ratio was decided a priori High risk of bias (at least one of the following criteria are met) • One or more of calibration measures, discrimination measures, and classification measures are not reported and cannot be calculated from the data available in the report The threshold used for calculating diagnostic odds ratio was decided by 'optimal cut-off' method Moderate risk of bias • It is not clear whether the threshold used for calculating diagnostic odds ratio was decided a priori

APPENDICES

Appendix 1. Glossary of terms

Acute: sudden
Aetiological: the factor that causes a particular disease, in this case, acute pancreatitis
Amylase: a digestive enzyme which is increased in many people with acute pancreatitis, though this can be elevated in normal people and other diseases
Bile: a fluid secreted by liver and stored in the gallbladder; it has digestive functions and also removes some toxins from the body
Bleeding diathesis: bleeding disorders which decreases the ability of the blood to clot
Colonic: of the large bowel
Contrast-enhanced computed tomography (CECT): a CT scan performed with a dye to improve the differentiation of different tissues in the body and differentiate between normal and abnormal tissues
Enzyme: substances that enable and speed up chemical reactions that are necessary for the normal functioning of the body
Epigastric pain: upper central abdominal pain
Epigastric: upper central abdomen
Heterogeneity: variability
Immunocompromise: decreased immunity (ability to fight against the bacteria and other micro-organisms)
Insulin: substance which helps regulate blood sugar
Interstitial: space in between
Lipase: a digestive enzyme which is increased in many people with acute pancreatitis, though this can be elevated in normal people and other diseases
Lymphatics: part of the circulatory system, comprising a network of (lymphatic vessels) that carry a clear fluid called lymph; the lymphatics are an important part of the immune system
Morphological: external features
Mortality: death
Necrosis: death and decomposition of living tissue usually caused by lack of blood supply but can be caused by other pathological insult
Necrotising: causing necrosis
Oedema: swelling of tissues (in this context)
Oedematous: excessive accumulation of serous fluid in the intercellular spaces of tissues
Pancreatic pseudocysts: fluid collections in the pancreas or the tissues surrounding the pancreas, surrounded by a well defined wall and contain only fluid with little or no solid material

Pancreatitis: inflammation of the pancreas
Pathologic insult: substance or mechanism that causes the condition
Perfusion: blood circulation
Peripancreatic tissues: tissues surrounding the pancreas
Peritonitis: inflammation of the peritoneum, the inner lining of the abdominal wall: this can occur because of various reasons including inflammation of an abdominal organ and rupture of an abdominal organ
Pleural effusion: collection of fluid around the lungs
Prognostic: to predict the likely outcome
Protease: an enzyme that digests protein
Pseudocyst: a fluid-filled cavity that resembles a cyst but lacks a wall or lining
Sensitivity: ability of the test to identify that people with disease correctly
Sepsis: life-threatening illness due to blood infection with bacteria, fungus, or virus
Serum: clear fluid that separates out when blood clots
Specificity: ability of the test to identify that people without disease correctly
Transabdominal ultrasonography: standard abdominal ultrasound (sound waves not audible to the ear)
Transient: temporary

Appendix 2. CENTRAL search strategy

#1 MeSH descriptor: [Pancreatitis] explode all trees and with qualifier(s): [Complications - CO, Mortality - MO]
#2 (acute near/3 pancrea*)
#3 #1 or #2
#4 MeSH descriptor: [Incidence] explode all trees
#5 MeSH descriptor: [Mortality] explode all trees
#6 MeSH descriptor: [Follow-Up Studies] explode all trees
#7 prognos* or predict* or course*
#8 MeSH descriptor: [ROC Curve] explode all trees
#9 Stratification or Discrimination or Discriminate or C-statistic or "c statistic" or "Area under the curve" or AUC or Calibration or Indices or Algorithm or Multivariable
#10 #4 or #5 or #6 or #7 or #8 or #9
#11 #3 and #10

Appendix 3. MEDLINE search strategy

1. Pancreatitis/co, mo
2. (acute adj3 pancrea*).ti,ab.
3. 1 or 2
4. incidence.sh.
5. exp mortality/
6. follow-up studies.sh.
7. prognos*.tw.
8. predict*.tw.
9. course*.tw.
10. exp ROC Curve/
11. (Stratification or Discrimination or Discriminate or C-statistic or "c statistic" or "Area under the curve" or AUC or Calibration or Indices or Algorithm or Multivariable).ti,ab.
12. 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. exp animals/ not humans.sh.
14. 12 not 13
15. 3 and 14

Appendix 4. Embase search strategy

1. exp acute pancreatitis/co
2. (acute adj3 pancrea*).ti,ab.
3. 1 or 2
4. follow-up.mp.
5. prognos*.tw.
6. ep.fs.
7. 4 or 5 or 6
8. 3 and 7

Appendix 5. Science Citation Index search strategy

1 TS=(acute near/3 pancrea*)
2 TS=(prognos* OR predict*)
3 #2 AND #1

Appendix 6. ClinicalTrials.gov search strategy

Condition: acute pancreatitis

Appendix 7. WHO ICTRP search strategy

Condition: acute pancreatitis

CONTRIBUTIONS OF AUTHORS

KG and TB wrote the protocol and will perform the analysis. KG and GR will select studies and extract data. KG will write the first draft of the review.

DECLARATIONS OF INTEREST

KSG: none known.

TD: none known.

GR: none known.

SOURCES OF SUPPORT

Internal sources

- University College London, Other.
Computer, online library

External sources

- No sources of support supplied