Cochrane Library

Cochrane Database of Systematic Reviews

Histologic scoring indices for evaluation of disease activity in Crohn's disease (Review)

Novak G, Parker CE, Pai RK, MacDonald JK, Feagan BG, Sandborn WJ, D'Haens G, Jairath V, Khanna R

Novak G, Parker CE, Pai RK, MacDonald JK, Feagan BG, Sandborn WJ, D'Haens G, Jairath V, Khanna R.
Histologic scoring indices for evaluation of disease activity in Crohn's disease.
Cochrane Database of Systematic Reviews 2017, Issue 7. Art. No.: CD012351.
DOI: 10.1002/14651858.CD012351.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Methodology Review]

Histologic scoring indices for evaluation of disease activity in Crohn's disease

Gregor Novak[1,2], Claire E Parker[3], Rish K Pai[4], John K MacDonald[5,6], Brian G Feagan[5,6,7], William J Sandborn[8], Geert D'Haens[9,10], Vipul Jairath[6,7], Reena Khanna[3,6]

[1]Department of Gastroenterology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. [2]Department of Gastroenterology and Hepatology, University Medical Centre, Ljubljana, Slovenia. [3]Robarts Clinical Trials, London, Canada. [4]Department of Pathology, Mayo Clinic, Scottsdale, AZ, USA. [5]Cochrane IBD Group, Robarts Clinical Trials, London, Canada. [6]Department of Medicine, University of Western Ontario, London, Canada. [7]Department of Epidemiology and Biostatistics, University of Western Ontario, London, Canada. [8]Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA. [9]Academic Medical Center, Amsterdam, Netherlands. [10]Robarts Clinical Trials, Amsterdam, Netherlands

Contact: Reena Khanna, Department of Medicine, University of Western Ontario, London, ON, Canada. reena.khanna@robartsinc.com.

Editorial group: Cochrane IBD Group.
Publication status and date: New, published in Issue 7, 2017.

Citation: Novak G, Parker CE, Pai RK, MacDonald JK, Feagan BG, Sandborn WJ, D'Haens G, Jairath V, Khanna R. Histologic scoring indices for evaluation of disease activity in Crohn's disease. *Cochrane Database of Systematic Reviews* 2017, Issue 7. Art. No.: CD012351. DOI: 10.1002/14651858.CD012351.pub2.

ABSTRACT

Background

Histologic assessment of mucosal disease activity has been increasingly used in clinical trials of treatment for Crohn's disease. However, the operating properties of the currently existing histologic scoring indices remain unclear.

Objectives

A systematic review was undertaken to evaluate the development and operating characteristics of available histologic disease activity indices in Crohn's disease.

Search methods

Electronic searches of MEDLINE, EMBASE, PubMed, and the Cochrane Library (CENTRAL) databases from inception to 20 July 2016 were supplemented by manual reviews of bibliographies and abstracts submitted to major gastroenterology meetings (Digestive Disease Week, United European Gastroenterology Week, European Crohn's and Colitis Organisation).

Selection criteria

Any study design (e.g. randomised controlled trial, cohort study, case series) that evaluated a histologic disease activity index in patients with Crohn's disease was considered for inclusion. Study participants included adult patients (≥ 16 years), diagnosed with Crohn's disease using conventional clinical, radiographic or endoscopic criteria.

Data collection and analysis

Two authors independently reviewed the titles and abstracts of the studies identified from the literature search. The full text of potentially relevant citations were reviewed for inclusion and the study investigators were contacted as needed for clarification. Any disagreements regarding study eligibility were resolved by discussion and consensus with a third author.

Two authors independently extracted and recorded data using a standard form. The following data were recorded from each eligible study: number of patients enrolled; number of patients per treatment arm; patient characteristics: age and gender distribution; description of

histologic disease activity index utilized; and outcomes such as content validity, construct validity, criterion validity, responsiveness, intra-rater reliability, inter-rater reliability, and feasibility.

Main results

Sixteen reports of 14 studies describing 14 different numerical histological indices fulfilled the inclusion criteria.

Inter-rater reliability was assessed in one study. For the Naini and Cortina Score, estimates of correlation were 'almost perfect', ranging from r = 0.94 to 0.96. The methodological quality of this study with respect to reliability was 'good'.

With respect to validity, correlation estimates between various histological scoring systems and Crohn's disease activity as measured by objective markers of inflammation (including C-reactive protein, erythrocyte sedimentation rate, fecal calprotectin and fecal lactoferrin); endoscopic disease activity scores; clinical disease activity scores; and quality of life questionnaires were reported. Comparisons between histologic scoring indices and endoscopic scoring indices ranged from no correlation to 'substantial' (r = 0.779). The methodological quality of the studies that explored validity ranged form 'poor' to 'good'.

Responsiveness data were available in seven studies. After subjects were administered a treatment of known efficacy, statistically significant change in the index score was demonstrated in five studies with respect to six indices. Two studies failed to indicate whether there was statistically significant change in the index score post-treatment. With regard to methodological quality, six of the studies were rated as 'poor' and one of the studies was rated as 'fair'.

Feasibility was assessed by one study. The Naini and Cortina Score was shown to be simple to use and feasible for every given case.

Authors' conclusions

Currently there is no fully validated histological scoring index for evaluation of Crohn's disease activity. Development of a validated histological scoring index for Crohn’s disease is a clinical and research priority.

PLAIN LANGUAGE SUMMARY

Histologic measurement tools for evaluation of disease activity in Crohn’s disease

What is Crohn's disease?

Crohn’s disease is a life-long (chronic), inflammatory disease of the gastrointestinal tract characterized by abdominal pain (cramping), rectal bleeding, diarrhoea, weight loss, and tiredness. The disease has a changing course with periods of symptoms (called 'active' disease or relapse) and periods without symptoms (called 'remission').

What is a histological scoring index?

A histological scoring index measures disease activity based on the examination of biopsy specimens from the bowel (small pieces of tissue removed from the bowel) under a microscope. Biopsy specimens are removed from the bowel during colonoscopy (a non-surgical procedure used to view the digestive tract using a camera) with biopsy forceps (instruments to grasp and remove pieces of tissue). Biopsies are then processed and assessed under microscope by a pathologist (a physician who interprets and diagnoses the changes caused by disease in tissues) who then rates disease activity using the index.

What did the researchers investigate?

It is important that histological scoring indices measure what they are supposed to measure (validity); that they detect change after treatment (responsiveness); that the scores are consistently reproducible (reliability); and that they can be easily utilized (feasibility). The researchers investigated whether studies have assessed the validity, responsiveness, reliability and feasibility of histological scoring indices.

What did the researcher find?

The researchers found that none of the existing histological scoring indices have been fully validated.

BACKGROUND

Crohn's disease is a chronic, systemic disorder involving immune-mediated inflammation of the gastrointestinal tract. While treatment goals have historically focused on achieving clinical remission, it is now understood that subclinical inflammation is capable of persisting in the absence of symptoms. Consequently, this inflammation can lead to progressive bowel damage and complications such as stricturing and penetrating disease (Arnott 2002; Hommes 2012). With the advent of more potent therapies such as immunosuppressives and biologics, treatment aims have advanced beyond symptom resolution to healing of the involved intestinal mucosa (D'Haens 2009, Bouguen 2013). However, the degree of mucosal healing necessary to prevent future complications from progressive bowel damage has yet to be defined. Previous research has established that the presence of endoscopic healing does not necessarily correlate with an absence of histologic inflammation, as up to one-third of biopsies from patients with Crohn's disease with endoscopically healed mucosa demonstrate evidence of ongoing histologic disease (Korelitz 1984; Molander 2013).

In order to evaluate the benefit of histologic mucosal healing as a treatment target in clinical trials, a scoring index is necessary to objectively quantify the degree of histologic inflammation. The development of such a validated index in Crohn's disease is challenging for several reasons. Not only does the potential exist for endoscopic biopsy sampling error due to the segmental, transmural nature of intestinal inflammation in Crohn's disease, but additionally, limited data exist regarding which histologic features are most relevant for assessing active inflammation in Crohn's disease (Mojtahed 2014). As such, the operating properties of the index must be rigorously demonstrated before it can be recommended for use in clinical trials. These operating properties include validity (the extent to which an instrument truly measures the outcome that it is intended to assess), responsiveness (the ability to detect a meaningful change in health status), reliability (the consistency or reproducibility of an instrument), and feasibility (the ease with which an instrument can be utilized in a given setting) (Kirshner 1985).

The first histologic disease activity index used in clinical trials for inflammatory bowel disease was the Truelove and Richards' Index which was developed for ulcerative colitis in 1956 (Truelove 1956). Since then, several Crohn's disease-specific histologic indices have been described, which are broadly categorized as either stepwise or numerical. Stepwise instruments separate disease activity into categories such as mild, moderate, and severe. Numerical instruments assign a point scale to individual biopsy findings which can then be summed to determine an overall score (Mojtahed 2014).

The features assessed for each biopsy specimen vary according to the histologic disease activity index. Examples of such features include acute and chronic inflammatory cell infiltrates, epithelial damage, crypt distortion, and the presence of granulomas, as well as the location from which the biopsy was collected. For example, the index developed by Ward and Webb in 1977 is based exclusively on rectal biopsies in Crohn's patients (Ward 1977), as compared to the instrument utilized by Colombel 2006 which is limited to ileal biopsies. The more widely used Global Histologic Disease Activity Score (GHAS) (D'Haens 1999), and the Naini and Cortina Score (Naini 2012), are numerical instruments which allow for separate grading of both ileal and colonic specimens.

With such variation among the histologic indices for Crohn's disease, the operating properties of these instruments require evaluation to identify the most reliable and accurate index for measuring histologic disease activity.

Why it is important to do this review

There are limited data available on the operating properties of indices used for the evaluation of histologic activity in Crohn's disease. This review examines the relative merits of each available index and identifies areas where further research is needed to develop an optimal evaluative instrument for use in clinical trials.

OBJECTIVES

The primary objective was to systematically review the current literature describing the development and operating characteristics of available histologic disease activity indices in Crohn's disease.

METHODS

Criteria for considering studies for this review

Types of studies

Any study design (e.g. randomised controlled trial, cohort study, case series) that evaluates a histologic disease activity index in patients with Crohn's disease was considered for inclusion. Study subjects included adult patients (≥ 16 years), diagnosed with Crohn's disease using conventional clinical, radiographic, or endoscopic criteria.

Types of data

Histologic scoring data obtained from eligible studies were considered for inclusion.

Types of methods

The methods used to construct and validate the histologic disease activity index (e.g. validity, responsiveness, reliability and feasibility) were examined in detail and described for each included study. The number of pathologists who scored the histologic index in each study was also reported, in addition to information on whether the pathologists were blinded from the other raters' scores and the participants' clinical and endoscopic disease severity.

Types of outcome measures

Reliability: Measures of reliability (intra-rater reliability, inter-rater reliability, test-retest reliability, or internal consistency) were assessed by evaluating correlation estimates, as expressed by the interclass correlation coefficient (ICC), kappa statistic or Pearson's r.

Validity: Each study was assessed to determine whether validity was measured, which is broadly defined as evidence that variations in Crohn's disease activity causally produce variations in the index measurement outcomes. Studies were reviewed for evidence of content validity, criterion validity, and construct validity for available histologic disease activity scores.

Content validation refers to a demonstration that the elements of the histology scoring index are sufficient to measure disease activity in Crohn's disease. Content validity is generally based on qualitative assessment. For example, evidence of content validity may include an expert panel opinion on the relevant components to include in an index, or a systematic review of the literature supporting index development.

Criterion validity refers to the degree to which index scores adequately reflect true Crohn's disease activity as compared to gold standard measurements of disease activity. Unfortunately, the lack of a single gold standard for measuring Crohn's disease activity is a limitation of these assessments. Studies of predictive criterion validity which compare whether the score predicts true Crohn's disease activity as measured by objective measures of inflammation including C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), fecal calprotectin and fecal lactoferrin or sequelae in the future, such as surgery or disability were reported. Statistical parameters reported for agreement between the histologic index and disease gold standards were assessed (i.e. sensitivity, specificity, receiver operating characteristic (ROC) curve, area under the curve, mean difference, weighted kappa, Spearman's r squared, and the ICC).

Construct validation acknowledges the lack of a gold standard measurement for Crohn's disease activity, and instead assesses whether the index is consistent with other available markers of disease activity. Studies of the construct validity of a histologic index were reported if the correlation between the histologic index score and endoscopic disease activity indices; or the correlation between histologic score and clinical disease activity indices (e.g. Crohn's Disease Activity Index (CDAI)); or the correlation between histologic score and quality of life measures were reported. The minimal clinically important difference and appropriate cut-off values to determine active and inactive disease states, clinical response and clinical remission were examined.

Responsiveness: Changes in the histologic disease activity scores following a treatment of known efficacy served as an assessment of responsiveness, and the ability of an index to detect change. Responsiveness was quantified using indicators of effect size or its functions (Zou 2005), or the use of ROC curves to describe how well various score changes can be used to distinguish between improved and unimproved patients (Deyo 1991).

Feasibility: Feasibility was assessed as rater evaluation of the ease of index administration and the time required for scoring.

The criteria proposed by Landis and Koch was used to interpret correlation estimates. A correlation estimate of less than 0.2 was considered 'slight', 0.21 to 0.40 was considered 'fair', 0.41 to 0.60 was considered 'moderate', 0.61 to 0.80 was considered 'substantial' and 0.81 to 1.00 was considered 'almost perfect' (Landis 1977).

Search methods for identification of studies

Electronic searches

A computer aided search to identify applicable studies was conducted using the following databases: MEDLINE (Ovid), EMBASE (Ovid), PubMed, and the Cochrane Library (CENTRAL) from inception to 20 July 2016. No language or document type restrictions were applied. The search strategies are listed in Appendix 1.

Searching other resources

We performed a manual review of bibliographies and abstracts submitted to major gastroenterology meetings (2000 to July 2016) including:

1. Digestive Disease Week;

2. United European Gastroenterology Week; and

3. European Crohn's and Colitis Organisation.

Reference lists from retrieved articles were scanned to identify additional citations that may have been overlooked by the database search.

Data collection and analysis

Selection of studies

Two authors (GN, CEP) independently reviewed the titles and abstracts of the studies identified from the literature search. The full text of potentially relevant citations were reviewed for inclusion and study investigators were contacted as needed for clarification. Case reports, editorials, commentary, letters to the editor and meeting reports were excluded. Potentially relevant publications were translated into English if necessary. Any disagreements regarding study eligibility were resolved by discussion and consensus with a third author (JKM).

A standardized form was used to assess eligibility of trials for inclusion in the study. Each item on the form was scored as yes, no or unclear. The following items were assessed:

a) Diagnosis of Crohn's disease; and

b) Use of a histologic disease activity index; and

c) Evaluation of validity, responsiveness, reliability, and feasibility of histologic disease activity index.

Data extraction and management

A standardized form was used to extract information from selected studies. Two authors (GN, CEP) independently extracted and recorded data. The following data was recorded from each eligible study:

a) Number of patients enrolled, number of patients per treatment arm;

b) Patient characteristics: age and gender distribution;

c) Description of histologic disease activity index utilized; and

d) Outcomes: validity, content validity, construct validity, criterion validity, responsiveness, intra-rater reliability, inter-rater reliability, and feasibility.

Assessment of risk of bias in included studies

We used the following criteria to appraise the risk of bias of included studies:

- Blinding to clinical and endoscopic information; and

- Independent observation by histopathologists.

Blinding to clinical information such as symptoms, physical examination or laboratory information, and endoscopic information, is necessary to objectively assess histologic data. Furthermore, independent observation is essential to ensure that we are confident in the inter-rater reliability coefficients.

We also assessed the methodology quality of the included studies using the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist. The checklist consists of 10 properties: (A) internal consistency, (B) reliability, (C) measurement error, (D) content validity, (E) structural validity (factor analysis), (F) hypothesis testing, (G) cross-cultural validity, (H) criterion validity, (I) responsiveness to change and (J) interpretability. Each property is rated on a four-point scale (1 = poor, 2 = fair, 3 = good, or 4 = excellent). The overall score for the assessment of an individual measurement property is obtained by taking the lowest score for any of the items in the box (i.e. if any item in the box is scored as 'poor' then the overall score for that property is 'poor'). Generalisability was also be assessed as part of the COSMIN checklist.

Measures of the effect of the methods

Descriptive statistics were used to report the validation of outcome data. Frequencies and percentages were shown for categorical variables.

Dealing with missing data

Where possible, authors were contacted to provide any missing information.

Sensitivity analysis

Given that this was a descriptive review, we did not conduct sensitivity analyses that excluded: (1) unpublished studies; and (2) studies of low methodological quality.

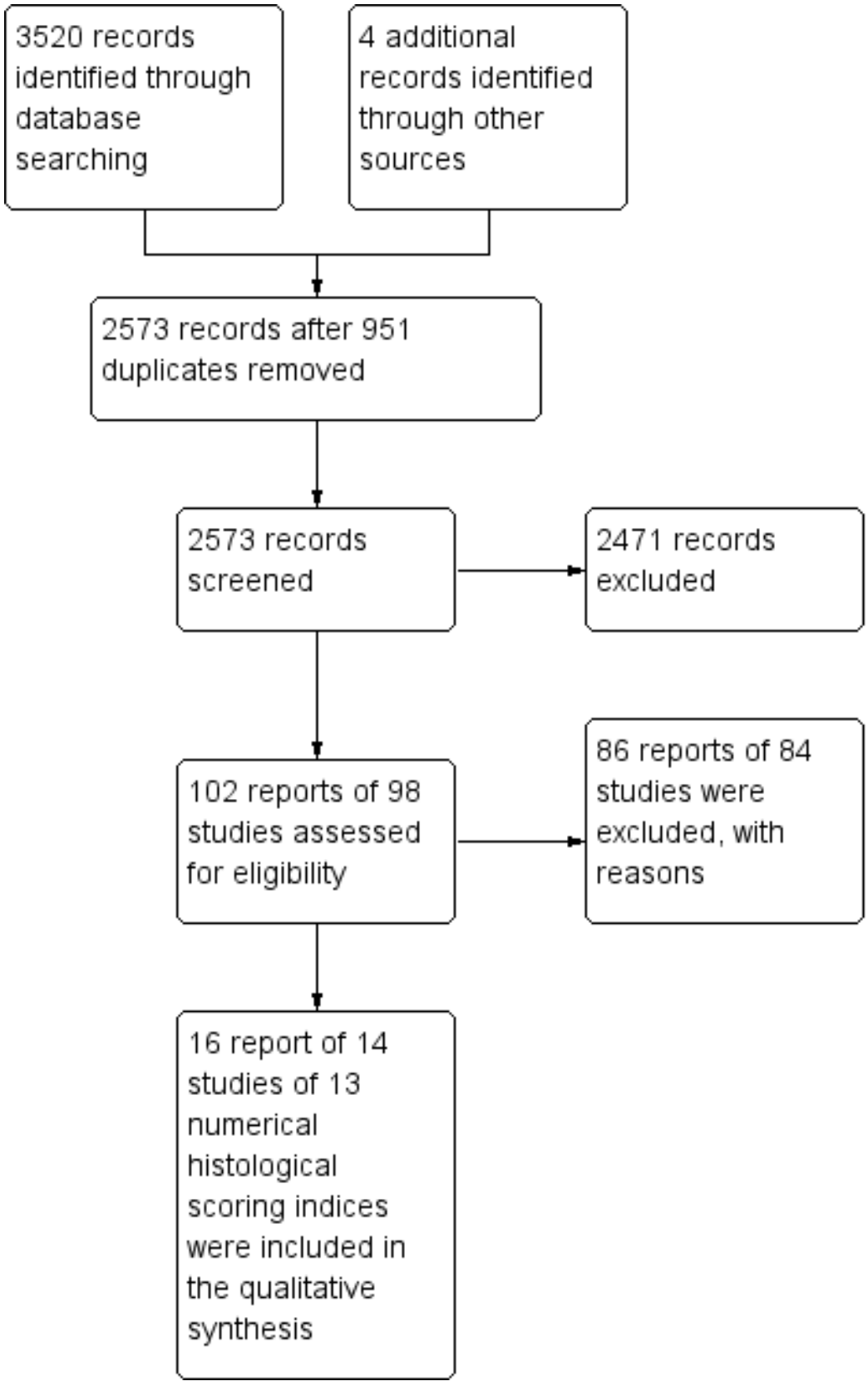
RESULTS

Description of studies

Results of the search

The literature search conducted on 20 July 2016 identified 3520 records. Four additional records were identified through other sources. After 951 duplicates were removed, a total of 2573 records were screened for inclusion. Of these, 102 reports were selected for full text review. Eighty-six reports of 84 studies were excluded (see Characteristics of excluded studies), leaving 16 reports of 14 studies evaluating numerical histological scores that met the pre-defined inclusion criteria (Figure 1). During the review process, we identified multiple overlapping stepwise histologic disease activity indices with the same or similar items. The index development process involves assessment of individual items. In our review, all of items of the stepwise histological disease activity indices from excluded studies were included among the items of numerical HDAI of the included studies. We decided to exclude all stepwise indices from further analysis.

Figure 1. Study flow diagram.



Included studies

Sixteen reports of 14 studies reported data on the operating characteristics of numerical histological indices (Agnholt 2003; Baert 1999; D'Haens 1999; Drews 2009; Geboes 2005; Gomes 1986; Laharie 2011; Mantzaris 2009; Naini 2012; Regueiro 2009; Sipponen 2008; Smith 2010a; Ward 1977; Yamamoto 2005). In total, 13 different numerical histologic indices were identified. These histologic indices included the Agnholt Score, the Global Histological Activity Score (GHAS), Drews Score, Colonic and Ileal Global Histological Activity Score, Gomes Score, Laherie Score, Average Histologic Score, Naini and Cortina Score, Regueiro Score, Sipponen Score, Dieleman Score, Ward Score and Saverymuttu Score. Six of these scores (the Agnholt Score, Colonic and Ileal Global Histological Activity Score, Laherie Score, Average Histologic Score, Regueiro Score and Sipponen Score) are modifications of the GHAS index (Table 1). The scoring index cut-offs of the histologic scoring indices that have undergone validation testing are listed in Table 2.

Setting

Eleven studies were prospective (Agnholt 2003; Baert 1999; D'Haens 1999; Geboes 2005; Gomes 1986; Laharie 2011; Mantzaris 2009; Regueiro 2009; Sipponen 2008; Smith 2010a; Yamamoto 2005), two studies were retrospective (Drews 2009; Ward 1977) and one study was composed of two retrospective phases (calibration and validation), followed by a prospective application phase (Naini 2012).

Histologic assessment was performed by a single pathologist in ten studies (Baert 1999; D'Haens 1999; Geboes 2005; Gomes 1986; Laharie 2011; Mantzaris 2009; Regueiro 2009; Sipponen 2008; Smith 2010a; Ward 1977), and by two pathologists in two studies (Agnholt 2003; Naini 2012). Drews 2009 and Yamamoto 2005 did not report how the histologic assessments were performed.

Description of included histologic scoring indices

One study specifically aimed to develop and assess the operating characteristics of a histologic scoring index. Naini 2012 described the development of the Naini and Cortina Score in a multi-phase study. The score was initially intended to standardize and reduce interpretive variability in histopathological assessment of mucosal biopsies for chronic ileocolitis and inflammatory bowel disease (IBD) diagnosis. The Naini and Cortina Score consists of 15 items which allows for separate scoring of the colon (ten items) and ileum (five items). For each item, scores range from zero to two. The score was developed in three phases. In phase one (calibration), the authors developed a histologic scoring index for chronic ileocolitis, measured agreement between two pathologists, and determined the ability of the score to distinguish ileocolitis from negative biopsies using retrospective specimens from ten patients. In phase two (validation phase), the investigators retrospectively included 164 patients who had undergone colonoscopy with biopsy procurement for clinical suspicion of ileocolitis to reaffirm agreement between pathologists, confirm the hypothesis that IBD cases would score high on the index, and to determine the likelihood of IBD. In phase three (application phase), the pathologists prospectively confirmed the accuracy and ease of application of the score in 30 patients.

Other included studies did not specifically aim to assess the operating properties of a histologic scoring index, but included some degree of validation testing. Furthermore, the construction and design of histological indices is not well reported. This could be due to word count limitations in articles.

In 1977, the first numerical histologic scoring index for Crohn's disease was described, which assessed the correlation between histology of rectal biopsies and prognosis (Ward 1977). The Ward Score consists of ten histologic features divided into two groups: first, non-graded features which can be rated as either present or absent (e.g. ulceration, fissures), and, second, graded features which can be scored on a semi-quantitative scale (e.g. infiltration of neutrophils). The total score is calculated by adding the scores of individual features with a maximum value of ten. The Ward Score is based exclusively on rectal biopsies which means that it is not applicable in patients with involvement of the colon, small intestine or upper-gastrointestinal tract. This score has not been used widely in clinical trials.

The most commonly used histologic disease activity index in clinical studies is the GHAS. In the initial index development study, ileal fecal fluids were infused into the neoterminal ileum of patients with a temporary diverting ileostomy. Biopsies from the ileum and colon were evaluated to determine whether the fecal stream causes recurrence of inflammation after surgery in the ileum (D'Haens 1998). The GHAS consists of eight items assessing acute and chronic inflammatory changes, epithelial damage and the extent of inflammation (i.e. the proportion of biopsy specimens affected). Each of the eight items is scored, with the totals subsequently added together (D'Haens 1999). The GHAS has been used in multiple studies, but only Baert 1999 and D'Haens 1999 have evaluated the operating properties of this index.

Many modifications of the GHAS have been proposed, and seven of these modifications have undergone validation testing. Geboes assessed the GHAS separately in the colon (CGHAS) and ileum (IGHAS) (Geboes 2005). Similarly, Sipponen 2008 counted ileal and colonic scores separately but colonic histologic activity was the sum of the total scores for four segments (right, transverse and left colon, and rectum). Thus, the maximum score for the ileum and colon was 16 and 64, respectively. For the Average Histology Score (AHS), the histologic disease activity of each explored colonic segment (rectum and sigmoid colon, descending colon, transverse colon, ascending colon, and cecum) and terminal ileum is scored with the GHAS. The final AHS is achieved by dividing the sum of individual intestinal segmental scores by the number of intestinal segments explored (Mantzaris 2009). Agnholt developed an index that includes only inflammatory changes from the GHAS (i.e. epithelial damage, mononuclear cells in lamina propria, polymorphonuclear cells in lamina propria and neutrophils in the epithelium). The final score is an average score of the individual biopsies (Agnholt 2003). The modification of the GHAS by Laharie 2011 omits the last item from the original GHAS (number of biopsy specimens affected), with total scores ranging from one to thirteen. Similarly, Regueiro replaced the 'number of biopsy specimens affected' item with 'pyloric gland metaplasia'. The maximum score of this histologic index, which focuses on the biopsies from the neoterminal ileum, is 14 (Regueiro 2009).

The second most commonly used histologic activity index in Crohn's disease was first described by Saverymuttu 1986. This numerical score was initially used to compare histologic disease

activity to indium 111-granulocyte scanning. Each biopsy specimen was assessed for the severity of changes in enterocytes and crypts, and for lamina propria involvement by mononuclear cells and neutrophils. These four items are scored on a scale from zero to three, depending on the severity of abnormalities, and added together. The average score from individual biopsy specimens from each region is converted to a grade (grade zero to three) with higher grades indicating greater disease activity (Saverymuttu 1986). This scoring system has been applied in multiple studies, and one study has reported on the operating characteristics of the Saverymuttu Score (Yamamoto 2005).

First described in a mouse study, the Dieleman Score assesses histologic disease activity based on the amount of inflammation (zero to three), depth of inflammation (zero to three), amount of crypt damage (zero to four) and crypt regeneration (zero to four). These changes are also quantified as a percentage of involvement on a scale from zero to four (Dieleman 1998). This histologic scoring index was subsequently applied and partially validated in a trial evaluating the opioid antagonist naltrexone in active Crohn's disease (Smith 2010a).

One of the included studies validated a histologic score that is both numerical and stepwise. The Gomes Score is achieved by adding stepwise scores that are graded on a scale of zero (normal) to four (most severe inflammation with active ulcerations) from individual biopsies taken from six colonic segments for a maximum total of 24 (Gomes 1986). Reminiscent of the Ward Score, this index is based exclusively on colonic biopsies thus limiting its use in Crohn's disease patients.

The Drews Score is based exclusively on ileal biopsies. It was developed in a study comparing histology to sonographically measured bowel wall vascularity and clinical disease activity (Drews 2009). Pathomorphological features of the score include infiltration by lymphocytes and plasma cells in the chronic inactive disease stage, infiltration of granulocytes in the active phase, crypt deformations, erosions and abscesses, and epithelioid cell granulomas. Histologic severity is ranked on a scale from zero to four.

Excluded studies

Eighty-six reports of 84 studies were excluded with reasons (see Characteristics of excluded studies).

Among the excluded reports we identified 53 different histologic scoring indices. Fifty studies (52 reports) described 41 stepwise instruments (Ajaj 2005; Allgayer 2007; Allison 1988; Baars 2010; Baars 2012; Beattie 1994; Bergeron 2010; Binder 1970; Bojic 2011; Breese 1995; Chiorean 2007; Chung-Faye 2011; Chung-Faye 2014; Colombel 2006; Cosin 2011; Di Sabatino 2009; Dieckgraefe 2002; Ellrichmann 2012; Ewe 1999; Fazio 1996; Ferrante 2006; Griga 1999; Haber 2002; Hanski 1999; Kaiser 2007; Kolkman 1996; Koutroubakis 2003; Kozarek 1989; Lasocki 2011; Leddin 1987; Lenze 2012; Levine 1993; Licata 2012; Maconi 2003; Matts 1961; Mazzucchelli 1994; Merra 2012; Neumann 2013; Nielsen 2003; Scheurlen 1998; Schunk 2001; Sciarretta 1993; Shoenut 1994; Shyn 2010; Smedh 1995; Smedh 1996; Sobhani 1992; Solem 2005; Vieira 2009; Wardle 1992). Twenty studies described 12 numerical instruments that had not undergone any validation testing (Ang 2000; Bariol 2002; Berni 2008; Bernstein 1993; Bruwer 2001; D'Haens 1998; D'Haens 2001; Folwaczny 1999; Girlich 2011; Hommes 1996; Iacucci 2015a; Iacucci 2015b; Mahmud 1996; Molnar 2009; Pullman 1988; Ripolles 2013; Saverymuttu 1986; Silva 2003; Silva 2004; Steward 2012). Many histological disease activity indices were incorporated into composite scores together with endoscopic, surgical and radiologic findings (e.g. Kolkman 1996; Koutroubakis 2003; Lasocki 2011; Leddin 1987; Lenze 2012).

Among studies using a numerical histological index, the main reasons for exclusion were absence of data on the development or assessment of operating properties of the histological scoring systems (Bernstein 1993; Bruwer 2001; D'Haens 1998; D'Haens 2001; Folwaczny 1999; Girlich 2011; Hommes 1996; Iacucci 2015b; Molnar 2009; Pullman 1988; Ripolles 2013; Silva 2003; Silva 2004; Steward 2012); inclusion of paediatric patients (Berni 2008; Silva 2003; Silva 2004); and inclusion of ulcerative colitis patients without separately reported data for patients with ulcerative colitis and Crohn's disease (Ang 2000; Bariol 2002; Iacucci 2015a; Mahmud 1996; Saverymuttu 1986).

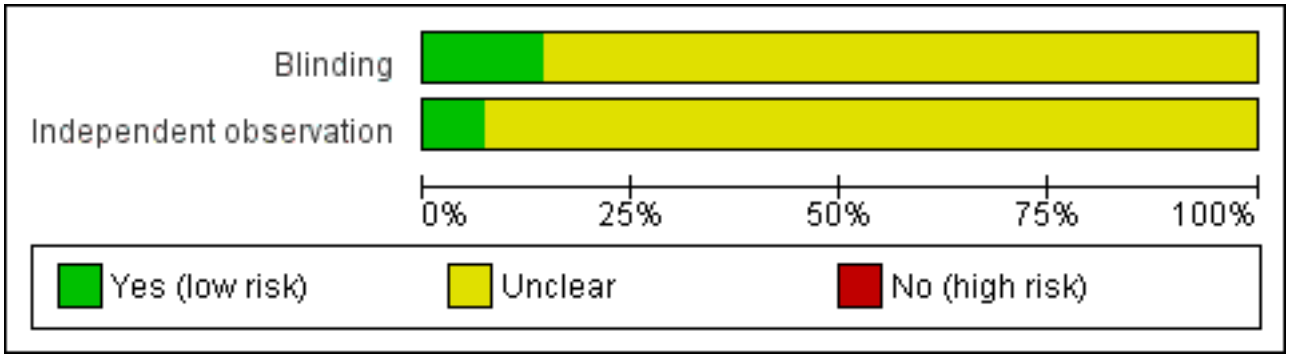
Among the studies that evaluated a non-numerical index, the main reasons for exclusion were usage of a stepwise histological scoring index; the absence of a specific description of the histological scoring index (Borody 2002; D'Haens 1997; Labaere 2013; Lofberg 2002; Macoritto 2012; Migaleddu 2009; Neaga 2009; Neumann 2012; Nicholls 1994; Plesec 2009; Schmitz-Moormann 1988; Seldenrijk 1991); and study design (i.e. animal study) (Abad 2014).

Risk of bias in included studies

Blinding

The presence or absence of blinding to clinical and endoscopic information was not routinely reported. In two studies, pathologists were fully blinded to clinical and endoscopic information (Geboes 2005; Gomes 1986). In three studies, pathologists were blinded to clinical information (Agnholt 2003; Naini 2012; Ward 1977), but it was not reported if they were blinded to endoscopic information as well. Four studies reported that pathologists were blinded but did not report more precisely if blinding was performed with respect to clinical and endoscopic information (D'Haens 1999; Mantzaris 2009; Regueiro 2009; Yamamoto 2005). In three studies, pathologists were blinded only to the treatment patients received in the clinical trial (Baert 1999; Laharie 2011; Smith 2010a). Finally, two studies did not report if the pathologists were blinded to clinical and endoscopic information (Drews 2009, Sipponen 2008) (Figure 2).

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



Independent observation

In the majority of studies the histological assessment of biopsy specimens was performed by a single pathologist (Baert 1999; D'Haens 1999; Geboes 2005; Gomes 1986; Laharie 2011; Mantzaris 2009; Regueiro 2009; Sipponen 2008; Smith 2010a; Ward 1977). As a result, independent observation by pathologists was not relevant, but differences in training and expertise could naturally account for variation in results. In phase one of Naini 2012, pathologists performed scoring independently, however it was unclear whether independent observation was performed in phase two. In phase three, scoring was performed by a single pathologist and independent observation was not relevant. Agnholt 2003 did not report on independent observation. There is a potential concern that scorings between the pathologists were not consistent because inter-rater reliability was not reported. Drews 2009 and Yamamoto 2005 did not report on the number of pathologists who scored the biopsies or whether the histological assessments were made independently. It is possible that a single pathologist examined the biopsy specimens, or that independent observation of histopathology was not considered in the clinical trial (Figure 2).

Methodological quality

The COSMIN tool was used to assess methodological quality of the included studies (see Table 3; Table 4).

In total, seven studies assessed criterion validity for any histological scoring index. With regard to methodological quality, one of these studies was rated as 'good' (Sipponen 2008), two were rated as 'fair' (Geboes 2005 (for the CGHAS); Laharie 2011) and five were rated as 'poor' (Geboes 2005 (for the IGHAS); Gomes 1986; Naini 2012; Regueiro 2009; Ward 1977).

Four studies assessed construct validity. With regard to methodological quality, two of these studies were rated as 'fair' (Geboes 2005 (for the CGHAS); Drews 2009) and three were rated as 'poor' (Agnholt 2003; Geboes 2005 (for the IGHAS); Gomes 1986).

One study assessed reliability and with regard to methodological quality, it was rated as 'good' (Naini 2012).

Seven studies assessed responsiveness for any histological scoring index. With regard to methodological quality, one of these studies was rated as 'fair' (Mantzaris 2009) and six were rated as 'poor' (Agnholt 2003; Baert 1999; D'Haens 1999; Geboes 2005; Smith 2010a; Yamamoto 2005).

One of the included studies assessed feasibility (Naini 2012).

Effect of methods

Reliability

One study assessed inter-rater reliability (Naini 2012). In phase one and phase two of the Naini and Cortina Score development paper, agreement between two pathologists was measured. In phase one, which included ten patients, there was 'almost perfect' agreement for both ileitis and colitis scores generated by the pathologists, as demonstrated by correlation coefficients of r = 0.96 and 0.95, respectively. Phase two, which included 164 patients and a minor modification of the score, reaffirmed 'almost perfect' agreement between pathologists. The correlation coefficients were r = 0.94 for scoring of both ileitis and colitis (Table 5).

None of the included studies assessed intra-rater reliability, test-retest reliability, or internal consistency.

Validity

Content validity

None of the included studies assessed the content validity of histological scoring indices for Crohn's disease.

Criterion validity

Correlation estimates between the histological scoring indices and objective biomarkers (i.e. CRP, ESR, fecal calprotectin and fecal lactoferrin) were 'fair' to 'moderate' (ranging from r = 0.25 to 0.56) and reported in two studies (Gomes 1986; Sipponen 2008; See Table 6).

CRP

One study investigated the correlation between a histological scoring system (the Gomes Score) and CRP (Gomes 1986). The estimate of correlation was 'fair', with r = 0.34 (P value not stated).

ESR

One study explored the relationship between the Gomes Score and ESR (Gomes 1986). The estimate of correlation was r = 0.25 (P value not reported), indicating 'fair' agreement.

Fecal calprotectin

For the Sipponen Score, the estimate of correlation was 'fair' (r = 0.31) for the modification of the ileal GHAS and 'moderate' (r = 0.56) for the modification of the colonic GHAS (Sipponen 2008). However, it should be noted that the P value of the former was not statistically significant, while P < 0.01 for the latter.

Fecal lactoferrin

The estimate of correlation between stool lactoferrin and the ileal Sipponen Score was 'fair', with r = 0.29, however this agreement was not statistically significant. Agreement between stool lactoferrin and the colonic component of the Sipponen Score was 'moderate', with r = 0.54 (P < 0.001) (Sipponen 2008).

Construct validity

A total of nine Crohn's disease scoring indices have been tested for construct validity (Agnholt 2003; Drews 2009; Geboes 2005; Gomes 1986; Laharie 2011; Naini 2012; Regueiro 2009; Sipponen 2008; Ward 1977). The correlation estimates between the histologic scoring indices and other measures of disease activity (i.e. clinical and endoscopic measurement tools) ranged from 'slight' to 'substantial'. The Agnholt Score, Drews Score and Geboes score were compared to the CDAI, the most commonly used clinical measure of disease activity. The estimate of correlation between the Aghholt Score and the CDAI was found to be r = 0.5 (P = 0.019) at week zero and r = 0.6 (P = 0.002) at week eight (Agnholt 2003). For the Drews Score, a non-statistically significant correlation of r = 0.25 was observed (Drews 2009). In Geboes 2005, the correlation estimates for the colonic GHAS and ileal GHAS were r = 0.43 and r = 0.10, respectively. While the correlation estimate for the colonic GHAS (P = 0.001) was statistically significant, this was not the case for the ileal GHAS (P = 0.700). The Gomes Score was compared to the Harvey Bradshaw index, however, no statistically significant agreement was observed (Gomes 1986). The colonic and ileal GHAS scores proposed by Geboes 2005 were also compared to the Inflammatory Bowel Disease Questionnaire (IBDQ). At weeks 10 and 54, the correlation estimates between the colonic GHAS and IBDQ were -0.33 (P = 0.037) and -0.04 (P = 0.873), respectively. For the ileal GHAS, the estimates of correlation with the IBDQ at weeks 10 and 54 were -0.29 (P = 0.132) and -0.35 (P = 0.246), respectively (Geboes 2005).

Four measures of endoscopic activity were used as comparisons: the Crohn's Disease Endoscopic Index of Severity (CDEIS), Gomes Score (endoscopic), Simple Endoscopic Score for Crohn's Disease (SES-CD), and Rugteerts Score. For the colonic GHAS described by Geboes 2005, the estimate of correlation with the CDEIS was r = 0.56. Although there was a statistically significant correlation between change in CDEIS and change in CGHAS from week 0 to week 10 was (r = 0.52, P = 0.001), the correlation coefficient was not statistically significant from week 0 to week 54. For the ileal GHAS, the estimate of correlation with the CDEIS was not statistically significant (Geboes 2005). With respect to the modified GHAS proposed by Laharie 2011, no significant correlation between the new index and the CDEIS or SES-CD was observed. For the ileal GHAS as modified by Sipponen 2008, the estimates of correlation with ileal SES-CD were r = 0.779, and r = 0.759 for the modified colonic GHAS. The correlation estimate between the modified GHAS developed by Regueiro and the Rutgeert's endoscopic score was r = 0.73. In Gomes 1986, novel histologic and endoscopic scores were compared; the estimate of correlation between the new scores was r = 0.76.

In the study by Ward 1977, the authors assessed whether the histological score could predict prognosis and disease sequelae in the future. At baseline, the Ward Score was capable of differentiating between symptomatic and asymptomatic patients, and those who died as a direct result of Crohn's disease, however no other statistically significant differences among groups were observed (Ward 1977).

The Naini and Cortina Score was not developed to assess histological disease activity but to standardise histological assessment of chronic ileocolitis and diagnose IBD. In the validation phase, the authors aimed to confirm that IBD cases would consistently receive high scores. Retrospective scoring of mucosal biopsies of 164 patients who had undergone colonoscopy for clinical suspicion of ileocolitis was performed. Subsequently, the clinical diagnosis was revealed and compared to the histologic score. Cases of IBD consistently scored higher than cases of suspected ileocolitis. For the diagnosis of IBD, an ileal score of five or greater had an 88% positive predictive value (PPV), whereas a colitis score of nine or greater had a 92% PPV. Ileal scores of three and four had a 53% PPV, and for scores of two or less, the PPV was 19%. Likewise, colonic scores of four and eight had a PPV of 52%, and for scores three or less, the PPV was 8% (Naini 2012) (Table 7).

Responsiveness

Responsiveness data regarding seven histological scoring indices is described in Table 8. After subjects were administered a treatment of known efficacy, there was a statistically significant change score observed with the GHAS (D'Haens 1999), the CGHAS and IGHAS (Geboes 2005), the Agnholt Score (Agnholt 2003), the AHS (Mantzaris 2009) and the Dieleman Score (Smith 2010a).

Although Baert 1999 and Yamamoto 2005 observed decreases in the GHAS and Saverymuttu Score, respectively, following a treatment of known efficacy, testing for statistical significance was not performed in these studies.

Feasibility

One of the included studies evaluated feasibility (Table 9). In the application phase of the development of the Naini and Coritna Score, pathologists rated the novel index as being simple to use and easily completed in less than 30 seconds (Naini 2012).

DISCUSSION

Summary of main results

Sixteen reports of 14 studies describing 13 numerical histological indices fulfilled the inclusion criteria. Inter-rater reliability was assessed in one study (Naini 2012). For scoring ileitis and colitis with the Naini and Cortina Score, ICCs ranged from r = 0.94 to r = 0.96. With regard to methodological quality, the study was rated as 'good'.

With respect to validity, correlation estimates between various histological scoring systems and Crohn's disease activity as measured by objective markers of inflammation (i.e. CRP, ESR, fecal calprotectin and lactoferrin), endoscopic disease activity scores, clinical disease activity scores and quality of life measures were

reported. For criterion validity, estimates of correlation ranged from r = 0.25 to r = 0.56 (Table 6).The estimates of correlation of histological scoring indices with endoscopic disease activity scores ranged from no correlation (Laharie 2011) to 'substantial' (r = 0.779) (Sipponen 2008). With respect to construct validity, the estimates of correlation between the histologic scoring indices and the clinical disease activity or quality of life measures ranged from no correlation to 0.78 (Table 7). It should be noted that no more than one study assessed each scoring index for validity.

Responsiveness data were available in seven studies. After subjects were administered a treatment of known efficacy, statistically significant changes in index scores were demonstrated in five studies (D'Haens 1999; Geboes 2005; Agnholt 2003; Mantzaris 2009; Smith 2010a). In two studies (Baert 1999; Yamamoto 2005), the histologic score decreased after a treatment of known efficacy was administered, but testing for statistical significance was not performed.

Feasibility was assessed by one study (Naini 2012). The Naini and Cortina Score was shown to be simple to use and feasible for every given case.

Overall completeness and applicability of evidence

One study specifically aimed to develop and validate a histological scoring index (Naini 2012). However, the Naini and Corina Score is not specific to Crohn's disease and was not developed to assess histological disease activity, but rather to diagnose IBD.

None of the histological disease activity indices identified in the current review including the extensively used GHAS have been formally validated, and many of the validation studies were of poor methodological quality.

Quality of the evidence

Four operating properties were assessed by the included studies: reliability, hypothesis testing (a facet of construct validity), criterion validity and responsiveness (Table 4). The single study that examined index reliability was rated as 'poor' with regard to methodological quality. For hypothesis testing, there were five evaluations conducted. Three of these studies were rated as 'poor' and two were rated as 'fair'. Eight studies reported on criterion validity, with five of these studies being rated as 'poor', two rated as 'fair' and one rated as 'good'. Finally, seven studies measured responsiveness. Six of these studies were rated as 'poor' methodological quality and one was rated as 'fair'. The Nainia and Cortina Score was assessed for feasibility, and found to be easily administered. Details on the interpretability and generalisability of the included studies are listed in Table 4.

Potential biases in the review process

It should be noted that stepwise histological scoring indices were excluded from this review. Most stepwise indices include the same or similar items and are thus subject to considerable overlap. While such scoring indices are easy to use, it is likely that they are less responsive to clinically meaningful changes in disease activity (Mojtahed 2014). However, it should be acknowledged that these indices exist and that these indices are not represented in the current review.

It is also worth noting that numerous issues regarding biopsy collection and processing remain. Biopsy specimens in the included studies were procured according to various protocols. More specifically, different numbers of biopsies were taken at different locations. In most studies, biopsy specimens were taken from the most macroscopically inflamed areas. This strategy seems sensible, but there is no evidence to support that this is the best method for procuring biopsy specimens. Efforts should be made to standardise and optimise biopsy sampling protocols while taking the anatomic heterogeneity of Crohn's disease into account to reduce sampling error. At present there is no uniform methodology on how to handle biopsy specimens (e.g. type of forceps, orientation of the sample before fixation, duration of fixation, or sectioning) (Langner 2014). Good sample quality is necessary for assessment of the sections (Geboes 2000; Mosli 2017), but it is not clear whether quality checks and exclusion of poor quality specimens was performed. Furthermore, specific training of pathologists on how to use scoring systems might be needed (Geboes 2000; Mosli 2017). Most of these issues were not assessed in the studies included in this systemic review.

Agreements and disagreements with other studies or reviews

Our review is in agreement with another systemic review on histological disease activity indices in Crohn's disease (Mojtahed 2014).

AUTHORS' CONCLUSIONS

Implication for methodological research

In summary, there are no currently available, fully-validated histologic scoring indices for use in Crohn's disease patients. There is a great need for future studies to fully validate and assess operating properties of existing histological scoring indices, or to create indices according to the currently accepted standards for index development.

ACKNOWLEDGEMENTS

Partial funding for the Cochrane IBD Review Group (April 1, 2016 - March 31, 2018) has been provided by Crohn's and Colitis Canada (CCC).

REFERENCES

References to studies included in this review

Agnholt 2003 *{published data only}*

Agnholt J, Dahlerup JF, Buntzen S, Tottrup A, Nielsen SL, Lundorf E. Response, relapse and mucosal immune regulation after infliximab treatment in fistulating Crohn's disease. *Alimentary Pharmacology and Therapeutics* 2003;**17**(5):703-10.

Baert 1999 *{published data only}*

Baert F J, D'Haens GR, Peeters M, Hiele MI, Schaible TF, Shealy D, et al. Tumor necrosis factor alpha antibody (infliximab) therapy profoundly down-regulates the inflammation in Crohn's ileocolitis. *Gastroenterology* 1999;**116**(1):22-8.

D'Haens 1999 *{published data only}*

D'Haens G, Deventer S, Hogezand R, Chalmers D, Kothe C, Baert F, et al. Endoscopic and histological healing with infliximab anti-tumor necrosis factor antibodies in Crohn's disease: A European multicenter trial. *Gastroenterology* 1999;**116**(5):1029-34.

Drews 2009 *{published data only}*

Drews BH, Barth TF, Hanle MM, Akinli AS, Mason RA, Muche R, et al. Comparison of sonographically measured bowel wall vascularity, histology, and disease activity in Crohn's disease. *European Radiology* 2009;**19**(6):1379-86.

Geboes 2005 *{published data only}*

Geboes K, Rutgeerts P, Opdenakker G, Olson A, Patel K, Wagner CL, et al. Endoscopic and histologic evidence of persistent mucosal healing and correlation with clinical improvement following sustained infliximab treatment for Crohn's disease. *Current Medical Research and Opinion* 2005;**21**(11):1741-54.

Gomes 1986 *{published data only}*

Gomes P, du Boulay C, Smith C L, Holdstock G. Relationship between disease activity indices and colonoscopic findings in patients with colonic inflammatory bowel disease. *Gut* 1986;**27**(1):92-5.

Laharie 2011 *{published data only}*

Laharie D, Reffet A, Belleannee G, Chabrun E, Subtil C, Razaire S, et al. Mucosal healing with methotrexate in Crohn's disease: a prospective comparative study with azathioprine and infliximab. *Alimentary Pharmacology and Therapeutics* 2011;**33**(6):714-21.

Mantzaris 2009 *{published data only}*

Mantzaris G J, Christidou A, Sfakianakis M, Roussos A, Koilakou S, Petraki K, et al. Azathioprine is superior to budesonide in achieving and maintaining mucosal healing and histologic remission in steroid-dependent Crohn's disease. *Inflammatory Bowel Diseases* 2009;**15**(3):375-82.

Naini 2012 *{published data only}*

Naini BV, Cortina G. A histopathologic scoring system as a tool for standardized reporting of chronic (ileo)colitis and independent risk assessment for inflammatory bowel disease. *Human Pathology* 2012;**43**(12):2187-96.

Regueiro 2009 *{published data only}*

Regueiro M, Palma-Diaz M, El-Hachem S, Kip KE, Schraut WH, Harrison J, et al. Correlation of histologic and endoscopic scores for evaluation of Crohn's disease recurrence after ileal resection and infliximab therapy. *Gastroenterology* 2010;**138**(5):S-358.

* Regueiro M, Schraut W, Baidoo L, Kip KE, Sepulveda AR, Pesci M, et al. Infliximab prevents Crohn's disease recurrence after ileal resection. *Gastroenterology* 2009;**136**(2):441-50.

Sipponen 2008 *{published data only}*

Sipponen T, Karkkainen P, Savilahti E, Kolho KL, Nuutinen H, Turunen U, et al. Correlation of faecal calprotectin and lactoferrin with an endoscopic score for Crohn's disease and histological findings. *Alimentary Pharmacology and Therapeutics* 2008;**28**(10):1221-9.

Smith 2010a *{published data only}*

Smith JP, Bingaman SI, Ruggiero F, Mauger DT, Mukherjee A, McGovern CO, et al. Naltrexone therapy improves activity and promotes mucosal healing in active Crohn's disease: A placebo-controlled trial. *Gastroenterology* 2010;**138**(5):S-85, S-86.

* Smith JP, Bingaman SI, Ruggiero F, Mauger DT, Mukherjee A, McGovern CO, et al. Therapy with the opioid antagonist naltrexone promotes mucosal healing in active Crohn's disease: a randomized placebo-controlled trial. *Digestive Diseases and Sciences* 2011;**56**(7):2088-97.

Ward 1977 *{published data only}*

Ward M, Webb JN. Rectal biopsy as a prognostic guide in Crohn's colitis. *Journal of Clinical Pathology* 1977;**30**(2):126-31.

Yamamoto 2005 *{published data only}*

Yamamoto T, Nakahigashi M, Umegae S, Kitagawa T, Matsumoto K. Impact of elemental diet on mucosal inflammation in patients with active Crohn's disease: cytokine production and endoscopic and histological findings. *Inflammatory Bowel Diseases* 2005;**11**(6):580-8.

References to studies excluded from this review

Abad 2014 *{published data only}*

Abad C, Cheung-Lau G, Coute-Monvoisin AC, Waschek JA. Vasoactive intestinal peptide-deficient mice exhibit reduced pathology in trinitrobenzene sulfonic acid-induced colitis. *Neuroimmunomodulation* 2015;**22**(3):203-12.

Ajaj 2005 *{published data only}*

Ajaj WM, Lauenstein TC, Pelster G, Gerken G, Ruehm SG, Debatin JF, et al. Magnetic resonance colonography for the detection of inflammatory diseases of the large bowel: quantifying the inflammatory activity. *Gut* 2005;**54**(2):257-63.

Allgayer 2007 *{published data only}*

Allgayer H, Roisch U, Zehnter E, Ziegenhagen DJ, Dienes HP, Kruis W. Colonic ornithine decarboxylase in inflammatory bowel disease: ileorectal activity gradient, guanosine triphosphate stimulation, and association with epithelial regeneration but not the degree of inflammation and clinical features. *Digestive Diseases and Sciences* 2007;**52**(1):25-30.

Allison 1988 *{published data only}*

Allison MC, Cornwall S, Poulter LW, Dhillon AP, Pounder RE. Macrophage heterogeneity in normal colonic mucosa and in inflammatory bowel disease. *Gut* 1988;**29**:1531-8.

Ang 2000 *{published data only}*

Ang YS, Mahmud N, White B, Byrne M, Kelly A, Lawler M, et al. Randomized comparison of unfractionated heparin with corticosteroids in severe active inflammatory bowel disease. *Alimentary Pharmacology and Therapeutics* 2000;**14**(8):1015-22.

Baars 2010 *{published data only}*

Baars JE, Vogelaar L, Wolfhagen FH, Biermann K, Kuipers EJ, van der Woude CJ. A short course of corticosteroids prior to surveillance colonoscopy to decrease mucosal inflammation in inflammatory bowel disease patients: Results from a randomized controlled trial. *Journal of Crohn's and Colitis* 2010;**4**(6):661-8.

Baars 2012 *{published data only}*

Baars JE, Nuij VJ, Oldenburg B, Kuipers EJ, Van Der Woude CJ. The majority of patients with inflammatory bowel disease in clinical remission has mucosal inflammation. *Gastroenterology* 2011;**1**:S776.

* Baars JE, Nuij VJ, Oldenburg B, Kuipers EJ, van der Woude CJ. Majority of patients with inflammatory bowel disease in clinical remission have mucosal inflammation. *Inflammatory Bowel Diseases* 2012;**18**(9):1634-40.

Bariol 2002 *{published data only}*

Bariol C, Meagher AP, Vickers CR, Byrnes DJ, Edwards PD, Hing M, et al. Thalidomide for inflammatory bowel disease: Early studies on the safety and efficacy of thalidomide for symptomatic inflammatory bowel disease. *Journal of Gastroenterology and Hepatology* 2002;**17**(2):135-9.

Beattie 1994 *{published data only}*

Beattie RM, Schiffrin EJ, Donnet-Hughes A, Huggett AC, Domizio P, MacDonald TT, et al. Polymeric nutrition as the primary therapy in children with small bowel Crohn's disease. *Alimentary Pharmacology and Therapeutics* 1994;**8**(6):609-15.

Bergeron 2010 *{published data only}*

Bergeron V, Nion-Larmurier I, Vienne A, Seksik P, Sokol H, Ruskone-Fourmestraux A, et al. Azathioprine (AZA) is associated with less histological inflammation of the colon in clinically inactive IBD. *Gastroenterology* 2010;**1**:S693.

Berni 2008 *{published data only}*

Berni Canani R, Terrin G, Rapacciuolo L, Miele E, Siani M C, Puzone C, et al. Faecal calprotectin as reliable non-invasive marker to assess the severity of mucosal inflammation in children with inflammatory bowel disease. *Digestive and Liver Disease* 2008;**40**(7):547-53.

Bernstein 1993 *{published data only}*

Bernstein CN, Robert ME, Eysselein VE. Rectal substance P concentrations are increased in ulcerative colitis but not in Crohn's disease. *American Journal of Gastroenterology* 1993;**88**:908-13.

Binder 1970 *{published data only}*

Binder V. A comparison between clinical state, macroscopic and microscopic appearances of rectal mucosa, and cytologic picture of mucosal exudate in ulcerative colitis. *Scandinavian Journal of Gastroenterology* 1970;**5**:627-32.

Bojic 2011 *{published data only}*

Bojic D, Bojic B, Protic M, Svorcan P, Gligorijevic V, Necic D, et al. Fecal calprotectin is reliable surrogate marker of endoscopic and histologic mucosal healing in Crohn's disease and ulcerative colitis. *Journal of Crohn's and Colitis* 2011;**5**(1):S34.

Borody 2002 *{published data only}*

Borody TJ, Leis S, Warren EF, Surace R. Treatment of severe Crohn's disease using antimycobacterial triple therapy--approaching a cure?. *Digestive and Liver Disease* 2002;**34**(1):29-38.

Breese 1995 *{published data only}*

Breese EJ, Michie CA, Nicholls SW, Williams CB, Domizio P, Walker-Smith JA, et al. The effect of treatment on lymphokine-secreting cells in the intestinal mucosa of children with Crohn's disease. *Alimentary Pharmacology and Therapeutics* 1995;**9**:547-52.

Bruwer 2001 *{published data only}*

Bruwer M, Schmid KW, Metz KA, Krieglstein CF, Senninger N, Schurmann G. Increased expression of metallothionein in inflammatory bowel disease. *Inflammation Research* 2001;**50**(6):289-93.

Chiorean 2007 *{published data only}*

Chiorean MV, Sandrasegaran K, Saxena R, Maglinte DD, Nakeeb A, Johnson CS. Correlation of CT enteroclysis with surgical pathology in Crohn's disease. *American Journal of Gastroenterology* 2007;**102**(11):2541-50.

Chung-Faye 2011 *{published data only}*

Chung-Faye G, Sandhu K, Logan RP, Sherwood RA. Fecal calproctectin is strongly predictive of clinical disease activity and histological severity in inflammatory bowel disease. *Gastroenterology* 2011;**1**:S421.

Chung-Faye 2014 *{published data only}*

Chung-Faye G, Rahman A, Sandhu K, Hayee B, Tumova J, Sherwood R. Fecal calprotectin is an accurate predictor of endoscopic and histological disease activity in IBD. *Journal of Crohn's and Colitis* 2014;**8**:S153-4.

Colombel 2006 *{published data only}*

Colombel JF, Solem CA, Sandborn WJ, Booya F, Loftus EV Jr, Harmsen WS, et al. Quantitative measurement and visual

assessment of ileal Crohn's disease activity by computed tomography enterography: correlation with endoscopic severity and C reactive protein. *Gut* 2006;**55**:1561-7.

Cosin 2011 *{published data only}*

Cosin J, Ortiz-Masia D, Quintana E, Cabrero M, Calatayud S, Nicolau MJ, et al. Histological damage correlates with the number of HIF-2 positive cells in the lamina propria of patients with inflammatory bowel disease. *Basic and Clinical Pharmacology and Toxicology* 2011;**109**:21.

D'Haens 1997 *{published data only}*

D'Haens G, Geboes K, Ponette E, Penninckx F, Rutgeerts P. Healing of severe recurrent ileitis with azathioprine therapy in patients with Crohn's disease. *Gastroenterology* 1997;**112**(5):1475-81.

D'Haens 1998 *{published data only}*

D'Haens GR, Geboes K, Peeters M, Baert F, Penninckx F, Rutgeerts P. Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* 1998;**114**:262-7.

D'Haens 2001 *{published data only}*

D'Haens G, Swijsen C, Noman M, Lemmens L, Ceuppens J, Agbahiwe H, et al. Etanercept in the treatment of active refractory Crohn's disease: a single-center pilot trial. *American Journal of Gastroenterology* 2001;**96**(9):2564-8.

Dieckgraefe 2002 *{published data only}*

Dieckgraefe BK, Crimmins DL, Landt V, Houchen C, Anant S, Porche-Sorbet R, et al. Expression of the regenerating gene family in inflammatory bowel disease mucosa: Reg Ialpha upregulation, processing, and antiapoptotic activity. *Journal of Investigative Medicine* 2002;**50**(6):421-34.

Di Sabatino 2009 *{published data only}*

Di Sabatino A, Saarialho-Kere U, Buckley MG, Gordon JN, Biancheri P, Rovedatti L, et al. Stromelysin-1 and macrophage metalloelastase expression in the intestinal mucosa of Crohn's disease patients treated with infliximab. *European Journal of Gastroenterology and Hepatology* 2009;**21**(9):1049-55.

Ellrichmann 2012 *{published data only}*

Ellrichmann M, Bethge J, Nikolaus S, Arlt A, Kuehbacher T, Schreiber S, et al. Endoscopic ultrasound of the colon can differentiate between crohn's disease and ulcerative colitis and reliably quantifies the degree of inflammation - A prospective, blinded, and controlled comparative study. *Gastrointestinal Endoscopy* 2012;**1**:AB444.

* Ellrichmann M, Wietzke-Braun P, Dhar S, Nikolaus S, Arlt A, Bethge J, et al. Endoscopic ultrasound of the colon for the differentiation of Crohn's disease and ulcerative colitis in comparison with healthy controls. *Alimentary Pharmacology and Therapeutics* 2014;**39**(8):823-33.

Ewe 1999 *{published data only}*

Ewe K, Bottger T, Buhr HJ, Ecker KW, Otto HF. Low-dose budesonide treatment for prevention of postoperative recurrence of Crohn's disease: a multicentre randomized placebo-controlled trial. German Budesonide Study Group. *European Journal of Gastroenterology and Hepatology* 1999;**11**(3):277-82.

Fazio 1996 *{published data only}*

Fazio VW, Marchetti F, Church M, Goldblum JR, Lavery C, Hull TL, et al. Effect of resection margins on the recurrence of Crohn's disease in the small bowel. A randomized controlled trial. *Annals of Surgery* 1996;**224**(4):563-71.

Ferrante 2006 *{published data only}*

Ferrante M, de Hertogh G, Hlavaty T, D'Haens G, Penninckx F, D'Hoore A, et al. The value of myenteric plexitis to predict early postoperative Crohn's disease recurrence. *Gastroenterology* 2006;**130**(6):1595-606.

Folwaczny 1999 *{published data only}*

Folwaczny C, Wiebecke B, Loeschke K. Unfractioned heparin in the therapy of patients with highly active inflammatory bowel disease. *American Journal of Gastroenterology* 1999;**94**(6):1551-5.

Girlich 2011 *{published data only}*

Girlich C, Jung EM, Huber E, Ott C, Iesalnieks I, Schreyer A, et al. Comparison between preoperative quantitative assessment of bowel wall vascularization by contrast-enhanced ultrasound and operative macroscopic findings and results of histopathological scoring in Crohn's disease. *Ultraschall in der Medizin* 2011;**32**(2):154-9.

Griga 1999 *{published data only}*

Griga T, Voigt E, Gretzer B, Brasch F, May B. Increased production of vascular endothelial growth factor by intestinal mucosa of patients with inflammatory bowel disease. *Hepatogastroenterology* 1999;**46**(26):920-3.

Haber 2002 *{published data only}*

Haber HP, Busch A, Ziebach R, Dette S, Ruck P, Stern M. Ultrasonographic findings correspond to clinical, endoscopic, and histologic findings in inflammatory bowel disease and other enterocolitides. *Journal of Ultrasound in Medicine* 2002;**21**(4):375-82.

Hanski 1999 *{published data only}*

Hanski C, Born M, Foss HD, Marowski B, Mansmann U, Arasteh K, et al. Defective post-transcriptional processing of MUC2 mucin in ulcerative colitis and in Crohn's disease increases detectability of the MUC2 protein core. *Journal of Pathology* 1999;**188**(3):304-11.

Hommes 1996 *{published data only}*

Hommes DW, Meenan J, de Haas M, ten Kate FJ, von dem Borne AE, Tytgat GN, et al. Soluble Fc gamma receptor III (CD 16) and eicosanoid concentrations in gut lavage fluid from patients with inflammatory bowel disease: reflection of mucosal inflammation. *Gut* 1996;**38**(4):564-7.

Iacucci 2015a *{published data only}*

Iacucci M, Fort Gasia M, Hassan C, Panaccione R, Kaplan G, Ghosh S, Gui X. Complete mucosal healing defined by endoscopic Mayo sub score still demonstrates abnormalities

by novel high definition colonoscopy and refined histological gradings. *Endoscopy* 2015;**47**(8):726-34.

Iacucci 2015b *{published data only}*

Iacucci M, Fort Gasia M, Panaccione R, Kaplan G, Ghosh S, Gui X. Endoscopic assessment of mucosal healing in Crohn's disease by novel iSCAN endoscopic and refined histological gradings. *Gastroenterology* 2015;**148**(4):S-638.

Kaiser 2007 *{published data only}*

Kaiser T, Langhorst J, Wittkowski H, Becker K, Friedrich A W, Rueffer A, et al. Faecal S100A12 as a non-invasive marker distinguishing inflammatory bowel disease from irritable bowel syndrome. *Gut* 2007;**56**(12):1706-13.

Kolkman 1996 *{published data only}*

Kolkman JJ, Falke TH, Roos JC, Van Dijk DH, Bannink IM, Den Hollander W, et al. Computed tomography and granulocyte scintigraphy in active inflammatory bowel disease. Comparison with endoscopy and operative findings. *Digestive Diseases and Sciences* 1996;**41**(4):641-50.

Koutroubakis 2003 *{published data only}*

Koutroubakis IE, Koukouraki SI, Dimoulios PD, Velidaki AA, Karkavitsas NS, Kouroumalis EA. Active inflammatory bowel disease: evaluation with 99mTc (V) DMSA scintigraphy. *Radiology* 2003;**229**(1):70-4.

Kozarek 1989 *{published data only}*

Kozarek RA, Patterson DJ, Gelfand MD, Botoman VA, Ball TJ, Wilske KR. Methotrexate induces clinical and histologic remission in patients with refractory inflammatory bowel disease. *Annals of Internal Medicine* 1989;**110**:353-6.

Labaere 2013 *{published data only}*

Labaere D, Smismans A, Van Olmen G, Christiaens P, D'Haens G, Moons V, et al. Clinical comparison of eight different calprotectin immunoassays for the diagnosis and follow-up of inflammatory bowel disease. *Journal of Crohn's and Colitis* 2013;**7**:S120.

Lasocki 2011 *{published data only}*

Lasocki A, Pitman A, Williams R, Lui B, Kalade A V, Farish S. Relative efficacy of different MRI signs in diagnosing active Crohn's disease, compared against a histological gold standard. *Journal of Medical Imaging and Radiation Oncology* 2011;**55**(1):11-9.

Leddin 1987 *{published data only}*

Leddin DJ, Paterson WG, DaCosta LR, Dinda PK, Depew WT, Markotich J, et al. Indium-111-labeled autologous leukocyte imaging and fecal excretion. Comparison with conventional methods of assessment of inflammatory bowel disease. *Digestive Diseases and Sciences* 1987;**32**(4):377-87.

Lenze 2012 *{published data only}*

Lenze F, Wessling J, Bremer J, Ullerich H, Spieker T, Weckesser M, et al. Detection and differentiation of inflammatory versus fibromatous Crohn's disease strictures: prospective comparison of 18F-FDG-PET/CT, MR-enteroclysis, and transabdominal ultrasound versus endoscopic/histologic evaluation. *Inflammatory Bowel Diseases* 2012;**18**(12):2252-60.

Levine 1993 *{published data only}*

Levine DS. Immunoglobulin therapy in inflammatory bowel disease. *Canadian Journal of Gastroenterology* 1993;**7**(2):187-95.

Licata 2012 *{published data only}*

Licata A, Randazzo C, Cappello M, Calvaruso V, Butera G, Florena AM, et al. Fecal calprotectin in clinical practice: A noninvasive screening tool for patients with chronic diarrhea. *Journal of Clinical Gastroenterology* 2012;**46**(6):504-8.

Lofberg 2002 *{published data only}*

Lofberg R, Neurath M, Ost A, Pettersson S. Topical NFkB p65 antisense oligonucleotides in patients with active distal colonic IBD. A randomised, controlled pilot trial. *Gastroenterology* 2002;**122**(4):A60.

Maconi 2003 *{published data only}*

Maconi G, Carsana L, Fociani P, Sampietro GM, Ardizzone S, Cristaldi M, et al. Small bowel stenosis in Crohn's disease: clinical, biochemical and ultrasonographic evaluation of histological features. *Alimentary Pharmacology and Therapeutics* 2003;**18**(7):749-56.

Macoritto 2012 *{published data only}*

Macoritto M, Hoang V, Thomson T, Laifenfeld D, Drubin D. A novel mechanistic biomarker panel predicts response to infliximab in patients with Crohn's disease. *Inflammatory Bowel Diseases* 2012;**18**:S61.

Mahmud 1996 *{published data only}*

Mahmud N, McDonald GS, Kelleher D, Weir DG. Microalbuminuria correlates with intestinal histopathological grading in patients with inflammatory bowel disease. *Gut* 1996;**38**(1):99-103.

Matts 1961 *{published data only}*

Matts SGF. The value of rectal biopsy in the diagnosis of ulcerative colitis. *QJM* 1961;**30**(4):393-407.

Mazzucchelli 1994 *{published data only}*

Mazzucchelli L, Hauser C, Zgraggen K, Wagner H, Hess M, Laissue J A, et al. Expression of interleukin-8 gene in inflammatory bowel disease is related to the histological grade of active inflammation. *American Journal of Pathology* 1994;**144**:997-1007.

Merra 2012 *{published data only}*

Merra G, Gasbarrini G, Laterza L, Pizzoferrato M, Poscia A, Scaldaferri F, et al. Propionyl-L-carnitine hydrochloride for treatment of mild to moderate colonic inflammatory bowel diseases. *World Journal of Gastroenterology* 2012;**18**(36):5065-71.

Migaleddu 2009 *{published data only}*

Migaleddu V, Scanu AM, Quaia E, Rocca PC, Dore MP, Scanu D, et al. Contrast-enhanced ultrasonographic evaluation of inflammatory activity in Crohn's disease. *Gastroenterology* 2009;**137**(1):43-52.

Molnar 2009 *{published data only}*

Molnar T, Farkas K, Nagy F, Szepes Z, Tiszlavicz L, Nemeth I, et al. Does the endoscopic appearance of the ileocecal valve suggest the severity of Crohn's disease in the terminal ileum?. *Journal of Crohn's and Colitis* 2009;**3**(4):287-90.

Neaga 2009 *{published data only}*

Neaga M, Nuti F, Del Giudice E, Civitelli F, Viola F, Aloi M, et al. Non invasive stool markers in assessing bowel inflammation in children with inflammatory bowel disease. *Digestive and Liver Disease* 2009;**41S (S3)**:S210.

Neumann 2012 *{published data only}*

Neumann H, Vieth M, Atreya R, Grauer M, Gunther C, Neurath M. Virtual chromoendoscopy with i-scan for diagnosis of mucosal healing in patients with inflammatory bowel disease a prospective randomized double-blind controlled study. *Journal of Crohn's and Colitis* 2012;**6**:S54.

Neumann 2013 *{published data only}*

Neumann H, Vieth M, Neurath MF, Atreya R. Endocytoscopy allows accurate in vivo differentiation of mucosal inflammatory cells in IBD: a pilot study. *Inflammatory Bowel Diseases* 2013;**19**(2):356-62.

Nicholls 1994 *{published data only}*

Nicholls S, Domizio P, Williams CB, Dawnay A, Braegger CP, MacDonald TT, et al. Cyclosporin as initial treatment for Crohn's disease. *Archives of Disease in Childhood* 1994;**71**:243-7.

Nielsen 2003 *{published data only}*

Nielsen OH, Kirman I, Rudiger N, Hendel J, Vainer B. Upregulation of interleukin-12 and -17 in active inflammatory bowel disease. *Scandinavian Journal of Gastroenterology* 2003;**38**(2):180-5.

Plesec 2009 *{published data only}*

Plesec TP, Shen B, Remzi F, Fazio V, Bennett AE, Goldblum JR. Ischemic pouchitis (IP): A histologic evaluation and comparison to Crohn's disease of the pouch (CDP) and antibiotic-responsive pouchitis (ARP). *Laboratory Investigation* 2009;**89**:146A.

Pullman 1988 *{published data only}*

Pullman WE, Sullivan PJ, Barratt PJ, Lising J, Booth JA, Doe WF. Assessment of inflammatory bowel disease activity by technetium 99m phagocyte scanning. *Gastroenterology* 1988;**95**(4):989-96.

Ripolles 2013 *{published data only}*

Ripolles T, Rausell N, Paredes JM, Grau E, Martinez MJ, Vizuete J. Effectiveness of contrast-enhanced ultrasound for characterisation of intestinal inflammation in Crohn's disease: a comparison with surgical histopathology analysis. *Journal of Crohn's and Colitis* 2013;**7**(2):120-8.

Saverymuttu 1986 *{published data only}*

Saverymuttu SH, Camilleri M, Rees H. Indium 111-granulocyte scanning in the assessment of disease extent and disease activity in inflammatory bowel disease. A comparison with colonoscopy, histology, and fecal Indium 111-granulocyte excretion. *Gastroenterology* 1986;**90**:1121-8.

Scheurlen 1998 *{published data only}*

Scheurlen C, Allgayer H, Hardt M, Kruis W. Effect of short-term topical corticosteroid treatment on mucosal enzyme systems in patients with distal inflammatory bowel disease. *Hepatogastroenterology* 1998;**45**(23):1539-45.

Schmitz-Moormann 1988 *{published data only}*

Schmitz-Moormann P, Brandes JW, Jesdinsky H, Knapp M. Relationships between blood chemistry and histology of the rectal mucosa in Crohn's disease. *Pathology, Research and Practice* 1988;**183**(1):35-8.

Schunk 2001 *{published data only}*

Schunk K, Reiter S, Kern A, Orth T, Wanitschke R. Hydro-MRI in inflammatory bowel diseases: a comparison with colonoscopy and histology. *RoFo* 2001;**173**(8):731-8.

Sciarretta 1993 *{published data only}*

Sciarretta G, Furno A, Mazzoni M, Basile C, Malaguti P. Technetium-99m hexamethyl propylene amine oxime granulocyte scintigraphy in Crohn's disease: diagnostic and clinical relevance. *Gut* 1993;**34**(10):1364-9.

Seldenrijk 1991 *{published data only}*

Seldenrijk CA, Morson BC, Meuwissen SG, Schipper NW, Lindeman J, Meijer CJ. Histopathological evaluation of colonic mucosal biopsy specimens in chronic inflammatory bowel disease: diagnostic implications. *Gut* 1991;**32**(12):1514-20.

Shoenut 1994 *{published data only}*

Shoenut JP, Semelka RC, Magro CM, Silverman R, Yaffe CS, Micflikier AB. Comparison of magnetic resonance imaging and endoscopy in distinguishing the type and severity of inflammatory bowel disease. *Journal of Clinical Gastroenterology* 1994;**19**(1):31-5.

Shyn 2010 *{published data only}*

Shyn PB, Mortele KJ, Britz-Cunningham SH, Friedman S, Odze RD, Burakoff R, et al. Low-dose 18F-FDG PET/CT enterography: improving on CT enterography assessment of patients with Crohn disease. *Journal of Nuclear Medicine* 2010;**51**(12):1841-8.

Silva 2003 *{published data only}*

Silva MA, Menezes J, Wizman S, Gendron R, Oligny L, Seidman EG. Cytokine tissue levels as markers of disease activity in pediatric Crohn disease. *Pediatric Research* 2003;**54**:456-61.

Silva 2004 *{published data only}*

Silva MA, Lopez CB, Riverin F, Oligny L, Menezes J, Seidman EG. Characterization and distribution of colonic dendritic cells in Crohn's disease. *Inflammatory Bowel Diseases* 2004;**10**(5):504-12.

Smedh 1995 *{published data only}*

Smedh K, Olaison G, Franzen L, Sjodahl R. Endoscopic and external bowel changes and histopathology in patients with Crohn's disease. *British Journal of Surgery* 1995;**82**(2):191-4.

Smedh 1996 *{published data only}*

Smedh K, Olaison G, Franzen L, Sjodahl R. The endoscopic picture reflects transmural inflammation better than endoscopic biopsy in Crohn's disease. *European Journal of Gastroenterology and Hepatology* 1996;**8**(12):1189-93.

Smith 2010b *{published data only}*

Smith JP, Bingaman SI, Ruggiero F, Mauger DT, Mukherjee A, McGovern CO, et al. Naltrexone therapy improves activity and promotes mucosal healing in active Crohn's disease: A placebo-controlled trial. *Gastroenterology* 2010;**138**(5):S85.

Sobhani 1992 *{published data only}*

Sobhani I, Hochlaf S, Denizot Y, Vissuzaine C, Rene E, Benveniste J, et al. Raised concentrations of platelet activating factor in colonic mucosa of Crohn's disease patients. *Gut* 1992;**33**(9):1220-5.

Solem 2005 *{published data only}*

Solem CA, Loftus EV Jr, Tremaine WJ, Harmsen WS, Zinsmeister AR, Sandborn WJ. Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. *Inflammatory Bowel Diseases* 2005;**11**(8):707-12.

Steward 2012 *{published data only}*

Steward MJ, Punwani S, Proctor I, Adjei-Gyamfi Y, Chatterjee F, Bloom S, et al. Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. *European Journal of Radiology* 2012;**81**(9):2080-8.

Vieira 2009 *{published data only}*

Vieira A, Fang CB, Rolim EG, Klug WA, Steinwurz F, Rossini LG, et al. Inflammatory bowel disease activity assessed by fecal calprotectin and lactoferrin: correlation with laboratory parameters, clinical, endoscopic and histological indexes. *BMC Research Notes* 2009;**2**:221.

Wardle 1992 *{published data only}*

Wardle TD, Hall L, Turnberg LA. Use of coculture of colonic mucosal biopsies to investigate the release of eicosanoids by inflamed and uninflamed mucosa from patients with inflammatory bowel disease. *Gut* 1992;**33**(12):1644-51.

Additional references

Arnott 2002

Arnott ID, Watts D, Ghosh S. Review article: is clinical remission the optimum therapeutic goal in the treatment of Crohn's disease?. *Alimentary Pharmacology and Therapeutics* 2002;**16**(5):857-67.

Bouguen 2013

Bouguen G, Levesque BG, Feagan BG, Kavanaugh A, Peyrin-Biroulet L, Colombel JF, et al. Treat to target: a proposed new paradigm for the management of Crohn's disease. Clinical Gastroenterology and Hepatology 2015; Vol. 13, issue 6:1042-50.

D'Haens 2009

D'Haens GR, Fedorak R, Lémann M, Feagan BG, Kamm MA, Cosnes J, et al. Endpoints for clinical trials evaluating disease modification and structural damage in adults with Crohn's disease. *Inflammatory Bowel Diseases* 2009;**15**(10):1599-604.

Deyo 1991

Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials* 1991;**12**(4):142S-58S.

Dieleman 1998

Dieleman LA, Palmen MJ, Akol H, Bloemena E, Peña AS, Meuwissen SG, et al. Chronic experimental colitis induced by dextran sulphate sodium (DSS) is characterized by Th1 and Th2 cytokines. *Clinical and Experimental Immunology* 1998;**114**(3):385-91.

Geboes 2000

Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;**47**(3):404-9.

Hommes 2012

Hommes D, Colombel JF, Emery P, Greco M, Sandborn WJ. Changing Crohn's disease management: need for new goals and indices to prevent disability and improve quality of life. *Journal of Crohn's and Colitis* 2012;**6**:S224-34.

Kirshner 1985

Kirshner B, Guyatt G. A methodological framework for assessing health indices. *Journal of Chronic Diseases* 1985;**38**(1):27-36.

Korelitz 1984

Korelitz BI, Sommers SC. Response to drug therapy in Crohn's disease: evaluation by rectal biopsy and mucosal cell counts. *Journal of Clinical Gastroenterology* 1984;**6**(2):123-8.

Landis 1977

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**(1):159–74.

Langner 2014

Langner C, Magro F, Driessen A, Ensari A, Mantzaris GJ, Villanacci V, et al. The histopathological approach to inflammatory bowel disease: a practice guide. *Virchows Archiv: an International Journal of Pathology* 2014;**464**(5):511-27.

Mojtahed 2014

Mojtahed A, Khanna R, Sandborn WJ, D'Haens GR, Feagan BG, Shackelton LM, et al. Assessment of histologic disease activity in Crohn's disease: a systematic review. *Inflamm Bowel Diseases* 2014;**20**(11):2092-103.

Molander 2013

Molander P, Sipponen T, Kemppainen H, Jussila A, Blomster T, Koskela R, et al. Achievement of deep remission during scheduled maintenance therapy with TNF α-blocking agents in IBD. *Journal of Crohn's and Colitis* 2013;**7**(9):730-5.

Mosli 2017

Mosli MH, Feagan BG, Zou G, Sandborn WJ, D'Haens G, Khanna R, et al. Development and validation of a histological index for UC. *Gut* 2017;**66**(1):50-8.

Truelove 1956

Truelove SC, Richards WC. Biopsy studies in ulcerative colitis. *British Medical Journal* 1956;**1**:1315-18.

Zou 2005

Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. *Quality of Life Research* 2005;**14**(6):1545-52.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Agnholt 2003

Methods	26 patients with Crohn's disease and complicating ano-rectal fistulae received infliximab 2 pre-treatment (week 0) and 2 post-treatment (week 8) biopsies were scored 2 gastrointestinal pathologists performed the histopathological examination
Data	Patient characteristics: Female/male: 14/12 Age (range), female/male: 33 years (27 to 50), 40 years (18 to 67) Duration of disease: 8.2 years (1 to 31) CDAI: 170 (16 to 441) Number of readers: 2
Comparisons	Construct validation, responsiveness
Outcomes	See Table 7; Table 8
Notes	Scoring index evaluated: Agnholt Score (Modified GHAS - including only inflammatory changes)

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to clinical information; it is unclear whether they were blinded to endoscopic information
Independent observation?	Unclear	It was not reported whether observations were made independently; it is concerning that >1 pathologist was involved and yet inter-rater reliability was not calculated

Baert 1999

Methods	18 patients with steroid-refractory moderate-to-severe Crohn's disease were randomised to different doses of infliximab (n = 13) or placebo (n = 5) 2 to 4 ileal and 4 to 6 colonic biopsies in the vicinity of lesions or ulcerations were taken pre- (week 0) and post-treatment (week 4). Half of the biopsy specimens was stained with H&E for histologic scoring All biopsies examined by a single gastrointestinal pathologist

Baert 1999 *(Continued)*

Data	Patient characteristics: Female/male: 12/6 Age (range): placebo arm 32.6 years (25 to 41), infliximab arm 32.07 years (20 to 47) CDAI at baseline (range): placebo arm 290 (230 to 355), infliximab arm 329 (219 to 395) Disease location: colonic 6, ileocolonic, 10, ileitis 2 Number of readers: 1
Comparisons	Responsiveness
Outcomes	See Table 8
Notes	Scoring index evaluated: GHAS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to the treatment that the patients received, but it is unclear whether they were blinded to clinical or endoscopic information
Independent observation?	Unclear	Not applicable (single pathologist)

D'Haens 1999

Methods	30 patients with active refractory Crohn's disease (CDAI 220 to 400) were randomised to different doses of infliximab or placebo 4 ileal and 4 colonic biopsies in the vicinity of the most prominent ulcerative lesion were taken pre- (week 0) and post-treatment (week 4) from only 9 patients All biopsies examined by a single gastrointestinal pathologist in random order
Data	Patient characteristics (for the whole cohort - not described for patients with histologic assessment only): Female/male: 18/12 Age: placebo arm 34.4 years (+/- 9.8), infliximab arm 31.4 years (+/- 7.1) CDAI at baseline: placebo arm 276.9 (+/- 20.3), infliximab arm 316.8 (+/- 11.4) Number of readers: 1
Comparisons	Responsiveness
Outcomes	See Table 8
Notes	Scoring index evaluated: GHAS

Risk of bias

Item	Authors' judgement	Description

D'Haens 1999 *(Continued)*

Blinding?	Unclear	Blinding occurred but it is unclear whether pathologists were unaware of clinical information, endoscopic information, or both
Independent observation?	Unclear	Not applicable (single pathologist)

Drews 2009

Methods	32 patients with confirmed Crohn's disease Biopsies were obtained from the terminal ileum It is not reported who performed the histologic assessment
Data	Patient characteristics: Female/male: 18/14 Mean age (range): 38.75 years (17 to 71) Mean duration of Crohn's disease: 8.9 years (0 to 26) Mean CDAI: 193.3 (54 to 524) Number of readers: not reported
Comparisons	Construct validity
Outcomes	See Table 7
Notes	Scoring index evaluated: Drews Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	It was not reported whether blinding occurred
Independent observation?	Unclear	It is unclear whether a single pathologist was used, if so, independent observation is non-applicable

Geboes 2005

Methods	48 patients with active Crohn's disease (CDAI 220 to 400) received 5 mg/kg infliximab at week 0, at week 2 they were randomised to either placebo or infliximab (at week 2, 6 and every 8 weeks) with possible episodic re-treatment in case of worsening (European endoscopic substudy of ACCENT 1) Biopsies of sufficient quality of the ileum and 4 segments of colon in the vicinity of any lesional or ulcerated area (if no lesions were present in the segment, biopsies were collected at random sites within the segment) were taken at baseline (week 0) from 44 patients, week 10 from 43 patients and week 54 from 31 patients All biopsies examined in random order by a single pathologist
Data	Patient characteristics: Female/male: 28/20

Geboes 2005 *(Continued)*

	Median age: 30 years Median CDAI at baseline: 306 Median CDEIS at baseline: 9.8 Number of readers: 1
Comparisons	Responsivnes and construct validity
Outcomes	See Table 7; Table 8
Notes	Scoring index evaluated: CGHAS and IGHAS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Pathologists were fully blinded including clinical and endoscopic information
Independent observation?	Unclear	Not applicable (single pathologist)

Gomes 1986

Methods	22 patients with Crohn's colitis undergoing routine colonoscopic assessment At least one biopsy was taken from the area of maximum inflammation within each of the 6 sections of colon (cecum, the hepatic flexure, splenic flexure, descending colon, sigmoid colon, rectum) All biopsies were examined by one gastrointestinal pathologist
Data	Patient characteristics: Female/male: not reported Age: not reported Therapy: not reported Number of readers: 1
Comparisons	Criterion and construct validity
Outcomes	See Table 6; Table 7
Notes	Scoring index evaluated: Gomes Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Pathologists were fully blinded including clinical and endoscopic information
Independent observation?	Unclear	Not applicable (single pathologist)

Laharie 2011

Methods	51 patients with clinical remission off steroids within last 3 months with a single immunomodulator (methotrexate, azathioprine or infliximab) as a maintenance treatment. Biopsy samples were available for 40 patients. Biopsies were taken in the most inflamed areas and for the screening of dysplasia, the most inflammatory lesions observed were retained for scoring. All biopsies were examined by a single experienced gastrointestinal pathologist blinded to treatment
Data	Patient characteristics (for the whole cohort - not described for patients with histologic assessment only): Female/male: 38/13 Median age (range): methotrexate arm (n = 18) 40 years (21 to 56), azathioprine arm (n = 18) 48 years (21 to 89), infliximab arm (n = 15) 46 years (30 to 79) Median disease duration: methotrexate arm 9.1 years (1.1 to 29.5), azathioprine arm 12.9 years (6.5 to 27.7), infliximab arm 10.8 years (5.8 to 29.8) Disease location: colonic 24, ileocolonic 22, ileitis 3 Characteristics of patients with biopsies: 12 methotrexate, 14 azathioprine, 14 infliximab) Number of readers: 1
Comparisons	Construct Validity
Outcomes	See Table 7
Notes	Scoring index evaluated: Laharie Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to the treatment that the patients received, but it is unclear whether they were blinded to clinical or endoscopic information
Independent observation?	Unclear	Not applicable (single pathologist)

Mantzaris 2009

Methods	77 patients with steroid-dependent Crohn's ileocolitis or proximal colitis who achieved clinical remission on conventional steroids were randomised to azathioprine and budesonide with ileocolonoscopies at baseline and at the end of the study 4 biopsies were taken from each explored colonic segment and the terminal ileum, if entered. All biopsies were examined by a single experienced gastrointestinal pathologist
Data	Patient characteristics: Female/male: 43/34 Median age (range): azathioprine arm (n = 38) 34.3 years (19 to 59), budesonide arm (n = 39) 34.5 years (19 to 62) Mean (SD) disease duration: azathioprine arm 1.8 years (0.6), budesonide arm 1.9 years (0.7)

Mantzaris 2009 *(Continued)*

	Disease location: ileocolitis 50, proximal colitis 27 Number of readers: 1
Comparisons	Responsiveness
Outcomes	See Table 8
Notes	Scoring index evaluated: AHS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Blinding occurred but it is unclear whether pathologists were unaware of clinical information, endoscopic information, or both
Independent observation?	Unclear	Not applicable (single pathologist)

Naini 2012

Methods	Development of a histologic scoring system for IBD in 3 phases. Restrospective phase I (calibration) included 10 patients with a diagnosis of chronic ileocolitis each with at least 5 ileocolonic biopsies, retrospective phase II (validation) included 164 patients who underwent colonoscopy for clinical suspicion of ileocolitis each with at least 5 ileocolonic biopsies, and prospective phase III (application) included 30 patients to assess feasibility Biopsy samples were scored by 2 pathologist
Data	Patient characteristics: -phase I: male/female: 3/7, age range: 17 to 64 years, 7 of 10 had a clinical diagnosis of IBD. Isolated ileitis/isolated colitis/ileocolitis: 1/7/2. -phase II: male/female: 59/105, age range: 11 to 89 years, 22 had Crohn's disease, 13 ulcerative colitis, IBD unclassified 6, non-IBD chronic ileocolitis 12, chronic ileocolitis not otherwise specified 2, and no chronic ileocolitis 109 -phase III: unclear Number of readers: 2
Comparisons	Reliability, construct validity and feasibility
Outcomes	See Table 5; Table 7; Table 9
Notes	Scoring index evaluated: Naini and Cortina Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to clinical information; it is unclear whether they were blinded to endoscopic information

Naini 2012 *(Continued)*

Independent observation?	Yes	Phase I - observations were independent; phase 2 - not explicitly stated, but it is assumed that independent observations were also performed during this portion of the study; phase 3 - not applicable

Regueiro 2009

Methods	24 patients with Crohn's disease who underwent ileocolonic resection were assigned to receive infliximab or placebo for 1 year, followed by endoscopy with biopsies 6 to 8 biopsy specimens were taken from neoterminal ileum from the sites of endoscopically visualized lesions or, if no lesions present, a random sample All biopsies were examined by a single blinded gastrointestinal pathologist
Data	Patient characteristics: Female/male: 8/16 Infliximab arm 11, placebo arm 13 Median age: infliximab arm 43, years placebo arm 32 years Median duration of Crohn's disease: infliximab arm 13 years, placebo arm 9 years Median CDAI at baseline: infliximab arm 202, placebo arm 112 Number of readers: 1
Comparisons	Construct validity
Outcomes	See Table 7
Notes	Scoring index evaluated: Regueiro Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Blinding occurred but it is unclear whether pathologists were unaware of clinical information, endoscopic information, or both
Independent observation?	Unclear	Not applicable (single pathologist)

Sipponen 2008

Methods	61 adult patients with Crohn's disease had 87 endoscopies with biopsy collection 4 biopsy specimens, targeted at the most severely diseased areas, were taken from the ileum, right, transverse and left colon and rectum. If no lesions were present, biopsies were collected from random sites of the segment All biopsies were examined by a single pathologist
Data	Patient characteristics: Female/male: 31/30

Sipponen 2008 *(Continued)*

	Median age (range): 33 years (19 to 70) Median duration of Crohn's disease (range): 8 years (0 to 31.1) Median CDAI at baseline: 93 Number of readers: 1
Comparisons	Criterion validity
Outcomes	See Table 6
Notes	Scoring index evaluated: Sipponen Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	It was not stated whether blinding occurred
Independent observation?	Unclear	Not applicable (single pathologist)

Smith 2010a

Methods	34 patients with active Crohn's disease were randomised to receive 4.5 mg naltrexone daily or placebo for 12 weeks Biopsies were collected at baseline and at week 12 from five segments of each patient's GI tract (ileum, right colon, transverse colon, left colon, rectum) Biopsies were examined by a single pathologist
Data	Patient characteristics: Female %: naltrexone arm 64.7%, placebo arm 62.5% Naltrexone arm 18, placebo arm 16 Mean age (range): naltrexone arm 40.5 years (21 to 60), placebo arm 44.8 years (26 to 67) Median CDAI at baseline (+/- standard error of the mean): naltrexone arm 365 +/- 16, placebo arm 327 +/- 19 Numer of readers: 1
Comparisons	Responsiveness
Outcomes	See Table 8
Notes	Scoring index evaluated: Dieleman Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to the treatment that the patients received, but it is unclear whether they were blinded to clinical or endoscopic information

Smith 2010a *(Continued)*

Independent observation?	Unclear	Not applicable (single pathologist)

Ward 1977

Methods	64 rectal biopsies from 27 patients with Crohn's disease limited to the large bowel were reviewed retrospectively 64 rectal biopsies but only the first biopsies from each of the 27 patients were compared for prognosis Biopsies were examined by a single pathologist
Data	Patient characteristics: Female/male: 12/15 Mean age at onset (range): group Good 34.7 years (19 to 68), group Moderate 22.4 years (17 to 36), group Colectomy 39.1 years (18 to 49), group Death 38.5 years (13 to 62) Mean disease duration (range): 4.3 years (1 to 12) Number of readers: 1
Comparisons	Construct validity
Outcomes	See Table 7
Notes	Scoring index evaluated: Ward Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Pathologists were blinded to clinical information; it is unclear whether they were blinded to endoscopic information
Independent observation?	Unclear	Not applicable (single pathologist)

Yamamoto 2005

Methods	28 patients with active Crohn's disease were treated with elemental diet. Mucosal biopsies were obtained before and after the treatment Mucosal biopsy specimens were taken from the terminal ileum and multiple sites in the colon, including the macroscopically most severely affected site Number of pathologist who scored was not reported
Data	Patient characteristics: Female/male:12/16 Median age (interquartile range): 28 years (22 to 36) Median duration of Crohn's disease (interquartile range): 24 months (10 to 42) Median CDAI at baseline (interquartile range): 315 (265 to 358)

Yamamoto 2005 *(Continued)*	Number of readers: not reported
Comparisons	Responsiveness
Outcomes	See Table 8
Notes	Scoring index evaluated: Saverymuttu Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Blinding occurred but it is unclear whether pathologists were unaware of clinical information, endoscopic information, or both
Independent observation?	Unclear	It is unclear whether a single pathologist was used, if so, independent observation is non-applicable

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abad 2014	Does not include patients with Crohn's disease It was an animal study
Ajaj 2005	Not a validation study and it uses a stepwise score Study compares magnetic resonance colonography to histological disease activity using simple categories (normal, slight, moderate, or severe inflammation), but no data on the development or evaluation of operating properties of the histological scoring system
Allgayer 2007	Not a validation study and it uses a stepwise score Study correlates mucosal ornithine decarboxylase activity to mucosal inflammation using a score described by Binder No data on the development or evaluation of operating properties of the histological scoring system
Allison 1988	Not a validation study and it uses a stepwise score Study evaluates macrophage heterogeneity in colonic mucosa Histological disease activity in Crohn's disease patients was scored by a score in order of increasing severity (1-least inflamed, 12-most inflamed), but no data on the development or evaluation of the operating properties of this histological scoring system were reported
Ang 2000	The vast majority of included patients have ulcerative colitis and only 3 had Crohn's colitis Hisotology was assessed by the score described by Saverymuttu but the data were not reported separately for patients with ulcerative colitis and Crohn's disease The study compares unfractionated heparin to corticosteroids fo r treatment of colonic IBD
Baars 2010	Not a validation study

Study	Reason for exclusion
	Histological disease activity was assessed on Geboes scale which was validated for ulcerative colitis No data on the development or evaluation of operating properties of this histological scoring system were reported
Baars 2012	Not a validation study and study uses a stepwise score Histological disease activity in Crohn's disease patients was scored on a four-point scale (0-no histological disease activity; 1-mild active; 2-moderate active; and 3- severe active inflammation), but no data on the development or evaluation of the operating properties of this histological scoring system were reported
Bariol 2002	Study also included patients with ulcerative colitis and data were not reported separately for patients with ulcerative colitis and Crohn's disease 'Not a validation study Histological disease activity in Crohn's disease patients was scored before and after treatment with thalidomide using a numerical scoring system (range 0 to 8) similar to that developed by Truelove and Richards
Beattie 1994	Not a validation study and participants were children Histological disease activity in ileal mucosal biopsies of children with Crohn's disease was scored before and after treatment with enteral nutrition using a stepwise scoring system (range 0 to 3) No precise description of the scoring system was provided No data were reported on the development of the histological scoring system
Bergeron 2010	Not a validation study and study uses a stepwise score Histological disease activity was assessed by a score which was developed for ulcerative colitis by Rutter No data were reported on the development or evaluation of operating properties of this histological scoring system
Berni 2008	All included patients were children (aged 1 to 16 years) Study uses a histological disease activity score described by Saverymuttu
Bernstein 1993	Not a validation study A numerical histological scoring system was used to compare histological disease activity to mucosal substance P concentration No data were reported on the development or evaluation of operating properties of the histological scoring system
Binder 1970	Does not include patients with Crohn's disease, only patients with ulcerative colitis The histologic index used in this study was later used also in studies with Crohn's disease
Bojic 2011	Study used a stepwise score Histological disease activity was scored on a four-point scale (0 for remission; 3 for highest degree of inflammation) and correlated with fecal calprotectin Abstract only

Study	Reason for exclusion
Borody 2002	No specific histologic index was described Histological disease activity in Crohn's disease patients was evaluated before and after treatment with triple antimycobacterial therapy, but no specific histological scoring system was discussed or validated
Breese 1995	Not a validation study and study uses a stepwise score in children Study evaluated effect of treatment on lymphokine-secreting cells in children with Crohn's disease A simple histological score (0-normal to 3-severely inflamed) to assess histological activity was applied No data were reported on the development or evaluation of operating properties of the histological scoring system
Bruwer 2001	Not a validation study Study compares expression of metallothionein to histological disease activity using a histological score described by Saverymuttu No data on the development or evaluation of operating properties of this histological scoring system
Chiorean 2007	Not a validation study and study uses a stepwise histological score Study compares accuracy of CT enteroclysis with a scoring system for inflammatory and fibrostenotic features of CD based on surgical macro- and microscopic pathology No data on the validation or evaluation of operating properties of a histological scoring system
Chung-Faye 2011	Study used a stepwise score Histological disease activity in Crohn's disease patients was scored on a four-point scale (0-normal; 3-severe) and correlated with fecal calprotectin, but no data on the development of this histological scoring system was reported Abstract only
Chung-Faye 2014	Study used a stepwise score Histological disease activity in Crohn's disease patients was scored on a simple four-point scale (0-normal; 3-severe) and correlated with fecal calprotectin and CRP, but no data on the development of this histological scoring system was reported Abstract only
Colombel 2006	Not a validation study and study use d a stepwise score A simple histological scoring index with 4 categories (0-no inflammation; 3-severe ileitis) was compared with CT enterography for evaluation of small bowel inflammation. No data on the development or evaluation of operating properties of this histological scoring system were reported
Cosin 2011	Not a validation study and study used a stepwise score Histological disease activity scored on a four-point scale (1-4) was compared to number of HIP-2 cells in the lamina propria There were no data reported on the development or evaluation of operating properties of this histological scoring system

Study	Reason for exclusion
	Abstract only
D'Haens 1997	No specific histologic index was described The effect of azathioprine on the inflammatory lesions in neoterminal ileum was assessed but no specific histological scoring system was discussed or validated
D'Haens 1998	Not a validation study To examine intestinal mucosal inflammation induced by fecal contact in the excluded neoterminal ileum a histological scoring system with 8 items was described (GHAS) but no data on the development, validation or evaluation of the operating properties of this histological scoring system were provided
D'Haens 2001	Not a validation study Study scored for histological inflammation with the previously described GHAS (D'Haens 1999) in Crohn's disease patients before and after therapy with etanercept Because eta nercept is not a treatment of known efficacy for Crohn's disease responsiveness could not be assessed No other data on the development of this histological scoring system was reported
Di Sabatino 2009	Not a validation study and the study used a stepwise score Mucosal expression of matrix metalloproteinase in Crohn's disease was compared to histological disease activity before and after treatment with infliximab A simple score described by Wardle was used No data were reported on the development and validation of the histological scoring system
Dieckgraefe 2002	Not a validation study and the study used a stepwise score Mucosal Reg expression was compared to histological scoring of mucosal inflammation Two stepwise histopathologic scores were used; one for acute inflammation (range 0 to 3) and one for chronic inflammatory changes (range 0 to 3) No data were reported on the development or evaluation of operating properties of the histological scoring system
Ellrichmann 2012	Not a validation study and the study used a stepwise score Endoscopic ultrasound of the colon is compared to a 5-point histological inflammation score No data were reported on the development or evaluation of operating properties of the histological scoring system
Ewe 1999	Not a validation study and the study used a stepwise score The study evaluated budesonide for prevention of postoperative recurrence A 4-point histological score was used after treatment, but no data on evaluation of operating properties of the histological scoring system were reported
Fazio 1996	Not a validation study and the study used a stepwise score Effect of resection margins on the recurrence of Crohn's disease in the small bowel was evaluated

Study	Reason for exclusion
	No data were reported on the development or evaluation of operating properties of this histological scoring system which was not intended to measure disease activity
Ferrante 2006	Not a validation study Assessment of different histologic IBD features was done in resection specimens of patients with Crohn's disease Grading of neural hypertrophy and plexitis was performed, but no specific histological disease activity scoring system was discussed or validated
Folwaczny 1999	Not a validation study Study assessed histological disease activity using the score described by Gomes in patients receiving unfractionated heparin No data were reported on the development or evaluation of operating properties of this histological scoring system
Girlich 2011	Not a validation study: contrast-enhanced ultrasound Preoperative measurement of bowel wall vascularisation by contrast-enhanced ultrasound was compared to an advanced numerical histological scoring system with 20 items No data were reported on the development or evaluation of operating properties of the histological scoring system
Griga 1999	Not a validation study and the study used a stepwise score The study compared VEGF production in colonic mucosa to histological disease activity scored on a scale 0 to 3 (normal mucosa-severe chronic colitis) No data were reported on the development or evaluation of operating properties of this histological scoring system
Haber 2002	It uses a stepwise score. It compares ultrasonographic findings to clinical, endoscopic and histologic findings. According to the degree of histological inflammation segments were classified on a three-point scale (no, mild, or severe inflammation). No data on the development of this histological scoring system.
Hanski 1999	Not a validation study and it uses a stepwise score. Expression of MUC2 mucin was compared to mucosal inflammation. A histological disease activity score described by Matts was used in this study with a minority of patients with Crohn's disease. No data on the development or evaluation of operating properties of the scoring system was provided.
Hommes 1996	Not a validation study Soluble Fc gamma receptor III and eicosanoid concentrations were compared to histological mucosal inflammation A numerical score with 14 relevant parameters and a maximal score of 22 points was used which was poorly described No data were reported on the development or evaluation of operating properties of this histological scoring system
Iacucci 2015a	The original study does not include patients with Crohn's disease, only patients with ulcerative colitis

Study	Reason for exclusion
	A numerical histological scoring system (ECAP system) was designed to reflect all histological changes in IBD, including minor changes
Iacucci 2015b	Not a true validation study In this abstract the numerical histologic scoring system ECAP-CD is compared to a novel high definition colonoscopy imaging technique iSCAN However, no data on the development or relevant evaluation of operating properties of this histological scoring system were presented
Kaiser 2007	A stepwise score was used Fecal calprotectin was compared to disease activity including histologic A simple four-point histologic scale is used (0-no inflammation; 3-severe inflammation) No data were reported on the development of this histological scoring system
Kolkman 1996	Not a validation study and the study used a stepwise score The study comp ared computer tomography and granulocyte scintigraphy to disease activity A simple four-point histological scale was applied (0-no abnormalities or plain fibrosis; 3-severe ulcerations and inflammation) and incorporated in a disease activity score with endoscopic and operative findings No data were reported on the development or evaluation of operating properties of this histological scoring system
Koutroubakis 2003	Not a validation study and the study used a stepwise score The study c ompared 99mTc (V) DMSA scintigraphy to disease activity A histological score was incorporated in a disease inflammatory activity score as described by Kolkman No data were reported on the development or evaluation of operating properties of this histological scoring system
Kozarek 1989	Not a validation study and the study used a stepwise score A simple histologic score was used to assess disease activity of patients with Crohn's disease treated with methotrexate No data were reporte d on the development of this histological scoring system
Labaere 2013	Not a validation study Histological and endoscopic findings were compared to 8 different calprotectin assays, but no specific histological scoring system was discussed or validated Abstract only
Lasocki 2011	Not a histological validation study and a stepwise score was used Different MRI signs of Crohn's disease are compared to a gold standard of severity of inflammation with a six-point scale, based on histologic, surgical and colonoscopic findings No data were reported on the development or evaluation of operating properties of the histological scoring system

Study	Reason for exclusion
Leddin 1987	Not a histological validation study and a stepwise score was used 111-Indium-labelled leukocyte imaging and fecal excretion were compared to a assessment of disease activity which included histology A simple four-point histological scale was applied No data were reported on the development or evaluation of operating properties of this histological scoring system
Lenze 2012	Not a histological validation study and the study used a stepwise score Noninvasive imaging methods were compared to endoscopic and histological evaluation A simple three-point histological scale was applied for the degree of inflammatory infiltrate and another one for the degree of fibrosis No data were reported on the development of the histological scoring system
Levine 1993	Not a validation study and the st udy used a stepwise score Immunoglobulin therapy in inflammatory bowel disease was evaluated A simple histological scoring index with 4 categories (no inflammation to severe inflammation) was used, but was not specific for Crohn's disease (most patients had ulcerative colitis) No data were reported on the development or evaluation of operating properties of this histological scoring system
Licata 2012	A stepwise score was used Fecal calprotectin values we re compared to histological and endoscopic findings A simple three-point histological scale (0-normal mucosa, 2-severe inflammation) was applied No data were reported on the development of this scoring system
Lofberg 2002	Not a validation study In this trial for antisense NFkB p65 oligonucleotide treatment also biopsies were scored, but no specific histological scoring system was discussed or validated Abstract only
Maconi 2003	A stepwise score was used Clinical, biochemical and ultrasonographic evaluation was compared to individual histological features of surgical specimens of small bowel stenosis Histopathology was assessed using the criteria proposed by Fazio with some modifications No data were reported on the development of this scoring system
Macoritto 2012	Not a validation study Response to infliximab was defined as complete mucosal healing with a drop of at least 3 points on the histologic score but no specific histological scoring system was discussed or validated Abstract only
Mahmud 1996	A half of included patients had ulcerative colitis and data were not reported separately for patients with ulcerative colitis and Crohn's disease

Study	Reason for exclusion
	The study compared microalbuminuria, CRP, ESR and clinical disease activity to histological disease activity assessed by a histological score described by Saverymuttu
Matts 1961	This study does not include patients with Crohn's disease, only patients with ulcerative colitis In a later study the histologic index was used in patients with Crohn's disease
Mazzucchelli 1994	Not a validation study and the study used a stepwise score The study desc ribed expression of IL-8 in relation to histological disease activity using values of mild, moderate, or severe No data were reported on the development or evaluation of operating properties of this scoring system
Merra 2012	Not a validation study and the study used a stepwise score Response to therapy was evaluated using a histological score on a four-point scale described by Mikhailova and modified from Truelove No data were reported on the development or evaluation of operating properties of this scoring system
Migaleddu 2009	No specific histologic disease activity score was described Contrast-enhanced ultrasonographic evaluation was compared to inflammatory activity as classified by endoscopy and histology
Molnar 2009	Not a validation study The appearance of the ileocaecal valve was compared to endoscopic and histological findings in the ileum Histology was assessed by a numerical score on a scale of 1 to 13 by adding up scores for different parameters No data were reported on the development or evaluation of operating properties of this scoring system
Neaga 2009	No specific histological scoring system was discussed or validated] Fecal calprotectin values were compared to the degree of intestinal inflammation, defined by endoscopy and histology Abstract only
Neumann 2012	No specific histologic disease activity score was described Abstract only
Neumann 2013	Not a validation study Endocytoscopy was compared to histopathological assessment in a cohort of patients with ulcerative colitis and Crohn's disease The histological degree of mucosal inflammation was evaluated as described by Riley for ulcerative colitis No data were reported on the development or evaluation of operating properties of this scoring system

Study	Reason for exclusion
Nicholls 1994	The study did not use a specific histological scoring system, but assessed the degree of inflammation in pairs before and after treatment as worse, no change, improved, or resolution of inflammation
Nielsen 2003	Not a validation study The study evaluated colonic expression of interleukin-12 and -17 in relation to histological disease activity Presence of chronic and acute inflammation was determined on a simple 4-point ordinal scale, as well as overall assessment of degree of inflammation No data were reported on the development or evaluation of operating properties of this scoring system
Plesec 2009	Not a validation study The study focused on patients with ileal pouch-anal anastomosis including some with Crohn's disease However, no specific histological scoring system was discussed or validated Abstract only
Pullman 1988	Not a validation study Technetium 99m-phagocyte scanning was compared to histologic inflammation of mucosal biopsies The histologic score out of 12 was calculated from changes noted in the epithelium and crypts, and from the degree of mononuclear and neutrophil cellular infiltration No data were reported on the development or evaluation of operating properties of this scoring system
Ripolles 2013	Not a validation study: ultrasound (US) Several US parameters for evaluation of mural inflammation were compared to histopathology of surgical specimens Acute inflammation was assessed according to the Borley method with a score up to a maximum of 13 and fibrostenosis was graded using a modification of the Chiorean method No data were reported on the development and evaluation of operating properties of the histological scoring system
Saverymuttu 1986	The majority of included patients had ulcerative colitis and validation data were not reported separately for patients with ulcerative colitis and Crohn's disease The study compared indium 111-granulocyte scanning with endoscopy, histology, and fecal 111In-granulocyte excretion A new histological scoring system evaluating severity of changes in the enterocytes and crypts, and the cellularity of lamina propria was described
Scheurlen 1998	Not a validation study Histologic disease activity was assessed before and after topical corticosteroid treatment using a stepwise histological score described by Binder No data were reported on the development of the histological scoring system

Study	Reason for exclusion
Schmitz-Moormann 1988	No specific histological disease activity score was described The relationship of blood chemistry and CDAI to various histological variables was analysed
Schunk 2001	Not a validation study and the study used a stepwise score Hydro-MRI was compared with colonoscopy and biopsy specimens regarding the assessment of inflammatory activity A histologic scoring system with five categories was used (0-normal, 4-severe activity) No data were reported on the development and evaluation of operating properties of the histological scoring system
Sciarretta 1993	Not a validation study and the study used a stepwise score Scintigraphic activity was compared to endoscopic and histologic disease activity A simple four-point histological scale was used regarding the granulocyte infiltration No data were reported on the development and evaluation of operating properties of the histological scoring system
Seldenrijk 1991	No specific histological disease activity index was discussed or validated Evaluation of multiple histopathological features in colonic biopsy specimens was performed to assess reproducibility and to determine features with highest discriminatory power to distinguish IBD from non-IBD, and Crohn's disease from ulcerative colitis The study did not evaluate histological disease activity
Shoenut 1994	A stepwise score was used The study compared magnetic resonance imaging and endoscopy to histology The histologic degree of activity was determined by estimating the amount of acute inflammation (mild, moderate, or severe) No data were re ported on the development of the histological scoring system
Shyn 2010	Not a validation study and a stepwise score was used The aim of the study was to evaluate (18)F-FDG PET/CTE compared to CTE in the detection and grading of active inflammation in Crohn's disease Each pathology specimen was graded for the degree of inflammation using a 5-point scale (0-normal, 4-severe activity) No data were reported on the development and evaluation of operating properties of the histological scoring system
Silva 2003	Not a validation study and the st udy included a paediatric population The study investigated the association between mucosal inflammation and colonic tissue levels of various cytokines A histological scoring index assessing 6 features graded on a 4-point scale with the final score being the numerical sum of all the partial scores was used No data were reported on the development and evaluation of operating properties of the histological scoring system

Study	Reason for exclusion
Silva 2004	Not a validation study and the study included a paediatric population The study investigated the distribution and state of maturation of dendritic cells in the colon in relation to the severity of inflammation The histological numeric score used was described by Silva 2003 No data were reported on the development and evaluation of operating properties of the histological scoring system
Smedh 1995	Not a validation study of a histological score The study compared intraoperative endoscopic findings, external bowel changes, and transmural histopathology Many histological variables were recorded as present or absent but no specific histological disease score was discussed only a global assessment of histological severity on a visual analogue scale graded 0 to 100 (stepwise)
Smedh 1996	Not a validation study of a histological score The study investigated how endoscopic score and histological findings of endoscopic biopsies were related to that of transmural sections Similarly to Smedh 1995, many histological variables were recorded as present or absent but no specific histological disease score was discussed only a global assessment of histological severity on a visual analogue scale graded 0 to 100 (stepwise)
Smith 2010b	Not a validation study The study aimed to test the efficacy and safety of an opioid antagonist (naltrexone) in patients with Crohn's disease Histological activity was graded according to the criteria for inflammation described by Dieleman No data were reported on the development and validation of the histological scoring system
Sobhani 1992	Not a validation study and the study used a stepwise score Platelet activating factor was determined in colonic mucosal biopsies in patients with Crohn's disease Mucosal infiltration by inflammatory cells was evaluated as: absent, moderate, mild, or intense No data were reported on the development and evaluation of operating properties of the histological scoring system
Solem 2005	A stepwise score was used The aim of the study was to compare CRP with clinical, endoscopic, histological and radiological disease activity Histology was graded based upon the greatest severity of inflammation found into 5 categories No data were reported on the development of the histological scoring system
Steward 2012	Not a validation study of a histological score: MRI enterography The aim of the study was to develop and validate a scoring system for MRI enterography of small bowel Crohn's disease based on histopathological scoring

Study	Reason for exclusion
	The study used a transmural histological scoring of mural acute inflammation in surgical resection specimens adopted from the method of Borely as well as a histological scoring system for acute inflammation in endoscopic biopsies with 6 histological variables However, the study did not try to validate or evaluate operating properties of the histological scoring system
Vieira 2009	A stepwise score was used Calprotectin and lactoferrin was correlated with laboratory parameters, clinical, endoscopic and histological indexes Histological degree of inflammation was graded (0-absent, 1-slight, 2-moderate, or 3-severe) based upon the presence of neutrophils, the presence of erosion or ulceration or both, and crypt aggression No data were reported on the development of operating properties of the histological scoring system
Wardle 1992	Not a validation study and the study used a stepwise score Eicosanoid production was measured in cultured biopsies of colonic mucosa and compared to histological disease activity A simple grading of inflammatory cell infiltrate on a 4-point scale was used No data were reported on the development and evaluation of operating properties of the histological scoring system

ADDITIONAL TABLES

Table 1. Indices that have been fully or partially validated

	Study ID	Histologic index evaluated	GHAS or modification
1	Agnholt 2003	Agnholt Score	Yes
2	D'Haens 1999	Global Histological Activity Score (GHAS)	Yes
3	Drews 2009	Drews Score	No
4	Geboes 2005	Colonic Global Histological Activity Score (CGHAS)	Yes
5	Geboes 2005	Ileal Global Histological Activity Score (IGHAS)	Yes
6	Gomes 1986	Gomes Score	No
7	Laharie 2011	Laharie Score	Yes
8	Mantzaris 2009	Average Histologic Score (AHS)	Yes
9	Naini 2012	Naini and Cortina Score	No
10	Regueiro 2009	Regueiro Score	Yes

Table 1. Indices that have been fully or partially validated *(Continued)*

11	Sipponen 2008	Sipponen Score	Yes
12	Smith 2010a	Dieleman Score	No
13	Ward 1977	Ward Score	No
14	Yamamoto 2005	Savermuttu Index	No

Table 2. Scoring Index Cut-offs

Study ID	Index	Outcome	Cut-off
Agnholt 2003	Agnholt Score	histologic inflammatory activity	Not reported
Baert 1999	GHAS	histologic disease activity	Not reported
D'Haens 1999	GHAS	histologic disease activity	Not reported
Drews 2009	Drews Score	histologic inflammatory activity	Stage 0 (score 0): no increase in inflammation stage 1 (1-3): chronic, non-active inflammation stage 2 (4-6): mild active inflammation stage 3 (7-9): moderate active inflammation stage 4 (10-14): severe active inflammation
Geboes 2005	Gobes Score	histologic disease activity	Not reported
Gomes 1986	Gomes Score	histologic disease activity	Not reported. Maximum score: 24.
Laharie 2011	Laharie Score	histologic disease activity	Not reported. Score ranges from 1 (no activity) to 13
Mantzaris 2009	AHS	histologic disease activity	Not reported
Naini 2012	Naini and Cortina Score	likelihood of IBD/ histopathologic support for IBD	Probability of chronic ileitis (max. score = 10) low 0-2 moderate 3-4 high 5-10 Probability of chronic colitis (max. score =17) low 0-3 moderate 4-8 high 9-17
Regueiro 2009	Regueiro Score	histologic disease activity	inactive: 0 mildly active: 1-5 moderately active: 6-10

Table 2. Scoring Index Cut-offs *(Continued)*

			severely active: 11-14
Sipponen 2008	Sipponen Score	histologic disease activity	Not reported
Smith 2010a	Dieleman Score	histologic inflammatory activity	Not reported
Ward 1977	Ward Score	prognosis of Crohn's disease	Not reported. Maximum score 10.
Yamamoto 2005	Saverymuttu Score	histologic disease activity	Total score of 4 items: 0-1 - Grade 0 (none inflammation) 2-4 - Grade 1 (mild inflammation) 5-8 - Grade 2 (moderate inflammation) 9-12 - Grade 3 (severe inflammation) Histologic healing was defined as grade of 0 Histologic improvement (including healing).was defined as a decrease of at least 1 grade

Table 3. Summary of operating properties of histologic scoring indices for Crohn's disease

Scoring index	Validity			Reliability				Respon-siveness	Feasibili-ty
	Content validity	Criterion validity	Construct validity	In-tra-rater	In-ter-rater	Test-retest	Internal consis-tency		
AHS	?	?	?	?	?	?	?	+	?
CGHAS	?	+	?	?	?	?	?	+	?
Dieleman Score	?	?	?	?	?	?	?	+	?
GHAS	?	?	+/-	?	?	?	?	+	?
Drews Score	?	?	-	?	?	?	?	?	?
Gomes Score	?	+/-	-	?	?	?	?	?	?
Saverymuttu Score	?	?	?	?	?	?	?	+	?
Ward Score	?	+/-	?	?	?	?	?	?	?
IGHAS	?	-	+/-	?	?	?	?	+	?
Sipponen Score	?	+	?	?	?	?	?	?	?
Laharie Score	?	-	?	?	?	?	?	?	?
Regueiro Score	?	+	?	?	?	?	?	?	?
Agnholt Score	?	?	+	?	?	?	?	+	?
Naini and Cortina Score	?	+	?	?	+	?	?	?	+

+ positive rating
? no information or indeterminate rating
- Negative rating

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist)

	A	B	C	D	E	F	G	H	I	J	
Study ID	**Internal consistency**	**Reliability**	**Measurement error**	**Construct validity**	**Factor analysis**	**Hypothesis testing**	**Cross-cultural validity**	**Criterion validity**	**Responsiveness**	**Interpretability**	**Generlisability**
Agnholt 2003	?	?	?	?	?	poor	?	?	poor	the distribution of scores: pre-treatment 2.0 (IQR 1-3) post-treatment 1.0 (IQR 0-3)	1) age (range): female/male: 33 (27-50), 40 (18-67) 2) female/male: 14/12 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: prospective single-centre trial at Aarhus university hospital 5) country: Denmark 6) language: not relevant
Baert 1999	?	?	?	?	?	?	?	?	poor	the distribution of scores: A) Colonic Mean decrease in colonic specimens: infliximab arm (n = 11): from 8.3 (1-12) to 3.4 (0-7) placebo arm (n = 2): from 6.5 (6-7) to 6 (5-7) B) Ileal:	1) Age (range): placebo arm 32.6 (25-41), infliximab arm 32.07 (20-47) 2) Female/male: 12/6 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: Single-centre, placebo-controlled, randomised trial at University hospital Gasthuisberg, Leuven 5) country: Belgium

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

										Mean decrease in ileal specimens: infliximab arm (n = 6): from 8.2 (2-12) to 2.3 (0-6) placebo arm (n = 4): from 6.8 (1-10) to 6.3 (4-10)	6) language: not relevant
D'Haens 1999	?	?	?	?	?	?	?	?	poor	the distribution of scores: A) Colonic infliximab arm (n = 7): before 8.8+/-1.7 (range, 2-10) and after infusion of infliximab 2.7+/-1.7 (range, 0-8) placebo arm (n = 4): before 11.0+/-2.3 and after infusion of placebo 9.0+/-1.9 B) Ileal: infliximab arm (n = 4): before 7.7+/-2.3 (range, 4-12) and after infusion of infliximab 3.3+/-2.0 (range, 0-7) placebo arm (n = 3): before 9.2+/-2.4 and after infusion of placebo 8.7+/-3.1	data for the whole cohort - not described for patients with histologic assessment only: 1) mean age: placebo arm 34.4 (+/- 9.8), infliximab arm 31.4 (+/- 7.1) 2) Female/male: 18/12 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: international multi-centre, randomised, placebo-controlled trial, biopsies specimens taken only in one centre (University hospital Gasthuisberg, Leuven) 5) countries: biopsies only in Belgium 6) language: not relevant

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

Drews 2009	?	?	?	?	?	fair	?	?	?	the distribution of scores (n = 32): stage 0: 8 (25%) stage 1: 5 (16%) stage 2: 10 (31%) stage 3: 3 (9%) stage 4: 6 (19%)	1) Mean age (range): 38.75 (17-71) 2) Female/male: 18/14 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: retrospective analysis from a single centre 5) countries: Germany 6) language: not relevant
Geboes 2005	?	?	?	?	?	Fair (CGHAS) and poor (IGHAS)	?	Fair (CGHAS) and poor (IGHAS)	poor	the distribution of scores: A) Median CGHAS SIngle dose/ episodic group: 10 (week 0), 4 (week 10) and 5 (week 54) Combined maintenance/episodic group: 9 (week 0), 2 (week 10) and 2 (week 54) B) Median IGHAS SIngle dose/ episodic group: 1 (week 0), 1 (week 10) and 2 (week 54)	1) Median age: 30 2) Female/male: 28/20 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: international multi-centre, randomised, controlled trial in European centres 5) countries: European (not reported which countries) 6) language: not relevant

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

										Combined maintenance/episodic group: 2 (week 0), 0 (week 10) and 0 (week 54)	
Gomes 1986	?	?	?	?	?	Poor	?	Poor	?	the distribution of histologic scores: range 0-12	1) Age: not reported 2) Female/male: not reported 3) disease characteristics and duration of treatment: not reported 4) setting: prospective, single centre trial 5) countries: USA 6) language: not relevant
Laharie 2011	?	?	?	?	?	?	?	fair	?	the distribution of scores: methotrexate arm (n = 12): median 2.5, range 1-12 azathioprine arm (n = 14): median 3, range 0-11 infliximab arm (n = 14): median 3.5, range 0-12	data for the whole cohort - not described for patients with histologic assessment only: 1) Median age (range): methotrexate arm (n = 18) 40 (21-56), azathioprine arm (n = 18) 48 (21-89), infliximab arm (n = 15) 46 (30-79) 2) Female/male: 38/13 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: prospective single-centre study at the Haut-Leveque hospital 5) country: France

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

											6) language: not relevant
Mantzaris 2009	?	?	?	?	?	?	?	?	fair	the distribution of scores: Azathioprine arm: from 5.92+/-1.7 at baseline to 2.92+/-1.93 at study termination Budesonide arm: from 5.72+/-1.63 at baseline to 6.01+/-1.72 at study termination	1) Median age (range): azathioprine arm (n = 38) 34.3 (19-59), budesonide arm (n = 39) 34.5 (19-62) 2) Female/male: 43/34 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: prospective single-centre, randomised, controlled study in Athens 5) country: Greece 6) language: not relevant
Naini 2012	?	good	?	?	?	?	?	poor	?	1) the distribution of scores (total and IBD): Ileal scores (max. of 10) ranged from 0 to 8 Colonic scores (max. of 17) ranged from 0 to 13 2) scores for relevant subgroups (ileal score: mean (SD), colonic score: mean (SD)) ulcerative colitis 0,65 (1.5), 9.4 (2.8)	1) mean age (range) for subgroups: ulcerative colitis 26 (11-43), Crohn's disease 32 (13-57), IBD unclassified 34 (17-51), non-IBD chronic ileocolitis 53 (25-81), non-specific chronic ileocolitis 53 (43-63), no chronic ileocolitis 47 (15-89) 2) male/female: 59/105 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: study with 2 retrospective phases (calibra-

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

										Crohn's disease 3.5 (2.4), 5.2 (3.6) IBD unclassified 0, 5.9 (3.7) non-IBD chronic ileocolitis 0.5 (0.7), 3.4 (2.9) nonspecific chronic ileocolitis 3.0 (3.5), 4.5 (1.7) no chronic ileocolitis 0.3 (0.8),1.0 (1.5)	tion and validation) and a prospective phase (application) in a single centre 5) country: USA 6) language: not relevant
Regueiro 2009	?	?	?	?	?	?	?	poor	?	the distribution of scores: score 0-3: 12 patients score 4-6: 3 patients score 7-14: 9 patients	1) Median age: infliximab arm 43, placebo arm 32 2) Female/male: 8/16 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: randomised, two-armed, double blinded, placebo-controlled trial at the University of Pittsburgh Medical Centre 5) country: USA 6) language: not relevant
Sipponen 2008	?	?	?	?	?	?	?	good	?	the distribution of scores: in ileocolonic or colonic disease, median total histology score was 15 (range 0-45),	1) Median age (range): 33 (19-70) 2) Female/male: 31/30 3) disease characteristics and duration of treatment: see table Characteristics of included studies

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

										ileal histology score 4 (0-11) and histology score of the colon 13 (0-38). In the ileal disease, ileal histology score was 6 (0-10).	4) setting: prospective single-center study 5) country: Finland 6) language: not relevant
Smith 2010a	?	?	?	?	?	?	?	?	poor	the distribution of scores: Naltrexone arm (n = 18): from 16 before treatment to 6 after treatment Placebo arm (n = 16): 18 before and after treatment	1) Mean age (range): naltrexone arm 40.5 (21-60), placebo arm 44.8 (26-67) 2) Female %: naltrexone arm 64.7%, placebo arm 62.5 % 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: prospective, double blind, randomised placebo-controlled trial at a single centre 5) countries: USA 6) language: not relevant
Ward 1977	?	?	?	?	?	?	?	poor	?	the distribution of scores (n = 27): group 1: 3.0 (0.7-6) group 2: 3.2 (0.7-5.3) group 3: 3.5 (1.7-6.7)	1) Mean age at onset (range): group Good 34.7 (19-68), group Moderate 22.4 (17-36), group Colectomy 39.1 (18-49), group Death 38.5 (13-62) 2) Female/male: 12/15 3) disease characteristics and duration of treatment: see table Characteristics of included studies

Table 4. The methodological quality of histologic index measurement properties as described in the original development articles (COSMIN checklist) *(Continued)*

										group 4: 4.9 (3.3-7.3)	4) setting: retrospective study at a single centre 5) countries: Scotland, United Kingdom 6) language: not relevant
Yamamoto 2005	?	?	?	?	?	?	?	?	poor	the distribution of scores: A) terminal ileum Before treatment (n = 28): grade 0: 2 (7%), grade 1: 8 (29%), grade 2: 11 (39%), grade 3: 77 (25%) After treatment (n = 28): grade 0: 7 (25%), grade 1: 8 (29%), grade 2: 10 (36%), grade 3: 3 (11%) B) colon Before treatment (n = 28): grade 0: 8 (29%), grade 1: 8 (29%), grade 2: 7 (25%), grade 3: 5 (18%) After treatment (n = 28): grade 0: 12 (43%), grade 1: 8 (29%), grade 2: 6 (21%), grade 3: 2 (7%)	1) Median age (interquartile range): 28 (22-36) 2) Female/male:12/16 3) disease characteristics and duration of treatment: see table Characteristics of included studies 4) setting: prospective, single centre, pilot trial 5) countries: Japan 6) language: not relevant

Table 5. Reliability

Study ID	Index	Inter-rater Kappa (between raters)	Inter-rater ICC (between raters)	Intra-rater Kappa (within rater)	Intra-rater ICC (within rater)	Internal Consistency
Naini 2012	Naini and Cortina score	-	Phase I (before score modification): Ileitis ICC: 0.96 Colitis ICC: 0.95 Phase II Ileitis ICC: 0.94 Colitis ICC: 0.94	-	-	-

Abbreviations: ICC, interclass correlation coefficient

Table 6. Criterion Validity

Study ID	Index	Outcome	Correlation
Gomes 1986	Histology score by Gomes	CRP	r = 0.34 (P valued not stated)
Gomes 1986	Histology score by Gomes	ESR	r = 0.25 (P valued not stated)
Sipponen 2008	Ileal GHAS (modified GHAS by Sipponen) Colonic GHAS (modified GHAS by Sipponen)	Faecal calprotectin (quantitative enzyme immunoassay, PhiCal Test, Calpro AS, Oslo, Norway, values < 100 mcg/g stoll considered normal)	Ileal GHAS: r = 0.311 (P valued not stated) Colonic GHAS: r = 0.563 (P < 0.01)
Sipponen 2008	Ileal GHAS (modified GHAS by Sipponen) Colonic GHAS (modified GHAS by Sipponen)	Faecal lactoferrin (quantitative enzyme immunoassay, IBD-SCAN, Inverness Medical, Princeton, NJ, USA, Techlab, Blacksburg, VA, USA, values < 7.25 mcg/g stoll considered normal)	Ileal GHAS: r = 0.291 (P valued not stated) Colonic GHAS: r = 0.543 (P < 0.01)

Table 7. Construct Validity

Study ID	Index	Comparison	Correlation
Agnholt 2003	Agnholt Score	CDAI	week 0: r = 0.5 (P = 0.019) week 8: r = 0.6 (P = 0.002)
Drews 2009	Drews Score	CDAI	no significant association (P = 0.2482)
Geboes 2005	Geboes Score (colonic)	CDAI	change from week 0 to week 10: r = 0.43 (P = 0.006, n = 40)

Table 7. Construct Validity *(Continued)*

			change from week 0 to week 54: r = 0.10 (P = 0.700, n = 17)
Geboes 2005	Geboes Score (ileal)	CDAI	change from week 0 to week 10: r = 0.27 (P = 0.155, n = 29) change from week 0 to week 54: r = 0.68 (P = 0.001, n = 13)
Geboes 2005	Geboes Score (colonic)	IBDQ	change from week 0 to week 10: r = -0.33 (P = 0.037, n = 40) change from week 0 to week 54: r = -0.04 (P = 0.873, n = 17)
Geboes 2005	Geboes Score (ileal)	IBDQ	change from week 0 to week 10: r = -0.29 (P = 0.132, n=29) change from week 0 to week 54: r = -0.35 (P = 0.246, n=13)
Geboes 2005	Geboes Score (colonic)	CDEIS	at week 0: r = 0.56 (P < 0.001, n = 45) change from week 0 to week 10: r = 0.52 (P = 0.001, n = 40) change from week 0 to week 54: r = 0.40 (P = 0.110, n = 17)
Geboes 2005	Geboes Score (ileal)	CDEIS	change from week 0 to week 10: r = 0.05 (P = 0.78, n = 29) change from week 0 to week 54: r = 0.42 (P = 0.153, n = 13)
Gomes 1986	Gomes Score	HBI	no correlation (P value not reported)
Gomes 1986	Gomes Score	Gomes Score (endoscopic)	r = 0.76 (P < 0.001)
Laharie 2011	Laharie Score	CDEIS	r = 0.154 (P = 0.344)
Laharie 2011	Laharie Score	SES-CD	No correlation but no numbers reported
Naini 2012	Naini and Cortina Score	likelihood of IBD/ histopathologic support for IBD	Positive predictive value for ileal involvement by IBD: ileal cut-off score of 5: 88% ileal scores of 3 or 4: 53% ileal scores of 2 or less: 19% Positive predictive value for colonic involvement by IBD: colonic cut-off score of 9: 92% colonic scores of 4 to 8: 52% colonic scores of 3 or less: 8%

Table 7. Construct Validity *(Continued)*

Regueiro 2009	Regueiro Score	Rutgeerts Score (endoscopic postoperative disease activity score)	r = 0.73 (P < 0.0001)
Sipponen 2008	Sipponen Score	SES-CD (ileal)	Ileal GHAS: r = 0.779 (P < 0.01)
Sipponen 2008	Sipponen Score	SES-CD (colonic - the sum of the scores for 4 colonic segments)	Colonic GHAS: r = 0.759 (P < 0.01)
Ward 1977	Ward Score	Prognosis of Crohn's disease as defined by 4 groups of clinical outcomes: group 1 (good): remaining asymptomatic at latest review group 2 (moderate): continued symptomatic activity together with evidence on sigmoidoscopy of continued inflammation group 3 (colectomy): disease process of severity sufficient to require colectomy but who subsequently remained in good general health group 4 (death): died as a direct result of the severity of the granulomatous colitis	Score of four clinical groups for first biopsies (n = 27): group 1: 3.0 (difference to group 4: P < 0.05) group 2: 3.2 (difference to group 4: P < 0.05) group 3: 3.5 group 4: 4.9

Abbreviations: CDAI, Crohn's Disease Activity Score; CDEIS, Crohn's Disease Endoscopic Index of Severity; HBI, Harvey-Bradshaw Index; IBDQ, Inflammatory Bowel Disease Questionnaire; SES-CD, Simple Endoscopic Score for Crohn's Disease

Table 8. Responsiveness

Study ID	Index	Treatment	Correlation with Responsiveness Meaure
Agnholt 2003	Agnholt Score	Infliximab 5 mg/kg body weight at week 0, 2 and 6	Median decrease: from 2.0 (interquartile range (IQR) = 1-3) to 1.0 (IQR 0-3), P = 0.011
Baert 1999	GHAS	Infliximab (5, 10 or 20 mg/kg body weight) as a single infusion or placebo	Mean decrease in colonic specimens: infliximab arm (n = 11): from 8.3 to 3.4, P = ? placebo arm (n = 2): from 6.5 to 6, P = ? Mean decrease in ileal specimens: infliximab arm (n = 6): from 8.2 to 2.3, P = ? placebo arm (n = 4): from 6.8 to 6.3, P = ?

Table 8. Responsiveness *(Continued)*

D'Haens 1999	GHAS	Infliximab (5, 10 or 20 mg/kg body weight) as a single infusion or placebo	Mean decrease in colonic specimens: infliximab arm (n = 7): from 8.8+/-1.7 (range, 2-10) to 2.7+/-1.7 (range, 0-8), P < 0.01 placebo arm (n = 4): from 11.0+/-2.3 to 9.0+/-1.9, P > 0.05 Mean decrease in ileal specimens: infliximab arm (n = 4): from 7.7+/-2.3 (range, 4-12) to 3.3+/-2.0 (range, 0-7), P < 0.01 placebo arm (n = 3): from 9.2+/-2.4 to 8.7+/-3.1, P > 0.05
Geboes 2005	CGHAS (colonic GHAS)	SIngle dose/episodic group (5 mg/kg infliximab at week 0, placebo at week 2, 6 and every 8 weeks with possible episodic infliximab re-treatment in case of worsening) Combined maintenance/episodic group (5 mg/kg at week 0, 2 and 6, after that every 8 weeks 5 or 10 mg/kg with possible episodic re-treatment in case of worsening)	Mean decrease in single dose/episodic group: from 10 (week 0, n = 15) to 4 (week 10, n = 15, P < 0.05) and 5 (week 54, n = 10, P > 0.05) Mean decrease in the combined maintenance/episodic group: from 9 (week 0, n = 25) to 2 (week 10, n = 25, P < 0.001) and 2 (week 54, n = 18, P < 0.001)
Geboes 2005	IGHAS (ileal GHAS)	SIngle dose/episodic group (5 mg/kg infliximab at week 0, placebo at week 2, 6 and every 8 weeks with possible episodic infliximab re-treatment in case of worsening) Combined maintenance/episodic group (5 mg/kg at week 0, 2 and 6, after that every 8 weeks 5 or 10 mg/kg with possible episodic re-treatment in case of worsening)	Mean decrease in single dose/episodic group: from 1 (week 0, n = 10) to 1 (week 10, n = 10, P > 0.05) and 2 (week 54, n = 9, P > 0.05) Mean decrease in the combined maintenance/episodic group: from 2 (week 0, n = 19) to 0 (week 10, n = 19, P < 0.001) and 0 (week 54, n = 14, P > 0.05)
Mantzaris 2009	AHS (average GHAS per intestinal segment)	Azathioprine 2-2.5 mg/kg a day (n = 38) or budesonide 6-9 mg/day (n = 39)	Azathioprine arm: from 5.92+/-1.7 at baseline to 2.92+/-1.93 at study termination (intention to treat analysis: P < 0.01) Budesonide arm: from 5.72+/-1.63 at baseline to 6.01+/-1.72 at study termination (ITT analysis: P = 0.31)
Smith 2010a	Dieleman Score	Naltrexone 4.5 mg orally or placebo for 12 weeks	Naltrexone arm: from 16 to 6 (P = 0.016) Placebo arm:

Table 8. Responsiveness *(Continued)*

			18 before and after treatment (N.S.) When former placebo-treated patients were subsequently administered naltrexone for 12 week in extended open-label study a significant improvement in histology was observed (P = 0.006, other data not reported)
Yamamoto 2005	Saverymuttu Score	Enteral nutrition Elentel (1 kcal/mL, 760 mOsm/L). Adaptation phase (concentration gradually increased from 1/3 to full strength) 7 days and maintenance phase (at the full strength) for 4 weeks	terminal ileum Before treatment (n = 28): grade 0: 2 (7%), grade 1: 8 (29%), grade 2: 11 (39%), grade 3: 77 (25%) After treatment (n = 28): grade 0: 7 (25%), grade 1: 8 (29%), grade 2: 10 (36%), grade 3: 3 (11%) colon Before treatment (n = 28): grade 0: 8 (29%), grade 1: 8 (29%), grade 2: 7 (25%), grade 3: 5 (18%) After treatment (n = 28): grade 0: 12 (43%), grade 1: 8 (29%), grade 2: 6 (21%), grade 3: 2 (7%)

Table 9. Feasibility

Study ID	Index	Feasibility Scoring
Naini 2012	Naini and Cortina score	The scoring worksheet was easy to use, could be appropriately applied and was easily completed in less than 30 seconds for each patient

APPENDICES

Appendix 1. Search Strategies for MEDLINE, EMBASE and CENTRAL databases

MEDLINE (1950 – current)

1. Crohn's Disease.mp. or ileitis.mp. or exp Crohn Disease/

2. (histopath* or histol* or pathology or immunopathology or immunohistochemistry or biops*or microscop*).mp.

3. exp Pathology, Clinical/

4. exp Immunohistochemistry/

5. exp Biopsy

6. or/2-5

7. ((score* or scori* or (scale* or scali*) or (index* or indice*) or (grade* or gradi*)).mp.

8. 1 and 6 and 7

EMBASE (1974 to current)

1. Crohn's Disease.mp. or ileitis.mp. or exp Crohn Disease/

2. (histopath* or histol* or pathology or immunopathology or immunohistochemistry or biops*or microscop*).mp.

3. exp Pathology, Clinical/

4. exp Immunohistochemistry/

5. exp Biopsy

6. or/2-5

7. ((score* or scori* or (scale* or scali*) or (index* or indice*) or (grade* or gradi*)).mp.

8. 1 and 6 and 7

Cochrane Library (CENTRAL)

1. MeSH descriptor: [Crohn Disease] explode all trees

2. Crohn*

3. #1 or #2

4. histol* or pathol* or immunohisto* or biops*

5. score* or scori* or scale* or scali* or index* or indice* or grade* or gradi*

6. #3 and #4 and #5

IBD/FBD Specialized Register

1. Crohn* AND (histol* or pathol* or immunohisto* or biops*)

CONTRIBUTIONS OF AUTHORS

All authors helped in the development of the protocol.

DECLARATIONS OF INTEREST

Several authors (BGF, RK, WJS) were involved in the development of the Robarts score. Robarts Clinical Trials began in 1986 as an academic research unit within the Robarts Research Institute which is affiliated with University Hospital and the University of Western Ontario. All profits from Robarts Clinical Trials, Inc. are directed towards academic research. The University of Western Ontario is the sole shareholder of Robarts Clinical Trials Inc. None of the authors with affiliation to Robarts Clinical Trials, Inc. have an equity position or any shares in the corporation.

Claire E Parker, Rish K. Pai, and John K MacDonald have no known conflicts.

Gregor Novak has received payment for lectures from Abbvie, Takeda, Ferring, Dr Falk Pharma and Merck.

Reena Khanna has received consulting fees from AbbVie, Janssen, Pfizer, Shire, and Takeda.

Brian Feagan has served on the board of Abbott/AbbVie, Amgen, Astra Zeneca, Avaxia Biologics Inc., Bristol-Myers Squibb, Celgene, Centocor Inc., Elan/Biogen, Ferring, JnJ/Janssen, Merck, Novartis, Novonordisk, Pfizer, Prometheus Laboratories, Protagonist, Salix Pharma, Takeda, Teva, Tillotts Pharma AG, and UCB Pharma; has received consultancy fees from Abbott/AbbVie, Actogenix, Albireo Pharma, Amgen, Astra Zeneca, Avaxia Biologics Inc., Axcan, Avir Pharma, Baxter Healthcare Corp., Biogen Idec, Boehringer-Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan/Biogen, EnGene, Ferring Pharma, Roche/Genentech, GiCare Pharma, Gilead, Given Imaging Inc., GSK, Ironwood Pharma, Janssen Biotech (Centocor), JnJ/Janssen, Kyowa Kakko Kirin Co Ltd., Lexicon, Lilly, Merck, Millennium, Nektar, Novonordisk, Pfizer, Prometheus Therapeutics and Diagnostics, Protagonist, Receptos, Salix Pharma, Serono, Shire, Sigmoid Pharma, Synergy Pharma Inc., Takeda, Teva Pharma, Tillotts, UCB Pharma, Vertex Pharma, Warner-Chilcott, Wyeth, Zealand, and Zyngenia; and has received lecture fees from Abbott/AbbVie, JnJ/Janssen, Takeda, Warner-Chilcott, and UCB Pharma.

William J Sandborn has received consultancy fees from Abbott Laboratories, ActoGeniX NV, AGI Therapeutics, Inc., Alba Therapeutics Corporation, Albireo,
Alfa Wasserman, Amgen, AM-Pharma BV, Anaphore, Astellas Pharma, Athersys, Inc., Atlantic Healthcare Limited, Axcan Pharma (now Aptalis), BioBalance Corporation, Boehringer-Ingelheim Inc, Bristol Meyers Squibb: (both money paid to WS and institution), Celegene, Celek Pharmaceuticals, Cellerix SL, Cerimon Pharmaceuticals, ChemoCentryx, CoMentis, Cosmo Technologies, Coronado Biosciences, Cytokine Pharmasciences, Eagle Pharmaceuticals,
Eisai Medical Research Inc., Elan Pharmaceuticals: (both money paid to WS and institution), EnGene, Inc., Eli Lilly, Enteromedics: (both money paid to WS and institution), Exagen Diagnostics, Inc., Ferring Pharmaceuticals, Flexion Therapeutics, Inc., Funxional Therapeutics Limited, Genzyme Corporation, Genentech (now Roche): (both money paid to WS and institution), Gilead Sciences, Given Imaging, Glaxo Smith Kline, Human Genome Sciences, Ironwood Pharmaceuticals (previously Microbia Inc.), Janssen (previously Centocor): (both money paid to WS and institution), KaloBios Pharmaceuticals, Inc., Lexicon Pharmaceuticals, Lycera Corporation, Meda Pharmaceuticals (previously Alaven Pharmaceuticals), Merck Research Laboratories, MerckSerono, Millennium Pharmaceuticals (subsequently merged with Takeda): (both money paid to WS and institution), Nisshin Kyorin Pharmaceuticals Co., Ltd., Novo Nordisk A/S,
NPS Pharmaceuticals, Optimer Pharmaceuticals, Orexigen Therapeutics, Inc., PDL Biopharma: (money paid to institution), Pfizer: (both money paid to WS and institution), Procter and Gamble: (both money paid to WS and institution), Prometheus Laboratories, ProtAb Limited, Purgenesis Technologies, Inc., Relypsa, Inc., Salient Pharmaceuticals, Salix Pharmaceuticals, Inc., Santarus, Schering Plough Corporation (acquired by Merck), Shire Pharmaceuticals: (money paid to institution), Sigmoid Pharma Limited, Sirtris Pharmaceuticals, Inc. (a GSK company), S.L.A. Pharma (UK) Limited, Takeda: (both money paid to WS and institution), Targacept, Teva Pharmaceuticals, Therakos, Tillotts Pharma AG (acquired by Zeria Pharmaceutical Co., Ltd), TxCell SA, UCB Pharma: (both money paid to WS and institution), Viamet Pharmaceuticals, Vascular Biogenics Limited (VBL), Warner Chilcott UK Limited, and Wyeth (now Pfizer); fees for expert testimony from Dickinson, Prud'Homme, Adams & Ingram; grants/grants pending from Abbott Laboratories, Bristol Meyers Squibb, Genentech, Glaxo Smith Kline, Janssen (previously Centocor), Millennium Pharmaceuticals (now Takeda), Novartis, Pfizer, Procter and Gamble Pharmaceuticals, Shire Pharmaceuticals, and UCB Pharma; payment for lectures from Abbott Laboratories, Bristol Meyers Squibb and Janssen (previously Centocor); and holds the following patents: Sandborn WJ. Use of topical azathioprine to treat inflammatory bowel disorders. United States patent number: 5,691,343. Date of patent: November 25, 1997., Sandborn WJ, Rhodes J. Colonic delivery of nicotine to treat inflammatory bowel disease. South African patent number: 97/1020. Date of patent: January 28, 1998., Sandborn WJ. Use of azathioprine to treat Crohn's disease. United States patent number: 5,733,915. Date of patent: March 31, 1998., Sandborn WJ, Rhodes J. Colonic delivery of nicotine to treat inflammatory bowel disease. United States patent number: 5,846,983. Date of patent: December 8, 1998., Sandborn WJ. Azathioprine compositions for colonic administration. New Zealand patent number: 306062. Date of Patent: February 11, 1999., Sandborn WJ. Azathioprine compositions for colonic administration. Singapore patent number: 45647. Date of Patent: March 14, 1999., Sandborn WJ, Rhodes J, Rhodes P, Evans BK. Colonic delivery of nicotine to treat inflammatory bowel disease. United States patent number: 5,889,028. Date of patent: March 30, 1999., Sandborn WJ. Topical formulations of azathioprine to treat inflammatory bowel disorders. United States patent number: 5,905,081. Date of Patent: May 18, 1999., Sandborn WJ. Azathioprine compositions for colonic administration. Australia patent number: 707168. Date of Patent: October 14, 1999. Sandborn WJ, Rhodes J, Evans BK. Intestinal absorption of nicotine to treat nicotine responsive conditions. Australia patent number: 718052. Date of patent: July 20, 2000., Sandborn WJ, Rhodes J. Colonic delivery of nicotine to treat inflammatory bowel disease. United States patent number: 6,166,044. Date of patent: December 26, 2000., Sandborn WJ. Use of topical azathioprine and thioguanine to treat colorectal adenomas. United States patent number: 6,166,024. Date of patent: December 26, 2000., Rhodes J, Evans BK, Rhodes P, Sandborn WJ. Intestinal absorption of nicotine to treat nicotine responsive conditions. United States patent number: 6,238,689. Date of patent: May 29, 2001., Sandborn, WJ. Azathioprine compositions for colonic administration. Czech Republic patent number: 290428. Date of patent: May 27, 2002., Sandborn, WJ, Rhodes J. Colonic delivery of nicotine to treat IBD. Mexico patent number: 209636. Date of Patent August 12, 2002., Sandborn WJ. Enema and enterically-coated oral dosage forms of azathioprine. United States Patent No.: 6,432,967. Date of patent: August 13, 2002., Sandborn WJ, Rhodes J. Colonic delivery of nicotine to treat nicotine responsive conditions. Europe patent number: 0954337. Date of patent: November 2, 2002., Sandborn WJ, Rhodes J, Rhodes P, Evans BK. Colonic delivery of nicotine to treat IBD. Europe patent number: 893998. Date of patent: April 15, 2003., Sandborn WJ, Rhodes J, Rhodes P, Evans BK. Colonic delivery of nicotine to treat inflammatory bowel disease. Hong Kong patent number: HK1019043. Date of patent: August 1, 2003., Sandborn WJ, Rhodes J, Rhodes P, Evans BK. Colonic delivery of nicotine to treat IBD. China patent number: ZL97192177. Date of patent: November 12, 2003., Sandborn W, Rhodes J, Rhodes P, Evans B. Colonic delivery of nicotine to treat inflammatory bowel disease. Czech patent number: 293616. Patent date: 2004., Rhodes J, Sandborn WJ, Rhodes P, Evans BK. Colonic deliver of nicotine to treat inflammatory bowel disease. Canada patent number: 2,246,235. Patent date: 2007., Sachetto JP, Sandborn WJ, Tremaine WJ. Pharmaceutical composition for the treatment of inflammatory bowel disease. United States patent number: 7341741. Patent date 2008., Rhodes J, Evans BK, Rhodes P, Sandborn WJ. Intestinal absorption of nicotine to treat nicotine responsive conditions. Canadian patent number: 2,260,909. Patent date 2008., and Levy MJ, Camilleri ML, Murray JA, Sandborn WJ. Obesity treatment and device. United States patent number: 7,803,195 B2. Date of patent September 28, 2010.

Geert D'Haens has received consultancy fees from Abbvie, Ablynx, Amakem, AM Pharma, Avaxia, Biogen, Bristol Meiers Squibb, Boerhinger Ingelheim, Celgene, Celltrion, Cosmo, Covidien, Ferring, DrFALK Pharma, Engene, Galapagos, Gilead, Glaxo Smith Kline, Hospira, Immunic, Johnson and Johnson, Lycera, Medimetrics, Millenium/Takeda, Mitsubishi Pharma, Merck Sharp Dome, Mundipharma, Novonordisk, Pfizer, Prometheus laboratories/Nestle, Protagonist, Receptos, Robarts Clinical Trials, Salix, Sandoz, Setpoint, Shire, Teva, Tigenix, Tillotts, Topivert, Versant and Vifor; has held grants/grants pending from Abbvie, Covidien, Ferring, DrFALK Pharma, Millenium/Takeda, Merck Sharp Dome, Mundipharma, Pfizer, and Prometheus laboratories/Nestle; has received payment for lectures from Abbvie, Ferring, Johnson and Johnson, Merck Sharp Dome, Mundipharma, Norgine, Pfizer, Shire, Millenium/Takeda, Tillotts and Vifor; and has stock/stock options with Engene.

Vipul Jairath has received scientific advisory board fees from Abbvie, Sandoz, Takeda and Janssen; lecture fees from Takeda, Janssen and Ferring; and travel support for conference attendance from Vifor pharmaceuticals.

All of the aforementioned financial activities are outside the submitted work.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

During the review process, we identified multiple overlapping stepwise histologic disease activity indices with the same or similar items. We decided to exclude stepwise indices from further analysis. Also, the original protocol did not include information on how estimates of correlation would be interpreted. The Landis and Koch criteria were used for this purpose, as explained in the current methods section. We also slightly modified the description of the COSMIN tool for clarity in the methods section.

INDEX TERMS

Medical Subject Headings (MeSH)

Biopsy; Colon [pathology]; Colonoscopy; Crohn Disease [*pathology]; Histological Techniques [methods] [standards]; Ileum [pathology]; Prospective Studies; Rectum [pathology]; Reproducibility of Results; Retrospective Studies

MeSH check words

Adult; Humans