Cochrane Library

Cochrane Database of Systematic Reviews

Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis (Review)

Mohammed Vashist N, Samaan M, Mosli MH, Parker CE, MacDonald JK, Nelson SA, Zou GY, Feagan BG, Khanna R, Jairath V

Mohammed Vashist N, Samaan M, Mosli MH, Parker CE, MacDonald JK, Nelson SA, Zou GY, Feagan BG, Khanna R, Jairath V.
Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis.
Cochrane Database of Systematic Reviews 2018, Issue 1. Art. No.: CD011450.
DOI: 10.1002/14651858.CD011450.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Methodology Review]

Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis

Nadia Mohammed Vashist[1], Mark Samaan[2], Mahmoud H Mosli[3], Claire E Parker[4], John K MacDonald[5,6], Sigrid A Nelson[7], GY Zou[5,8], Brian G Feagan[5,6,8], Reena Khanna[4,6], Vipul Jairath[4,6,8]

[1]PRA Health Sciences, Victoria, Canada. [2]Academic Medical Center, Amsterdam, Netherlands. [3]King Abdulaziz University Hospital, King Abdulaziz University, Jeddah, Saudi Arabia. [4]Robarts Clinical Trials, London, Canada. [5]Cochrane IBD Group, Robarts Clinical Trials, London, Canada. [6]Department of Medicine, University of Western Ontario, London, Canada. [7]Robarts Clinical Trials, San Diego, CA, USA. [8]Department of Epidemiology and Biostatistics, University of Western Ontario, London, Canada

Contact address: Vipul Jairath, Department of Medicine, University of Western Ontario, London, ON, Canada. vjairath@uwo.ca.

Editorial group: Cochrane IBD Group.
Publication status and date: New, published in Issue 1, 2018.

Citation: Mohammed Vashist N, Samaan M, Mosli MH, Parker CE, MacDonald JK, Nelson SA, Zou GY, Feagan BG, Khanna R, Jairath V. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database of Systematic Reviews* 2018, Issue 1. Art. No.: CD011450. DOI: 10.1002/14651858.CD011450.pub2.

ABSTRACT

Background

Endoscopic assessment of mucosal disease activity is routinely used to determine eligibility and response to therapy in clinical trials of ulcerative colitis. The operating properties of the existing endoscopic scoring indices are unclear.

Objectives

A systematic review was undertaken to evaluate the development and operating characteristics of endoscopic scoring indices for the evaluation of ulcerative colitis.

Search methods

We searched MEDLINE, Embase and CENTRAL from inception to 5 July 2016. We also searched references and conference proceedings (Digestive Disease Week, United European Gastroenterology Week, European Crohn's and Colitis Organization).

Selection criteria

Any study design (e.g. randomized controlled trials, cohort studies, case series) that evaluated endoscopic indices for evaluation of ulcerative colitis disease activity were considered for inclusion. Eligible participants were adult patients (> 16 years), diagnosed with ulcerative colitis using conventional clinical, radiologic and endoscopic criteria.

Data collection and analysis

Two authors independently reviewed the studies identified from the literature search. These authors also independently extracted and recorded data on the number of patients enrolled; number of patients per treatment arm; patient characteristics including age and gender distribution; endoscopic index; and outcomes such as reliability (intra-rater and inter-rater), validity (content, construct, criterion), responsiveness and feasibility. Any disagreements regarding study inclusion or data extraction were resolved by discussion and consensus with a third author. Risk of bias was assessed by determining whether assessors were blinded to clinical information and whether assessors scored the endoscopic index independently. We also assessed the methodological quality of the validation studies using the COSMIN checklist

Main results

A total of 23 reports of 20 studies met the pre-defined inclusion criteria and were included in the review. Of the 20 included validation studies, 19 endoscopic scoring indices were assessed, including the Azzolini Classification, Baron Score, Blackstone Endoscopic Interpretation, Chinese Grading System of Ulcerative Colitis, Endoscopic Activty Index, Jeroen Score, Magnifying Colonoscopy Grade, Matts Score, Mayo Clinic Endoscopic Subscore, Modified Baron Score, Modified Mayo Clinic Endoscopic Subscore, Osada Score, Rachmilewtiz Endoscopic Score, St. Mark's Index, Ulcerative Colitis Colonoscopic Index of Serverity (UCCIS), endoscopic component of the Ulcerative Colitis Disease Activity Index (UCDAI), Ulcerative Colitis Endoscopic Index of Severity (UCEIS), Witts Sigmoidoscopic Score and Watson Grade. The individuals who performed the endoscopic scoring were blinded to clinical and/or histologic information in ten of the included studies, not blinded to clinical and/or histologic information in one of the included studies, and it was unclear whether blinding occurred in the remaining nine included studies. Independent observation was confirmed in four of the included studies, unclear in five of the included studies, and non-applicable (since inter-rater reliability was not assessed) in the remaining eleven included studies. The methodological quality (COSMIN checklist) of most of the included studies was rated as 'good' or 'excellent'. One study that assessed responsiveness was rated as 'fair'. The inter-rater reliability of nine endoscopic scoring indices including the Baron Score, Blackstone Endoscopic Interpretation, Endoscopic Activity Index, Matts Score, Mayo Clinic Endoscopic Subscore, Osada Score, UCCIS, UCEIS, Watson Grade was assessed in seven studies, with estimates of correlation, ƙ, ranging from 0.44 to 0.97. The iIntra-rater reliability of seven endoscopic scoring indices including the Baron Score, Blackstone Endoscopic Interpretation, Matts Score, Mayo Clinic Endoscopic Subscore, Osada Score, UCCIS and UCEIS was assessed in three studies, with estimates of correlation, ƙ, ranging from 0.41 to 0.86. No studies assessed content validity. Three studies evaluated the criterion validity of three endoscopic scoring indices including the Rachmilewitz Endoscopic Score, Magnifying Colonoscopy Grade and the UCCIS. These indices were correlated with objective markers of disease activity including albumin, blood leukocytes, C-reactive protein, fecal calprotectin, hemoglobin, mucosal interleukin-8 concentration and platelet count. Correlation estimates ranged from r = -0.19 to 0.83. Thirteen endoscopic scoring indices were tested for construct validity in 13 studies. Estimates of correlation between the endoscopic scoring indices and other measures of disease activity ranged from r = 0.27 to 0.93. Two studies explored the responsiveness of four endoscopic scoring indices including the Mayo Endoscopic Subscore, Modified Baron Score, Modified Mayo Endoscopic Subscore and UCEIS. One study concluded that the Modified Baron Score, Modified Mayo Endoscopic Subscore and UCEIS had similar responsiveness for detecting disease change in ulcerative colitis. The other included study concluded that the UCEIS may be the most accurate endoscopic scoring tool. None of the included studies formally assessed feasibility.

Authors' conclusions

While the UCEIS, UCCIS and Mayo Clinic Endoscopic Subscore have undergone extensive validation, none of these instruments have been fully validated and only two studies assessed responsiveness. Further research on the operating properties of these indices is needed given the lack of a fully-validated endoscopic scoring instrument for the evaluation of disease activity in ulcerative colitis.

PLAIN LANGUAGE SUMMARY

Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis

What is ulcerative colitis?

Ulcerative colitis is an inflammatory bowel disease characterized by long-term (chronic) inflammation and ulcers (sores) in the inner most lining of the large intestine and the rectum. Common symptoms include diarrhea, abdominal pain and cramping, weight loss and tiredness.

What is an endoscopic scoring index?

An endoscopic scoring index measures disease activity based on what a physician can see during an endoscopy procedure. An endoscopy is a non-surgical procedure whereby a small camera is used to view the digestive tract. The physician who performs the endoscopy may rate disease activity using the index, or this may be done by another physician if the procedure was video recorded or photographs were taken.

Commonly used endoscopic indices include the Baron Score, Rachmilewitz Index, Ulcerative Colitis Endoscopic Index of Severity, Mayo Clinic Endoscopic Subscore, and the Ulcerative Colitis Colonoscopic Index of Serverity.

What did the researchers investigate?

It is important for endoscopic indices to be valid, meaning that they accurately evaluate what they are intended to measure. The researchers investigated the validity of various endoscopic indices for assessing disease activity in ulcerative colitis. While the Ulcerative Colitis Endoscopic Index of Severity, Mayo Clinic Endoscopic Subscore, and the Ulcerative Colitis Colonoscopic Index of Serverity have undergone extensive validation compared to the other indices, none of these instruments have been fully validated,

What did the researchers find?

The researchers found that none of the currently used endoscopic indices have been fully validated. Further research on the operating properties of these indices is needed given the lack of a fully-validated endoscopic scoring instrument for the evaluation of disease activity in ulcerative colitis.

BACKGROUND

Ulcerative colitis (UC) is an idiopathic inflammatory disease that primarily affects the colonic mucosa with a tendency towards involving the distal part of the colon. The disease can present at any age with symptoms of bloody diarrhea and abdominal pain with a relapsing-remitting course. UC mainly affects the superficial layers of the colonic lining which translates into endoscopic findings such as mucosal edema, erythema, granularity, friability, and ulcers. Disease activity may be classified as mild, moderate, severe, or fulminant based on combined clinical and endoscopic assessments. The aim of therapy is to induce and maintain clinical and endoscopic remission to prevent long-term complications such as uncontrolled bleeding, colorectal cancer and colectomy (Abraham 2009; Baumgart 2007). The evaluation of therapy in clinical trials is highly dependent on the use of well-defined endpoints (Hanauer 2004).

Evaluating the distal colon in patients with suspected colitis using a sigmoidoscope was first described by Bargen 1935. Since then, several endoscopic, clinical and composite indices have been developed to evaluate disease activity in clinical trials of medical therapy for UC (Cooney 2007). It is now apparent that a poor correlation exits between clinical symptoms, as assessed by an instrument such as the Truelove and Witts Severity Index, and endoscopic measures (Truelove 1955). The first index used to evaluate endoscopic activity in UC was the Matts Score (Matts 1961). Developed shortly thereafter was the Baron score, which was first used in a clinical trial assessing the efficacy of prednisolone for the treatment of active UC. In this trial, endoscopic evaluation was limited to the use of a rigid sigmoidoscope and patients were scored from zero to three based on degree of inflammation (Baron 1964).

Feagan 2005 modified the Baron Score by assessing patients on a scale from zero to four based on degree of inflammation. This index is known as the Modified Baron Score, or Feagan Score. The Dick Score is a sigmoidoscopic grading system that was initially used in a randomized controlled trial of sulfasalazine for the treatment of UC. The Dick score is relatively subjective as it categorizes patients as worse, unchanged, improved or much improved (Dick 1964). The Powell-Tuck Score, also known as the St. Mark's Index, was also developed using rigid sigmoidoscopy (Powell-Tuck 1982). The Sutherland Index, also known as the Ulcerative Colitis Disease Activity Index (UCDAI), contains an endoscopic sub score and was introduced in a randomised controlled trial of rectal 5-aminosalicylic acid for the treatment of UC (Sutherland 1987). One of the most commonly used endoscopic measures, the Mayo Endoscopic Subscore (a component of the Mayo Clinic Score), is a four-point scoring system in which patients with normal or inactive, mild, moderate or severe disease are given scores of zero, one, two or three, respectively (Schroeder 1987). The Rachmilewitz Score, otherwise known as the Endoscopic Index, was developed for a randomised clinical trial comparing coated mesalazine to sulfasalazine for the treatment of active UC and has been widely used as an outcome in clinical trials (Rachmilewitz 1989). Other less commonly used scores include the Truelove and Witts Sigmoidoscopic Assessment (Truelove 1955), the Lemann Score, also known as the Sigmoidoscopic Inflammation Grade Score (Lemann 1995), and the Sigmoidoscopic Index (Hanauer 1993).

Over the last decade, several widely used endoscopic scores have been developed: the Endoscopic Activity Index (EAI; Naganuma 2010), Ulcerative Colitis Endoscopic Index of Severity (UCEIS; Travis 2009), and Ulcerative Colitis Colonoscopic Index of Severity (UCCIS; Samuel 2013). The EAI is a novel endoscopic scoring system developed to facilitate treatment options for patients with severe UC. The EAI consists of six items (size of ulcers, depth of ulcers, redness, bleeding, mucosal edema, mucosal exudate) that can be given a maximum score of two or three (Naganuma 2010). The UCEIS was developed using a linear mixed regression model. It assesses the extent of endoscopic severity using three variables: vascular pattern (normal (1), patchy obliteration (2) or obliterated (3)); bleeding (none (1), mucosal (2), luminal mild (3), luminal moderate or severe (4)); and erosions and ulcers (none (1), erosions (2), superficial ulcer (3) or deep ulcer (4)) (Travis 2012). The UCCIS is an endoscopic index that assesses endoscopic severity according to four variables: vascular pattern, granularity, friability, and ulceration (Samuel 2013).

Why it is important to do this review

Increasing importance has been placed on the use of endoscopic indices as outcome measures in clinical research as these indices may function as a more objective measure of disease activity compared to symptom-based indices. However, the operating properties of these endoscopic indices need to be clearly defined. In particular, an endoscopic index must be valid (i.e. it must measure the outcome that it is intended to assess), responsive (i.e. it must be capable of detecting a meaningful change in health status); and reliable (i.e. consistent results should be obtained in patients with a stable clinical status). Furthermore, an ideal instrument is feasible for use in clinical trials. This review will evaluate the relative merits of the existing endoscopic scoring indices and identify areas where further research is needed.

OBJECTIVES

The primary objective was to systematically review the current literature describing the development and operating characteristics of endoscopic scoring indices in UC.

METHODS

Criteria for considering studies for this review

Types of studies

Any study design (e.g. randomized controlled trials, cohort studies, case series) evaluating an endoscopic index in UC was considered for inclusion. Study subjects included adult patients (> 16 years) diagnosed with UC using conventional clinical, radiographic, histologic and endoscopic criteria.

Types of data

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

Types of methods

The methods used to construct and validate the endoscopic indices (e.g. reliability, validity, responsiveness and feasibility) were examined in detail and described for each eligible study. We also reported on the number of endoscopists who scored the

endoscopic indices in each study and whether these endoscopists were aware of other raters' scores.

Types of outcome measures

Reliability: Measures of reliability including intra-rater and inter-rater reliability, test-retest reliability, or internal consistency, were evaluated by assessing the reported correlation estimates (interclass correlation coefficients (ICCs), kappa statistics (ƙ), or Pearson's r statistic).

Validity: Studies were reviewed for whether content validity, criterion validity and construct validity was evaluated.

If the components of an index are sufficient to measure disease activity in UC, the study is thought to have content validity. Content validation is generally based on qualitative assessment. For example, evidence of content validity includes expert panel opinion on face validity, or a systematic review of the literature supporting the development of an endoscopic index.

Criterion validity refers to the degree to which the endoscopic index score is an adequate reflection of true UC activity as assessed against gold standard measurements of disease activity. The lack of a single gold standard for UC activity is a limitation of these assessments. In the current study, studies were considered to test criterion validity if they compared the score to objective biomarkers of inflammation (e.g. fecal calprotectin) or sequelae in the future (e.g. surgery or disability). Statistical parameters reporting agreement between the endoscopic index and disease gold standards were recorded (i.e. sensitivity, specificity, receiver operating characteristic (ROC) curve, area under the curve, mean difference, weighed ƙ, Spearman's rank correlation coefficient (ρ), Pearson's correlation coefficient (r) and ICCs).

Construct validation acknowledges the lack of a gold standard for disease activity. Rather than comparing the index to a gold standard, the index is compared to another hypothesis of true disease activity. Studies reporting on the correlation between the endoscopic index and measures of clinical disease activity were evaluated.

Responsiveness: Following a period of known endoscopic change (e.g. after a treatment of known efficacy), the relationship between pre-change and post-change scores was assessed to determine index responsiveness. 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 distinguish improved from unimproved patients (Deyo 1991).

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

The interpretation of correlation estimates for observer agreement in this systematic review was based on the criteria proposed by Landis and Koch. Using this system, a correlation coefficient of < 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). For the interpretation of correlation coefficients in circumstances other than observer agreement, we will use the criteria proposed by Cohen. The effect size indicated by a correlation coefficient of 0.10 was considered 'small', 0.30 was considered 'medium' and 0.50 was considered 'large' (Cohen 1992).

Search methods for identification of studies

Electronic searches

We searched the following databases from inception to 5 July 2016:
1. MEDLINE (1966);
2. Embase (1980); and
3. CENTRAL.

The search strategies are reported in Appendix 1.

Searching other resources

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

1. Digestive Disease Week;
2. United European Gastroenterology Week; and
3. European Crohn's and Colitis Organization.

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 (NV and MM) independently reviewed the titles and abstracts of the studies identified by the literature search. The full text of potentially relevant citations was reviewed for inclusion. Any disagreements regarding scores identified or included studies were resolved by discussion and consensus with a third author (CEP).

Data extraction and management

A standardized form was used to extract information from selected studies. Two authors (NV, MM) independently extracted and recorded data. The following data were recorded from each eligible study:
a) Number of patients enrolled, number of patients per treatment arm;
b) Patient characteristics including age and gender distribution;
c) The endoscopic index; and
d) Outcomes including intra-rater reliability, inter-rater reliability, responsiveness, validity, feasibility, construct validity and criterion validity.

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 information; and
- Independent observation.

We also assessed the methodological quality of the included studies using the COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments) checklist. The checklist consists of ten properties: internal consistency, reliability, measurement error, content validity, structural validity (factor analysis), hypothesis testing, cross-cultural validity, criterion validity, responsiveness to change and interpretability. A four-point scale is used to rate each property (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'). Generalizability was also assessed as part of the COSMIN checklist.

Measures of the effect of the methods

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

Dealing with missing data

In the case of missing data, the original study authors were contacted if possible.

Sensitivity analysis

This was a descriptive systematic review, therefore we did not conduct sensitivity analyses.

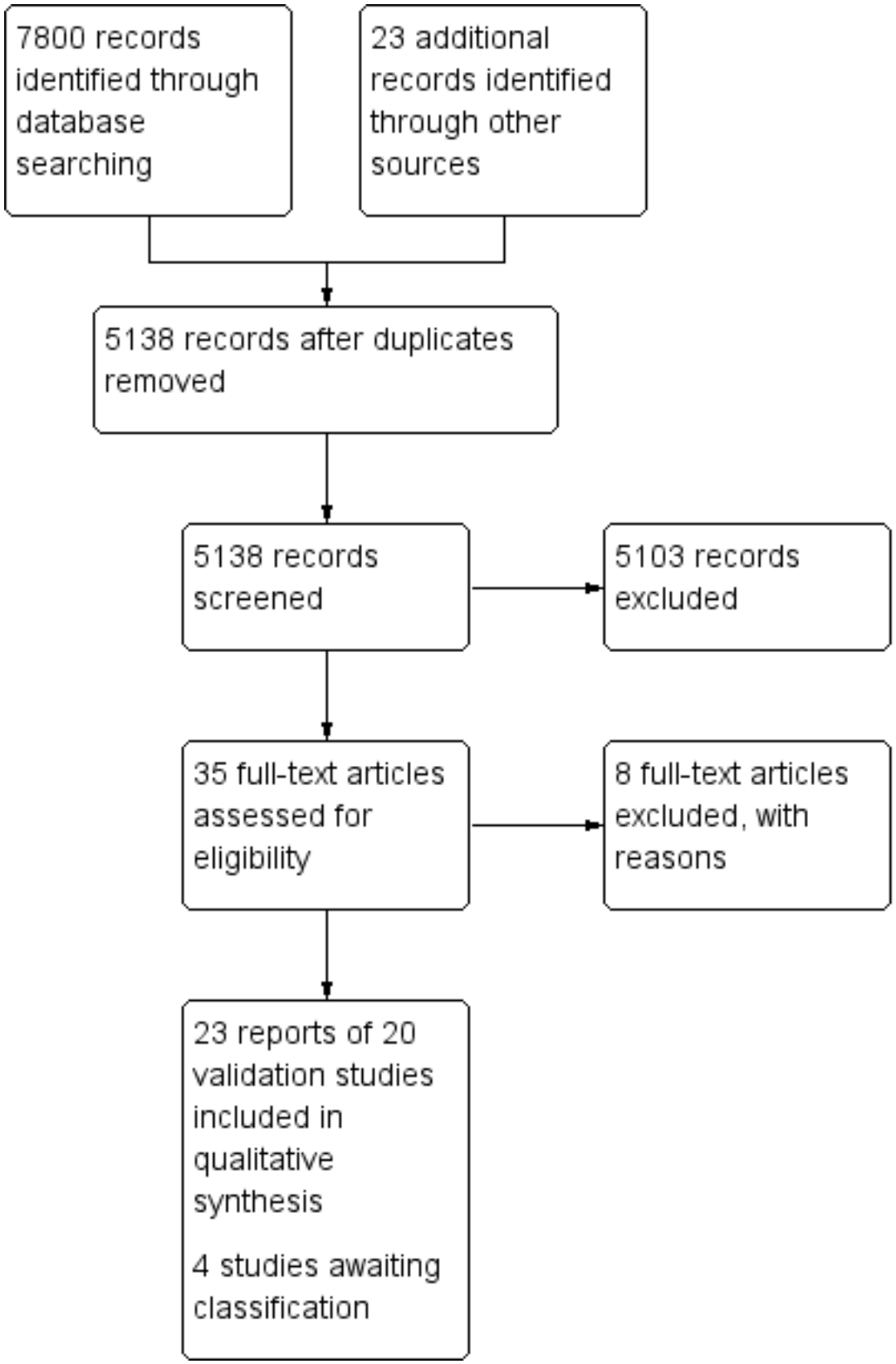
RESULTS

Description of studies

Results of the search

The literature search performed on 5 July 2016 identified 7800 records. An additional 23 records were identified through other sources including reference lists. After duplicates were removed, a total of 5138 records were screened for inclusion. Of these, 35 were selected for full text review. Eight articles were excluded with reasons (see Characteristics of included studies), leaving 23 reports of 20 studies that met pre-defined inclusion criteria (see Figure 1). Four studies are awaiting classification.

Figure 1. Study flow diagram.



Included studies

Twenty studies reported validation results (Burger 2011; Daperno 2011; Daperno 2014; de Lange 2004; Dhanda 2012; Higgins 2005a; Hirai 2010; Ikeya 2016; Jun 2008; Kiesslich 2012; Levesque 2014; Naganuma 2010; Nishio 2006; Osada 2010; Rubin 2012; Samuel 2013; Schoepfer 2009; Thomas 2009; Travis 2013; Walsh 2009).

The 20 included studies evaluated 19 different scoring indices (Table 1). One study evaluated the Azzolini Classification (Jun 2008), six studies evaluated the Baron Score (Burger 2011; Hirai 2010; Jun 2008; Osada 2010; Thomas 2009; Walsh 2009), one study evaluated the Blackstone Endoscopic Interpretation (Osada 2010), one study evaluated the Chinese Grading System of Ulcerative Colitis (CGSUC) (Jun 2008), two studies evaluated the Endoscopic Activity Index (de Lange 2004; Naganuma 2010), one study evaluated the Jeroen Score (Jun 2008), one study evaluated the Magnifying Colonoscopy Grade (Nishio 2006), two studies evaluated the Matts Score (Naganuma 2010; Osada 2010), six studies evaluated the Mayo Clinic Endoscopic Subscore (Daperno 2011; Dhanda 2012; Ikeya 2016; Osada 2010; Rubin 2012; Walsh 2009), three studies evaluated the Modified Baron Score (Jun 2008; Levesque 2014; Walsh 2009), one study evaluated the Modified Mayo Clinic Endoscopic Subscore (Levesque 2014), one study evaluated the Osada Score (also known as the Modified 6-Point Activity Index) (Osada 2010), three studies evaluated the Rachmilewitz Endoscopic Score (Hirai 2010; Naganuma 2010; Schoepfer 2009), one study evaluated the St. Mark's Index (Higgins 2005a), one study evaluated the UCCIS (Samuel 2013), one study evaluated the UCDAI (Higgins 2005a), three studies evaluated the UCEIS (Levesque 2014; Ikeya 2016; Travis 2013), one study evaluated the Truelove and Witts Sigmoidoscopic Score (Jun 2008), and one study evaluated the Watson Grade (Kiesslich 2012).

Excluded studies

Eight studies were excluded after full-text review as these studies did not meet the inclusion criteria (Blonski 2011; Hameed 2001; Kato 2011; Neumann 2012; Ohkusa 2006; Powell-Tuck 1982; Travis 2009; Travis 2011).

Eighteen additional endoscopic scoring indices were identified but not included in the current review as these indices have not undergone any form of validation testing (Table 2).

Risk of bias in included studies

Blinding

Blinding to clinical information such as symptoms, physical examination or laboratory information is important for the objective assessment of endoscopic data (Feagan 2013). However, the presence or absence of blinding was not routinely reported in the included studies.

Raters were blinded to clinical information in ten of the included studies (Daperno 2011; Jun 2008; Kiesslich 2012; Levesque 2014; Nishio 2006; Osada 2010; Samuel 2013; Schoepfer 2009; Travis 2013; Walsh 2009). In one study, the endoscopic raters were not blinded to clinical information (Osada 2010). It was unclear whether the raters were blinded to clinical information in the remaining nine studies (Burger 2011; Daperno 2014; de Lange 2004; Dhanda 2012; Higgins 2005a; Hirai 2010; Ikeya 2016; Rubin 2012; Thomas 2009).

Independent Observation

Eleven of the included studies did not assess inter-rater reliability (Burger 2011; Dhanda 2012; Higgins 2005a; Hirai 2010; Ikeya 2016; Levesque 2014; Naganuma 2010; Nishio 2006; Schoepfer 2009; Thomas 2009; Walsh 2009), therefore observation by independent endoscopic raters was not relevant. Of the remaining eight included studies, independent observation was conducted in four instances (Jun 2008; Osada 2010; Samuel 2013;Travis 2013). It was unclear whether independent observation was performed in the other five studies (Daperno 2011; Daperno 2014; de Lange 2004; Kiesslich 2012; Rubin 2012).

Effect of methods

Reliability

Seven studies assessed endoscopic scoring index reliability, with estimates of inter-rater reliability reported in all seven studies (Daperno 2011; Daperno 2014; de Lange 2004; Kiesslich 2012; Osada 2010; Samuel 2013; Travis 2013), and intra-rater reliability reported in three studies (Osada 2010; Samuel 2013; Travis 2013) (Table 3).

Mayo Clinic Endoscopic Subscore

Estimates of inter-rater reliability for the Mayo Clinic Endoscopic Subscore ranged between κ = 0.45 and κ = 0.75, indicating moderate to substantial agreement (Daperno 2011; Daperno 2014; Osada 2010). In Daperno 2011, 171 gastroenterologists rated five endoscopic videos before and after receiving training specific to the Mayo Clinic Endoscopic Subscore. The κ statistic improved with training, increasing from 0.45 to 0.71. In Daperno 2014, 13 endoscopic videos were evaluated by 14 expert gastroenterologists. A subset of five videos were also evaluated by 30 general gastroenterologists with no experience in endoscopic scoring. Interestingly, the 'non-expert' inter-rater reliability estimate was higher (κ = 0.71) compared to the 'expert' inter-rater reliability estimate (κ = 0.53). In Osada 2010, 279 endoscopic images were shown to four expert and four trainee endoscopists and assessed using five endoscopic scoring indices. For the Mayo Clinic Endoscopic Subscore, the inter-rater reliability estimates were κ = 0.74 for experts and κ = 0.46 for trainees. With respect to intra-rater reliability, Osada 2010 reported reliability estimates of κ = 0.75 for experts and κ = 0.48 for trainees.

EAI

In de Lange 2004, five 30-second endoscopic video clips were scored by an audience of expert (n = 15) and inexperienced endoscopists (n = 21) using the Endoscopic Activity Index. The inter-rater reliability estimate was higher in the expert group (κ = 0.97, 95% CI 0.92 to 1.00) compared to the non-expert group (κ = 0.79, 95% CI 0.71 to 0.49).

Osada Score

The inter-rater and intra-rater reliability of the Osada Score was assessed in Osada 2010. The inter-rater reliability estimates for experts and trainees were κ = 0.65 and κ = 0.54, respectively. The intra-rater reliability estimates for experts and trainees were κ = 0.79 and κ = 0.64, respectively.

Matts Score

The inter-rater and intra-rater reliability of the Matts Score was assessed in Osada 2010. The inter-rater reliability estimates for

experts and trainees were ƙ = 0.76 and ƙ = 0.44, respectively. The intra-rater reliability estimates for experts and trainees were ƙ = 0.78 and 0.41, respectively.

Baron Score

The inter-rater and intra-rater reliability of the Baron Score was assessed in Osada 2010. The inter-rater reliability estimates for experts and trainees were ƙ = 0.61 and ƙ = 0.47, respectively. The intra-rater reliability estimates for experts and trainees were ƙ = 0.62 and ƙ = 0.46, respectively.

Blackstone Score

The inter-rater and intra-rater reliability of the Blackstone Score was assessed in Osada 2010. The inter-rater reliability estimates for experts and trainees were ƙ = 0.57 and ƙ = 0.46, respectively. The intra-rater reliability estimates for experts and trainees were ƙ = 0.73 and ƙ = 0.51, respectively.

UCCIS

To determine the inter-rater reliability of the four variables (granularity, vascular pattern, bleeding/friability and ulcerations) that comprise the UCCIS, Samuel 2013 had eight gastroenterologists score 250 30-second video recordings representing an equal number of colonic segments. Estimates of inter-rater reliability for each colonic segment (measured by ƙ) ranged from moderate (ICC = 0.56) to substantial (ICC = 0.88) (see Table 3).

UCEIS

In Travis 2013, 57 sigmoidoscopic videos were scored by 25 gastroenterologists using the UCEIS (28 videos were scored by each individual). The inter-rater and intra-rater reliability estimates were ICC = 0.50 and ICC = 0.72, respectively. Internal consistency, as measured by Cronbach's alpha, was estimated to be 0.86.

Watson Grade

Kiesslich 2012 conducted a prospective pilot study in which 58 patients with inactive inflammatory bowel disease underwent confocal laser endomicroscopy. A total of 232 endoscopic images (four images per patient) were obtained and graded by two blinded assessors. Inter-rater reliability, quantified using Cohen's ƙ statistic was estimated to be 0.87.

Validity

Content validity

None of the included studies assessed content validity.

Criterion validity

Estimates of correlation between three endoscopic scoring indices (the Rachmilewitz Endoscopic Score, Magnifying Colonoscopy Grade and UCCIS) and objective biomarkers of inflammation (albumin, blood leukocytes, C-reactive protein (CRP), hemoglobin, mucosal interleukin-8 concentration and platelet count) ranged from small to large effect sizes (r = 0.19 to r = 0.83) and were reported in three studies (Nishio 2006; Samuel 2013; Schoepfer 2009) (Table 4).

Albumin

One study explored the relationship between the UCCIS and albumin levels (Samuel 2013). The effect size for the correlation estimate was large with r = -0.55 (P < 0.001).

Blood leukocytes

The correlation between the Rachmilewitz Endoscopic Score and blood leukocytes had a large effect size: r = 0.46 (P < 0.001) (Schoepfer 2009).

CRP

Two studies explored the relationship between the Rachmilewitz Endoscopic Score and CRP. Both Samuel 2013 and Schoepfer 2009 determined the correlation coefficient to have a large effect size (r = 0.56, P < 0.001 and r = 0.50, P < 0.001, respectively).

Hemoglobin

Samuel 2013 investigated the association between the UCCIS and hemoglobin. The correlation coefficient had a medium effect size with r = -0.39 (P < 0.001).

Interleukin-8 concentration

Nishio 2006 explored the relationship between the Magnifying Colonscopy Grade and mucosal interleukin-8 activity. Spearman's rank test was used to estimate correlation. While the investigators reported that a statistically significant association was observed (P = 0.001), no correlation coefficient was reported.

Platelet count

The correlation between the UCCIS and platelet count was small (r = 0.19, P > 0.050) (Samuel 2013).

Construct validity

A total of 13 endoscopic scoring indices were tested for construct validity in 13 studies (Burger 2011; Dhanda 2012; Higgins 2005a; Hirai 2010; Jun 2008; Naganuma 2010; Nishio 2006; Rubin 2012; Samuel 2013; Schoepfer 2009; Thomas 2009; Travis 2013; Walsh 2009) (Table 5). The effect size of the correlation between the endoscopic scoring indices and other measures of disease activity (e.g. clinical and histologic measurement tools) ranged from medium (r = 0.27) to large (r = 0.93).

Azzolini Score

Jun 2008 compared the Azzolini Score to five other endoscopic indices including the Baron Score, CGSUC, Jeroen Score, Modified Baron Score, and the Truelove and Witts Score. The effect size of the correlation estimates was large (ρ = 0.69 to 0.79, P < 0.001).

CGSUC

In Jun 2008, the CGSUC was compared to five other endoscopic indices (Azzolini Score, Baron Score, Jeroen, Modified Baron Score and Truelove and Witts Score). The effect size of the correlation estimates was large (ρ = 0.74 to 0.80, P < 0.001).

Baron Score

The Baron Score is the most studied endoscopic scoring instrument with respect to construct validity. Five studies (Burger 2011; Hirai 2010; Jun 2008; Thomas 2009; Walsh 2009), assessed the correlation between the Baron Score and three clinical indices (the Seo Index (Seo 1992); Simple Clinical Colitis Activity Index (SCCAI; Walmsley 1998); and UCDAI), two histologic indices (the Truelove and Richards Index (Truelove 1956); and the Lichtiger Index; Langholz 1992), and six other endoscopic indices (Azzolini Score, CGSUC, Jeroen Score, Modified Baron Score, Truelove and Witts Score, and the Rachmilewitz Endoscopic Score). The effect size of the correlation estimates ranged from medium ($\hat{k}$ = 0.27) to large ($\hat{k}$ = 0.89) (Table 5).

Endoscopic Activity Index

Naganuma 2010 examined the relationship between the Endoscopic Activity Index and one clinical index (the Lichtiger Index) and two other endoscopic indices (the Matts Score and Rachmilewitz Endoscopic Score). Large correlation estimates of r = 0.77, 0.91 and 0.87 (P < 0.001) were observed.

Jeroen Score

In Jun 2008, the Jeroen Score was compared to five other endoscopic indices (Azzolini Score, Baron Score, CGSUC, Modified Baron Score and the Truelove and Witts Score). The correlation estimates were large (ρ = 0.76 to 0.83, P < 0.001).

Magnifying Colonoscopy Grade

The Magnifying Colonscopy Grade was compared to a histologic measure of disease activity (the Riley Score) in Nishio 2006. Spearman's rank test was used to estimate correlation. While the investigators reported that a statistically significant association was observed (P = 0.001), no correlation coefficient was reported.

Mayo Clinic Endoscopic Subscore

The Mayo Clinic Endoscopic Subscore was compared to two histologic indices (the Riley Score and Rubin Histologic Score) in two studies (Dhanda 2012; Rubin 2012). Correlation estimates with a large effect size were reported (r = 0.55 and r = 0.60 respectively). Rubin 2012 also compared the Mayo Clinic Endoscopic Subscore to another clinical index containing an endoscopic component (the SCCAI) and found a large effect size (r = 0.53, P < 0.001).

Modified Baron Score

Jun 2008 compared the Modified Baron Score to five other endoscopic indices including Azzolini Score, Baron Score, CGSUC, Jeroen Score and the Truelove and Witts Score. The effect size of the correlation estimates was large (ρ = 0.69 to 0.76, P < 0.001).

Rachmilewitz Endoscopic Score

In Hirai 2010 estimates of correlation between the Rachmilewitz Endoscopic Score and the Rachmilewitz Score, UCDAI, Seo Index and Lichtiger Index were calculated. The effect sizes of the correlation estimates ranged from medium (r = 0.28) to large (r = 0.89).

St. Mark's Index

Higgins 2005a explored the relationship between the St. Mark's Index and the UCDAI, the SCCAI and the Seo Index. The St. Mark's Index failed to be significantly associated with any of the indices. Correlation estimates of r = 0.88, 0.91, 0.80 were observed, respectively (P > 0.05).

Truelove and Witts Score

Jun 2008 compared the Truelove and Witts Score to five other endoscopic indices including the Azzolini Score, Baron Score, CGSUC, Jeroen, and the Modified Baron Score. The correlation estimates had a large effect size (ρ = 0.75 to 0.81, P < 0.001).

UCCIS

In Samuel 2013 the UCCIS was examined in relationship to the SCCAI, Rachmilewitz Score and Patient-Defined Remission Score. The effect size of the correlation estimates was large with r = 0.5, 0.43 and 0.67 (P < 0.01), respectively.

UCEIS

Travis 2013 compared the UCEIS to a Visual Analogue Scale (VAS; 0 = completely normal and 100 = worst ever seen). The effect size of the correlation estimate was large (r = 0.93, P < 0.005).

Responsiveness

Two of the included studies assessed responsiveness.

Levesque 2014 evaluated the responsiveness of three endoscopic scoring indices (the Modified Mayo Endoscopic Subscore (Lobatón 2015), Modified Baron Score and UCEIS) after a treatment of known efficacy (mesalamine) was administered to patients with mild-to-moderate ulcerative colitis. Four central readers independently scored 121 endoscopic videos taken from patients who were both clinically changed and unchanged following mesalamine therapy. The effect sizes and Guyatt's responsiveness statistics for the Modified Mayo Endoscopic Subscore, Modified Baron Score and UCEIS were 0.49 (95% CI 0.28 to 0.71), 0.49 (95% CI 0.28 to 0.71) and 0.58 (95% CI 0.36 to 0.81), and 0.32 (95% CI 0.11 to 0.53), 0.33 (95% CI 0.13 to 0.54) and 0.47 (95% CI 0.25 to 0.69), respectively. The area under the ROC curve for the three endoscopic scoring indices was also similar (Modified Mayo Endoscopic Subscore: 0.66 (95% CI 0.55 to 0.78), Modified Baron Score: 0.65 (95% CI 0.54 to 0.77), UCEIS: 0.68 (95% CI 0.58 to 0.79)). The authors concluded that while the UCEIS had a slightly larger effect size, the three endoscopic scoring indices had similar responsiveness (medium effect size) for detecting change in ulcerative colitis disease activity (Table 6).

In Ikeya 2016, the Mayo Clinic Endoscopic Subscore and the UCEIS were used to score colonoscopies performed in ulcerative colitis patients before and after receiving tacrolimus therapy. The mean change in the Mayo Clinic Endoscopic Subscore and the UCEIS was recorded. The mean UCEIS score significantly improved after tacrolimus therapy among patients who achieved remission (6.2 (+/- 0.9) to 3.4 (+/- 2.1), P < 0.001) and response (6.6 +/- 0.5 to 5.4 +/- 0.8, P = 0.005), while there was no significant decrease in the UCEIS among the non-responders (5.3 +/- 1.5 to 5.7 +/- 1.5). For the Mayo Clinic Endoscopic Subscore, no significant decreases were observed in the response or remission groups. The investigators concluded that the UCEIS may be a more accurate scoring index than the Mayo Clinic Endoscopic Subscore (Table 6).

Feasibility

While it has been suggested that the Baron Score, Mayo Clinic Endoscopic Subscore, Modified Baron Score, Rachmilewitz Index and UCEIS are relatively simple to use (Paine 2014), none of the indices included in this review have been formally assessed for feasibility.

Methodological Quality

The COSMIN tool was used to assess the methodological quality of the included studies (see Table 7).

In total, seven studies assessed the reliability of an endoscopic scoring index (Daperno 2011; Daperno 2014; de Lange 2004; Osada 2010; Rubin 2012; Samuel 2013; Travis 2013). With regard to methodological quality, two of these studies were rated as 'excellent' (Osada 2010; Samuel 2013), and five studies were rated as 'good' (Daperno 2011; de Lange 2004; Kiesslich 2012; Rubin 2012; Travis 2013).

Three studies assessed criterion validity (Nishio 2006; Samuel 2013; Schoepfer 2009). Nishio 2006 received a rating of 'good' and two studies received a rating of 'excellent' using the COSMIN tool (Samuel 2013; Schoepfer 2009).

Thirteen studies assessed construct validity (Burger 2011; Dhanda 2012; Higgins 2005a; Hirai 2010; Jun 2008; Naganuma 2010; Nishio 2006; Rubin 2012; Samuel 2013; Schoepfer 2009; Thomas 2009; Travis 2013; Walsh 2009). Four studies were rated as 'excellent' (Dhanda 2012; Naganuma 2010; Samuel 2013; Schoepfer 2009) and nine studies were rated as 'good' with respect to methodological quality (Burger 2011; Higgins 2005a; Hirai 2010; Jun 2008; Nishio 2006; Rubin 2012; Thomas 2009; Travis 2013).

Two studies assessed responsiveness (Ikeya 2016; Levesque 2014). One study was rated as 'fair' (Ikeya 2016) and one study was rated as 'excellent' (Levesque 2014) methodological quality.

DISCUSSION

Summary of main results

In total, 23 reports of 20 studies that validated 19 different endoscopic scoring indices were identified by the literature search (Table 1). Eighteen endoscopic scoring indices that have not undergone any form of validation testing were also identified (Table 2). Correlation estimates for intra-rater reliability for seven of the endoscopic scoring indices ranged from 'moderate' to 'substantial'. Inter-rater reliability was assessed in nine of the partially validated indices, with correlation estimates ranging from 'moderate' to 'almost perfect' (Table 3). Three of the included studies assessed criterion validity by calculating correlation estimates between an endoscopic scoring index (the Magnifying Colonoscopy Grade, Rachmilewitz Endoscopic Score and UCCIS) and various biomarkers of inflammation (i.e. C-reactive protein, albumin, hemoglobin, platelet count, fecal calprotectin, interleukin-8 concentration and blood leukocytes). The effect size of the correlation estimates ranged from small to large (Table 4). Twelve of the included studies explored construct validity by comparing a total of 13 endoscopic scoring indices with other measures of disease activity (clinical, endoscopic and histologic). The effect size of the correlation estimates ranged from small to large (Table 5). Two of the included studies measured the responsiveness of a total of four endoscopic scoring indices (i.e. the Mayo Clinic Endoscopic Subscore, Modified Baron Score, Modified Mayo Clinic Endoscopic Subscore and UCEIS). In Levesque 2014, effect size, Guyatt's responsiveness statistic and area under the ROC ranged from 0.49 to 0.58, 0.32 to 0.47 and 0.66 to 0.68, respectively. In Ikeya 2016, the mean Mayo Clinic Endoscopic Subscore changed from 2.9 to 2.0 after tacrolimus therapy, while the mean UCEIS score changed from 6.2 to 3.4 (Table 6).

Overall completeness and applicability of evidence

Three endoscopic scoring indices, the UCCIS, UCEIS and Mayo Clinic Endoscopic Subscore, have undergone the most validation testing. The UCCIS has been evaluated for reliability (inter-rater), criterion validity and construct validity, while the UCEIS and the Mayo Clinic Endoscopic Subscore have been evaluated for reliability (inter-rater and intra-rater), construct validity and responsiveness. None of the currently available endoscopic scoring indices for ulcerative colitis have been fully validated (Table 8).

Quality of the evidence

The COSMIN tool was used to assess the methodological quality of the included studies (Table 7). The 20 included studies received scores ranging from 'fair' to 'excellent' with respect to the 10 operating properties incorporated into this instrument.

Potential biases in the review process

We performed an extensive search of the literature using electronic databases and handsearching of conference abstracts. However, we did not perform a formal search of the grey literature.

Agreements and disagreements with other studies or reviews

The current systematic review was based on an earlier literature review that identified a total of 31 endoscopic scoring indices (Samaan 2014). In addition to identifying four additional endoscopic scoring indices, the current review provides a more thorough examination of the validation testing that has been performed by reporting on reliability, validation, responsiveness and feasibility testing separately.

Several other literature reviews have also addressed the topic of endoscopic scoring indices for the evaluation of disease activity in ulcerative colitis, including D'Haens 2007, Ket 2015 and Paine 2014. The data presented in these publications are consistent with the results published in the current systematic review.

AUTHORS' CONCLUSIONS

Implication for methodological research

While three indices (the UCEIS, UCCIS and Mayo Clinic Endoscopic Subscore) have undergone extensive validation, none of these instruments are fully validated and only two studies assessed responsiveness. Further research on the operating properties of these indices is needed given the lack of a fully-validated endoscopic scoring instrument for the evaluation of disease activity in ulcerative colitis.

ACKNOWLEDGEMENTS

Partial funding for the Cochrane IBD 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

Burger 2011 *{published data only}*

Burger DC, Thomas SJ, Walsh AJ, Von Herbay A, Buchell OC, Keshav S, et al. Depth of remission may not predict outcome of UC over 2 years. Gut. Elsevier, 2011; Vol. 60, issue Suppl 1:A133.

Daperno 2011 *{published data only}*

Daperno M, Comberlato M, Bossa F, Biancone L, Bonanomi A, Cassinotti A, et al. Interobserver agreement in IBD scores requires expertise and education: Preliminary results from an ongoing IG-IBD study. *Digestive and Liver Disease* 2011;**46**(11):969-73.

Daperno 2014 *{published data only}*

Daperno M, Comberlato M, Bossa F, Bonanomi AG, Lombardi G, Biancone L, et al. Increasing interobserver agreement on IBD endoscopic scoring systems: results from the IGIBD Endo Educational Program. *Gastroenterology* 2014;**146**(Supplement 1):S-234.

de Lange 2004 *{published data only}*

de Lange T, Larsen S, Aabakken L. Inter-observer agreement in the assessment of endoscopic findings in ulcerative colitis. *BMC Gastroenterology* 2004;**4**:9.

Dhanda 2012 *{published data only}*

Dhanda AD, Creed TJ, Greenwood R, Sands BE, Probert CS. Can endoscopy be avoided in the assessment of ulcerative colitis in clinical trials?. *Inflammatory Bowel Diseases* 2012;**18**:2056-62.

Dhanda AD, Greenwood R, Creed TJ, Probert CS. Endoscopy can be avoided in the assessment of ulcerative colitis in clinical trials. Gut. BMJ Publishing Group, 2011; Vol. 60, issue Suppl 1:A140-1.

Higgins 2005a *{published data only}*

Higgins PD, Schwartz M, Mapili J, Zimmermann EM. Is endoscopy necessary for the measurement of disease activity in ulcerative colitis?. *American Journal of Gastroenterology* 2005;**100**:355-61.

Hirai 2010 *{published data only}*

Hirai F, Matsui T, Aoyagi K, Inoue N, Hibi T, Oshitani N, et al. Validity of activity indices in ulcerative colitis: comparison of clinical and endoscopic indices. *Digestive Endoscopy* 2010;**22**:39-44.

Ikeya 2016 *{published data only}*

Ikeya K, Hanai H, Sugimoto K, Osawa S, Kawasaki S, Iida T, et al. The ulcerative colitis endoscopic index of severity more accurately reflects clinical outcomes and long-term prognosis than the mayo endoscopic score. *Journal of Crohn's and Colitis* 2016;**10**(3):286-95.

Jun 2008 *{published data only}*

Jun S, Hua RZ, Lu TJ, Xiang C, Dong XS. Are endoscopic grading and scoring systems in inflammatory bowel disease the same?. *Saudi Medical Journal* 2008;**29**:1432-7.

Kiesslich 2012 *{published data only}*

Kiesslich R, Duckworth CA, Moussata D, Gloeckner A, Lim LG, Goetz M, et al. Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. *Gut* 2012;**61**(8):1146-53.

Levesque 2014 *{published data only}*

Levesque BG, Loftus EV, Panaccione R, McDonald JW, Van Assche G, Zou G. Responsiveness of endoscopic indices in the evaluation of ulcerative colitis. *Gastroenterology* 2014;**146**(5):S226.

Naganuma 2010 *{published data only}*

Naganuma M, Ichikawa H, Inoue N, Kobayashi T, Okamoto S, Hisamatsu T, et al. Novel endoscopic activity index is useful for choosing treatment in severe active ulcerative colitis patients. *Journal of Gastroenterology* 2010;**45**:936-43.

Nishio 2006 *{published data only}*

Nishio Y, Ando T, Maeda O, Ishiguro K, Watanabe O, Ohmiya N, et al. Pit patterns in rectal mucosa assessed by magnifying colonoscope are predictive of relapse in patients with quiescent ulcerative colitis. *Gut* 2006;**55**(12):1768-73.

Osada 2010 *{published data only}*

Osada T, Ohkusa T, Okayasu I, Yoshida T, Hirai S, Beppu K, et al. Correlations among total colonoscopic findings, clinical symptoms, and laboratory markers in ulcerative colitis. *Journal of Gastroenterology and Hepatology* 2008;**23 Suppl 2**:S262-7.

Osada T, Ohkusa T, Yokoyama T, Shibuya T, Sakamoto N, Beppu K, et al. Comparison of several activity indices for the evaluation of endoscopic activity in UC: inter- and intraobserver consistency. *Inflammatory Bowel Diseases* 2010;**16**:192-7.

Rubin 2012 *{published data only}*

Rubin D, Keyashian K, Bunnag A, Dave A, Williams J, Hanauer S, et al. Correlation between clinical, endoscopic, and histologic disease activity in ulcerative colitis. American Journal of Gastroenterology. Nature Publishing Group, 2012; Vol. 107:S694.

Samuel 2013 *{published data only}*

Samuel S, Bruining DH, Loftus EV Jr, Thia KT, Schroeder KW, Tremaine WJ, et al. Validation of the ulcerative colitis colonoscopic index of severity and its correlation with disease activity measures. *Clinical Gastroenterology Hepatology* 2013;**11**:49-54.

Schoepfer 2009 *{published data only}*

Schoepfer AM, Beglinger C, Straumann A, Trummler M, Renzulli P, Seibold F. Ulcerative colitis: correlation of the Rachmilewitz endoscopic activity index with fecal calprotectin, clinical activity, C-reactive protein, and blood leukocytes. *Inflammatory Bowel Diseases* 2009;**15**:1851-8.

Thomas 2009 *{published data only}*

Thomas SJ, Walsh A, Von Herbay A, Burchell O, Brain O, Keshav S, et al. How much agreement is there between histological, endoscopic and clinical assessments of remission

in ulcerative colitis?. Gut. BMJ Publishing Group, 2009; Vol. 58:A101-2.

Travis 2013 *{published data only}*

Travis SP, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, et al. Reliability and Initial Validation of the Ulcerative Colitis Endoscopic Index of Severity. *Gastroenterology* 2013;**145**(5):987-95.

Walsh 2009 *{published data only}*

Walsh AJ, Brain AOS, Keshav S, Buchel OC, Jacobovits S, Merrin B, et al. How variable is the Mayo score between observers and might this affect trial recruitment or outcome?. Journal of Gastroenterology and Hepatology. Blackwell Publishing, 2009; Vol. 24:A230.

Walsh AJ, Brain AOS, Keshav S, Buchel OC, Jacobovits S, Merrin B, et al. Which activity index for ulcerative colitis? Evaluation of inter-observer variation in clinical, endoscopic and composite indices. Gut. BMJ Publishing Group, 2009; Vol. 58, issue Suppl 1:A15.

References to studies excluded from this review

Blonski 2011 *{published data only}*

Blonski W, Osterman MT, Lin MV, Brensinger CM, Sonu I, Lichtenstein G. An update: which endoscopic or clinical factor is most predictive of future disease course in patients with ulcerative colitis?. Gastroenterology. W.B. Saunders, 2011; Vol. 140, issue Suppl 1:S358-9.

Hameed 2001 *{published data only}*

Hameed K, Khan IU, Farooqui JI, Shah S. Correlation of endoscopic extent and severity with the clinical presentation of ulcerative colitis. Journal of the College of Physicians and Surgeons Pakistan. Pakistan: College of Physicians and Surgeons Pakistan, 2001; Vol. 11, issue 9:551-4.

Kato 2011 *{published data only}*

Kato J, Kuriyama M, Hiraoka S, Yamamoto K. Is sigmoidoscopy sufficient for evaluating inflammatory status of ulcerative colitis patients?. *Journal of Gastroenterology and Hepatology* 2011;**26**:683-7.

Neumann 2012 *{published data only}*

Neumann H, Neurath MF. Ulcerative colitis: UCCIS - A reproducible tool to assess mucosal healing. *Nature Reviews Gastroenterology and Hepatology* 2012;**9**:692-4.

Ohkusa 2006 *{published data only}*

Ohkusa T, Osada T, Terai T, Sato N, Okayasu I. Current opinion of activity index for colonoscopic and histological findings in ulcerative colitis: A proposal of new activity assessment by cumulative method of seven points scores. Gastroenterological Endoscopy. Japan: Japan Gastroenterological Endoscopy Society, 2006; Vol. 48, issue 4:977-86.

Powell-Tuck 1982 *{published data only}*

Powell-Tuck J, Day DW, Buckell NA, Wadsworth J, Lennard-Jones JE. Correlations between defined sigmoidoscopic appearances and other measures of disease activity in ulcerative colitis. *Digestive Diseases and Sciences* 1982;**27**(6):533-7.

Travis 2009 *{published data only}*

Travis S, Sandborn WJ, Hanauer SB, Lemann M, Sands BE, Marteau P, et al. Identification of items to be included in an ulcerative colitis endoscopic index of severity (UCEIS). Gastroenterology. W.B. Saunders, 2009; Vol. 136, issue 5 Suppl 1:A160.

Travis 2011 *{published data only}*

Travis SP, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2011;**61**:535-42.

References to studies awaiting assessment

Iacucci 2017 *{published data only}*

Iacucci M, Daperno M, Lazarev M, Arsenascu R, Tontini GE, Akinola O, et al. Development and reliability of the new endoscopic virtual chromoendoscopy score: the PICaSSO score. *Gastrointestinal Endoscopy* 2017;**86**(6):1118-27.

Kim 2016 *{published data only}*

Kim DB, Lee KM, Lee JM, Chung YY, Sung HJ, Paik CN, et al. Correlation between histological activity and endoscopic, clinical, and serologic activities in patients with ulcerative colitis. *Gastroenterology Research and Practice* 2016;**2016**:5832051.

Lee 2016 *{published data only}*

Lee YJ, Kim ES, Cho KB, Han S, Kim SK, Lee HS, et al. Validation of ulcerative colitis endoscopic index of severity (UCEIS) in Korea. *Gastrointestinal Endoscopy* 2016;**10**(Supp 1):S157.

Songur 2009 *{published data only}*

Songur Y, Ensari A, Savas B, Senol A, Percinel S. Quantitative endoscopic and histologic activity assessment of ulcerative colitis. *Acta Gastro-enterologica Belgica* 2009;**72**:225-9.

Additional references

Abraham 2009

Abraham C, Cho JH. Inflammatory bowel disease. *New England Journal of Medicine* 2009;**361**(21):2066-78.

Azzolini 2005

Azzolini F, Pagnini C, Camellini L, Scarcelli A, Merighi A, Primerano AM, et al. Proposal of a new clinical index predictive of endoscopic severity in ulcerative colitis. *Digestive Disease and Sciences* 2005;**50**(2):246-51.

Bargen 1935

Bargen JA. The management of colitis. New York: National Medical Book Company Inc, 1935.

Baron 1964

Baron JH, Connell AM, Lennard-Jones JE. Variation between observers in describing mucosal appearances in proctocolitis. *British Medical Journal* 1964;**1**(5375):89-92.

Baumgart 2007

Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007;**369**(9573):1641-57.

Beattie 1996

Beattie RM, Nicholls SW, Domizio P, Williams CB, Walker-Smith JA. Endoscopic assessment of the colonic response to corticosteroids in children with ulcerative colitis. *Journal of Pediatric Gastroenterology and Nutrition* 1996;**22**(4):373-9.

Binder 1970

Binder V. A comparison between clinical state, macroscopic and microscopic appearances of rectal mucosa, and cytologic picture of mucosal exudate in ulcerative colitis. *Scandanavian Journal of Gastroenterology* 1970;**5**(7):627-32.

Blackstone 1984

Blackstone MO. Inflammatory bowel disease. In: Blackstone MO editor(s). Endoscopic interpretation: normal and pathologic appearances of the gastrointestinal tract. New York: Raven Press, 1984.

Carbonnel 1994

Carbonnel F, Lavergne A, Lemann M, Bitoun A, Valleur P, Hautefeuille P, et al. Colonoscopy of acute colitis. A safe and reliable tool for assessment of severity. *Digestive Diseases and Sciences* 1994;**39**(7):1550-7.

Cohen 1992

Cohen J. A power primer. *Psychological Bulletin* 1992;**112**(1):155-9.

Cooney 2007

Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SP. Outcome measurement in clinical trials for Ulcerative Colitis:towards standardisation. *Trials* 2007;**8**:17.

D'Haens 2007

D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology* 2007;**132**(2):763-86.

Danielsson 1987

Danielsson A, Hellers G, Lyrenas E, Lofberg R, Nilsson A, Olsson O, et al. A controlled randomized trial of budesonide versus prednisolone retention enemas in active distal ulcerative colitis. *Scandanavian Journal of Gastroenterology* 1987;**22**(8):987-92.

Darr 2017

Darr U, Khan N. Treat to target in inflammatory bowel disease: an updated review of the literature. *Current Treatment Options in Gastroenterology* 2017;**15**(1):116-25.

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 Suppl):142S-58.

Dick 1964

Dick AP, Grayson MJ, Carpenter RG, Petrie A. Controlled trial of sulphasalazine in the treatment of ulcerative colitis. *Gut* 1964;**5**:437-42.

Feagan 2005

Feagan BG, Greenberg GR, Wild G, Fedorak RN, Pare P, McDonald JW, et al. Treatment of ulcerative colitis with a humanized antibody to the alpha4beta7 integrin. *New England Journal of Medicine* 2005;**352**(24):2499-507.

Feagan 2013

Feagan BG, Sandborn WJ, D'Haens G, Pola S, McDonald JW, Rutgeerts P, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;**145**(1):149-57.

Friedmann 1986

Friedman LS, Richter JM, Kirkham SE, DeMonaco HJ, May RJ. 5-Aminosalicylic acid enemas in refractory distal ulcerative colitis: a randomized, controlled trial. *American Journal of Gastroenterology* 1986;**81**(6):412-8.

Froslie 2007

Frøslie KF, Jahnsen J, Moum BA, Vatn MH. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. *Gastroenterology* 2007;**133**(2):412-22.

Hanauer 1993

Hanauer S, Schwartz J, Robinson M, Roufail W, Arora S, Cello J, et al. Mesalamine capsules for treatment of active ulcerative colitis: results of a controlled trial. Pentasa Study Group. *American Journal of Gastroenterology* 1993;**88**(8):1188-97.

Hanauer 2004

Hanauer SB. Medical therapy for ulcerative colitis 2004. *Gastroenterology* 2004;**126**(6):1582-92.

Higgins 2005b

Higgins PD, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous endpoints for remission and clinical improvement in ulcerative colitis. *Gut* 2005;**54**(6):782-8.

Jeroen 2002

Jeroen D, van Bergeijk JD, Wilson JH, Nielsen OH, vonTirpitz C, Karvonen AL, et al. Octreotide in patients with active ulcerative colitis treated with high dose corticosteroids (OPUS I). *European Journal of Gastroenterology and Hepatology* 2002;**14**(3):243-8.

Ket 2015

Ket SN, Palmer R, Travis S. Endoscopic disease activity in inflammatory bowel disease. *Current Gastroenterology Reports* 2015;**17**(12):50.

Landis 1977

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

Langholz 1992

Langholz E, Munkholm P, Davidsen M, Binder V. Colorectal cancer risk and mortality in patients with ulcerative colitis. *Gastroenterology* 1992;**103**(5):1444–51.

Lemann 1995

Lemann M, Galian A, Rutgeerts P, Van Heuverzwijn R, Cortot A, Viteau JM, et al. Comparison of budesonide and 5-aminosalicylic acid enemas in active distal ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 1995;**9**(5):557-62.

Levine 2002

Levine DS, Riff DS, Pruitt R, Wruble L, Koval G, Sales D, et al. A randomised, double blind, dose response comparison of balsalazide (6.75 g), balsalazide (2.25 g), and mesalamine (2.4 g) in the treatment of active, mild-to-moderate ulcerative colitis. *American Journal of Gastroenterology* 2002;**97**(6):1398-407.

Lindgren 2002

Lindgren S, Löfberg R, Bergholm L, Hellblom M, Carling L, Ung KA, et al. Effect of budesonide enema on remission and relapse rate in distal ulcerative colitis and proctitis. *Scandanavian Journal of Gastroenterology* 2002;**37**(6):705-10.

Lobatón 2015

Lobatón T, Bessissow T, De Hertogh G, Lemmens B, Maedler C, Van Assche G, et al. The modified Mayo endoscopic score (MMES): a new index for the assessment of extension and severity of endoscopic activity in ulcerative colitis patients. *Journal of Crohn's and Colitis* 2015;**9**(10):846-52.

Löfberg 1994

Löfberg R, Ostergaard Thomsen O, Langholz E, Schiöler R, Danielsson A, Suhr O, et al. Budesonide versus prednisolone retention enemas in active distal ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 1994;**8**(6):623-9.

Maier 1988

Maier K, von Gaisberg U, Kraus B. Ulcerative colitis. Activity index for the clinical and histological classification of inflammatory activity. *Schweizerische Medizinische Wochenschrift* 1988;**118**(20):763-6.

Matts 1961

Matts SG. The value of rectal biopsy in the diagnosis of ulcerative colitis. *Quarterly Journal of Medicine* 1961;**30**:393–407.

McPhee 1987

McPhee MS, Swan JT, Biddle WL, Greenberger NJ. Proctocolitis unresponsive to conventional therapy. Response to 5-aminosalicylic acid enemas. *Digestive Diseases and Sciences* 1987;**32**(12 Suppl):76S-81S.

Paine 2014

Paine ER. Colonoscopic evaluation in ulcerative colitis. *Gastroentoerology Report* 2014;**2**(3):161-8.

Rachmilewitz 1989

Rachmilewitz D. Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. *BMJ* 1989;**298**(6666):82-6.

Rutter 2004

Rutter M, Saunders B, Wilkinson K, Rumbles S, Schofield G, Kamm M, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology* 2004;**126**(2):451-9.

Samaan 2014

Samaan MA, Mosli MH, Sandborn WJ, Feagan BG, D'Haens GR, Dubcenco E, et al. A systematic review of the measurement of endoscopic healing in ulcerative colitis clinical trials: recommendations and implications for future research. *Inflammatory Bowel Diseases* 2014;**20**(8):1465-71.

Saverymuttu 1986

Saverymuttu SH, Camilleri M, Rees H, Lavender JP, Hodgson HJ, Chadwich VS. 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**(5 Part 1):1121-8.

Schroeder 1987

Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *New England Journal of Medicine* 1987;**317**(26):1625-9.

Seo 1992

Seo M, Okada M, Yao T, Ueki M, Arima S, Okumura M. An index of disease activity in patients with ulcerative colitis. *American Journal of Gastroenterology* 1992;**87**(8):971-6.

Sutherland 1987

Sutherland LR, Martin F, Greer S, Robinson M, Greenberger N, Saibil F, et al. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. *Gastroenterology* 1987;**92**(6):1894-8.

Travis 2012

Travis SP, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2012;**61**(4):535-42.

Truelove 1955

Truelove SC, Witts LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *British Medical Journal* 1955;**2**(4947):1041-8.

Truelove 1956

Truelove SC, Richards WC. Biopsy studies in ulcerative colitis. *British Medical Journal* 1956;**1**(4979):1315–8.

van der Heide 1987

van der Heide H, Mulder C, Wiltnik E. Comparison of enemas containing beclomethasone-di-propionate (BDP) or

prednisolone 21-phosphate (PF) in the treatment of distal ulcerative colitis. *Gastroenterology* 1987;**92**(5 Part 2):A1679.

Walmsley 1998

Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998;**43**(1):29-32.

Zou 2005

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

CHARACTERISTICS OF STUDIES

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

Burger 2011

Methods	Consectuive patients were assessed by 4 gastroenterologists using clinical and endoscopic scoring indices Histologic activity was scored by 2 pathologists Fleiss' ƙ was used to evaluate interobserver variation
Data	Number of patients: 91 Number of readers: 4/2
Comparisons	SCCAI (clinical) Truelove and Richards Index (histologic)
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring index validated: the Baron Score Study published in abstract form only; methods indicates interobserver variation study, but only construct validity is reported

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Yes	4 gastroenterologists scored sigmoidoscopy videos of consecutive patients independently (although the rates of interrater agreement were not reported)

Daperno 2011

Methods	171 gastroenterologists were shown 5 video clips of an endoscopy procedure from a patient with UC All participants rated the video using an iPad system after extensive discussion of scoring modalities
Data	Agreement differed significantly ($P < 0.001$) after scoring training was conducted for 3/5 video clips
Comparisons	Interrater reliability was measured before training and after training
Outcomes	Interrater reliability (see Table 3)
Notes	Endoscopic scoring index validated: Mayo Clinic Endoscopic Subscore

Daperno 2011 *(Continued)*

	Study published in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Each video was blindly reviewed
Independent Observation?	Unclear	Not adequately described

Daperno 2014

Methods	14 expert gastroenterologists reviewed 13 UC videos (in addition to 10 postoperative and 8 luminal Crohn's disease videos) A subset of 5 of the endoscopic clips were also reviewed by 30 general gastroenterologists without experience performing endoscopic scoring
Data	Expert gastroenterologists: belonged to tertiary referral centres, had previous experience using IBD scores, median duration of practice was 21 years, median number of patients followed was 1750 Non-expert gastroenterologists: belonged to primary/secondary referral centres, had basic experience in endoscopy but no formal training in scoring instruments (they were briefly introduced to the indices before being asked to score videos)
Comparisons	Interrater reliability for expert gastroenterologists and non-expert gastroenterologists
Outcomes	Interrater reliability (see Table 3)
Notes	Endoscopic scoring index evaluated: The Mayo Clinic Endoscopic Subscore

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Unclear	After every round of video scoring, the raters were permitted to discuss, but not change their scores

de Lange 2004

Methods	30 second video clips (N = 5) of ulcerative colitis were shown to an audience of experienced (n = 15) and inexperienced (n = 21) endoscopists on a high resolution video projector Both groups were asked to assess eight endoscopic features and the overall mucosal inflammation on the Visual Analogue Scale
Data	The 15 experienced gastroenterologists had performed > 750 endoscopies The 21 inexperienced gastroenterologists had performed < 200 endoscopies
Comparisons	Inter-observer reliability

de Lange 2004 *(Continued)*

Outcomes	See Table 3
Notes	Endoscopic Scoring Index evaluated: EAI

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Unclear	The ratings were performed in the same room based on a projection. It is unclear whether this may have affected scoring

Dhanda 2012

Methods	Post-hoc analysis of data from a multicenter randomised controlled trial in steroid-refractory moderate to severe UC (NCT00430898) (N = 149)
Data	Clinical and endoscopic assessment of disease activity was performed at baseline, week 4, week 8 Histologic assessed of disease activity was performed as an optional sub study Biopsies were scored by a single blinded pathologist
Comparisons	Riley Score (histopathology)
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring index evaluated: Mayo Clinic Endoscopic Subscore Correlation was measured using Spearman's rho Correlation estimate for endoscopic and histologic measures only reported at week 4

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	It is unclear whether the endoscopist was blinded to clinical information
Independent Observation?	Unclear	Not relevant (construct validity)

Higgins 2005a

Methods	74 consecutive patients requiring endoscopy were prospectively identified by searching an endoscopy schedule (4 patients did not participate)
Data	Prior to endoscopy, UCDAI scores were calculated After each endoscopy, the endoscopist (15 total) were asked to perform scoring using the St. Mark's Index and UCDAI
Comparisons	UCDAI (clinical)

Higgins 2005a *(Continued)*

	SCCAI (clinical)
	Seo Index (clinical symptoms, hemoglobin, albumin, erthrocyte sedimentation rate)
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring index evaluated: St. Mark's Index
	Correlation was measured using Spearman's ρ and Pearson's r

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Primary gastroenterologists or endoscopists scored disease activity prior to endoscopy; it is unclear whether the endoscopists were blinded to clinical information when endoscopic assessments were performed
Independent Observation?	Unclear	Not relevant (construct validity)

Hirai 2010

Methods	74 patients with moderate to severe UC from 8 institutes
Data	Patients received medical therapy and were evaluated clinically and endoscopically at weeks 2, 4, 8 and post-treatment
Comparisons	Rachmilewitz Score (clinical)
	UCDAI (clinical)
	Lichtiger Index (clinical)
	Seo Index (clinical)
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring index evaluated: Baron Score, Rachmilewtiz Endoscopic Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Unclear	Not relevant (construct validity)

Ikeya 2016

Methods	A responsiveness study based on a treatment of known efficacy
Data	40 patients had colonoscopies performed pre- and post- treatment
Comparisons	Treatment of known efficacy (tacrolimus)

Ikeya 2016 *(Continued)*

Outcomes	Responsiveness (see Table 6)
Notes	Endoscopic scoring index evaluated: Mayo Clinic Endoscopic Subscore, UCEIS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Unclear	Not relevant (responsiveness)

Jun 2008

Methods	Two experienced endoscopists scored Baron Scale and Jeroen Classification independently. The correlation and difference between the two indices were assessed using Kendall's coefficient of concordance and Spearman correlations.
Data	Patient characteristics: 80 UC patients Mean age: 41.14 years
Comparisons	6 endoscopic scoring indices were compared
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring indices evaluated: CGSUC, Truelove and Witts Sigmoidoscopic Score, Baron Score, Modified Baron Score, Jeroen Score, Azzolini Score Both patients with UC and CD were included in this study (80 UC patients, 31 CD patients)

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Two endoscopists were blinded to clinical and histologic findings
Independent Observation?	Yes	Two endoscopists evaluated endoscopic findings independently

Kiesslich 2012

Methods	A prospective pilot study
Data	58 patients with UC or Crohn's disease in clinical remission
Comparisons	232 Endoscopic images (4 per patient) graded determined using confocal endomicroscopy by two blinded raters
Outcomes	Inter-rater reliability (see Table 3)
Notes	Endoscopic scoring index evaluated: Watson Grade

Kiesslich 2012 *(Continued)*

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Observers were blinded
Independent Observation?	Unclear	Not adequately described

Levesque 2014

Methods	A prospective validation study based on previously collected RCT data
Data	Four central readers evaluated endoscopic videos captured during a placebo-controlled trial (Feagan 2013)
Comparisons	Treatment of known efficacy (mesalamine)
Outcomes	Responsiveness (see Table 6)
Notes	Reported in abstract form only Endoscopic Scoring indices evaluated: Modified Mayo Clinic Endoscopic Subscore, Modified Baron Score, UCEIS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Central reading was employed
Independent Observation?	Unclear	Not relevant (responsiveness)

Naganuma 2010

Methods	A novel endoscopic scoring index was developed, the Endoscopic Activity Index (EAI) Inpatients and outpatients from a gastroenterology clinic between 13-71 years with active, moderate to severe UC were eligible to participate
Data	396 patients with UC (454 colonoscopies) The endoscopic score was calculated by a single endoscopist
Comparisons	EAI (endoscopic) Matts Score (endoscopic) Rachmilewitz Endoscopic Score (endoscopic) Lichtiger Index (clinical)
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic scoring indices assessed: EAI, Matts Score, Rachmilewitz Endocopic Score

Naganuma 2010 *(Continued)*

Risk of bias

Item	Authors' judgement	Description
Blinding?	No	Clinical symptoms and endoscopic videos were assessed
Independent Observation?	Unclear	Not relevant (construct validity)

Nishio 2006

Methods	A novel grading system was developed for use when high-resolution video-magnifying colonoscopy is performed
Data	113 patients with UC
Comparisons	Riley Score (histologic) Mucosal interluekin-8 activity (inflammatory cytokine activity measured as picograms per microgram)
Outcomes	Criterion validity, construct validity (see Table 4 and Table 5)
Notes	Endoscopic scoring index evaluated: Magnifying Colonoscopy Grade Spearman's rank correlation coefficient value not given (only P value)

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Pathologist was blinded to clinical data
Independent Observation?	Unclear	Not relevant (construct validity)

Osada 2010

Methods	An inter- and intra-observer agreement study that assessed 4 established endoscopic scoring indices and one novel index
Data	279 endoscopic images of inflamed lesions from 93 UC patients Endoscopic images were displayed twice to 4 expert and 4 trainee endoscopists over an 1 month interval
Comparisons	5 endoscopic scoring indices were assessed
Outcomes	Inter-rater and intra-rater reliability (see Table 3)
Notes	Endoscopic scoring indices evaluated: the Matts Score, Mayo Endoscopic Subscore, Baron Score and Blackstone Score were compared to a new Modified 6-point Activity Index (Osada Score)

Risk of bias

Item	Authors' judgement	Description

Osada 2010 *(Continued)*

Blinding?	Yes	4 expert and 4 trainee endoscopists assessed endoscopic pictures
Independent Observation?	Yes	The images were displayed to the endoscopists independently

Rubin 2012

Methods	A prospective study of UC patients measuring the correlation between endoscopic, clinical and histologic measurement tools
Data	86 UC patients undergoing standard colonoscopy or sigmoidoscopy Static endoscopic images and corresponding biopsies of the mucosa of the distal colon were obtained
Comparisons	SCCAI (clinical) Rubin Histologic Score
Outcomes	Construct validity (see Table 5)
Notes	Endoscopic index evaluated: Mayo Clinic Endoscopic Subscore Study published in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Unlcear whether the endoscopist and histologist who performed scoring were blinded to patient information
Independent Observation?	Unclear	Not adequately described

Samuel 2013

Methods	Prospective validation study of the UCCIS 50 patients with a spectrum of UC disease activity underwent a video recorded colonoscopy
Data	250 video clips (30 seconds in length) representative of an equal number of colonic segments were graded by 8 investigators (2000 evaluations of 50 patients)
Comparisons	Rachmilewitz Score (clinical) SCCAI (clinical) Patient-Defined Remission (clinical) (Higgins 2005b) C-reactive protein albumin hemoglobin platelet count

Samuel 2013 *(Continued)*

Outcomes	Criterion validity, construct validity (see Table 4 and Table 5)
Notes	Endoscopic scoring index evaluated: UCCIS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	8 gastroenterologists blindly rated mucosal lesions
Independent Observation?	Yes	Gastroenterologists independently assessed mucosal lesions

Schoepfer 2009

Methods	115 UC patients requiring colonoscopy were prospectively enrolled The clinical and endoscopic portions of the Rachmilewitz Endoscopic Score were assessed Fecal and blood samples were obtained after colonoscopy 4 trained gastroenterologists graded the endoscopic findings
Data	19 patients underwent 2 colonoscopies, therefore there were 134 colonoscopies performed
Comparisons	Rachmilewitz Score (clinical) Fecal calprotectin C-reactive protein Blood leukocytes
Outcomes	Criterion validity, construct validity (see Table 4 and Table 5)
Notes	Endoscopic scoring index assessed: Rachmilewitz Endoscopic Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	All gastroenterologists performing the colonoscopies were unaware of clinical and biomarker data to avoid bias The clinical score was performed by a different physician than the one that performed the colonoscopy
Independent Observation?	Unclear	Not adequately described (not necessary for construct and criterion validation)

Thomas 2009

Methods	Consecutive UC patients were evaluated using clinical, endoscopic and histological indices in an effort to validate each index Endoscopic activity was assessed independently by 4 specialist gastroenterologists

Thomas 2009 *(Continued)*

	Histological activity was scored by 2 specialist pathologists
Data	91 patients with mild, moderate or severe UC
Comparisons	SCCAI (clinical) Truelove and Richards Score (histologic)
Outcomes	Construct validity (Table 5)
Notes	Endoscopic scoring index evaluated: Baron Score Study published in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Not adequately described
Independent Observation?	Yes	Endoscopic activity was assessed independently by 4 specialist gastroenterologists

Travis 2013

Methods	Videos were retrospectively obtained from a library of videos from clinical trials of patients with active UC
Data	57 sigmoidoscopic videos, stratified based on disease severity, were assessed by 25 investigators The investigators read 28 videos each (4 of which were duplicates, so that intra-rater reliability could be assessed)
Comparisons	Visual Analogue Scale
Outcomes	Inter-rater reliability, intra-rater reliability, construct validity (see Table 3 and Table 5)
Notes	Endoscopic scoring index evaluated: UCEIS

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Investigators were assigned videos randomly and were blinded to clinical details of patients
Independent Observation?	Yes	Investigators assessed videos independently

Walsh 2009

Methods	Purpose was to determine the impact of inter-rater reliability on inclusion criteria and outcomes in clinical trials
Data	100 patients with UC were seen independently, on the same day, by 4 gastroenterologists

Walsh 2009 *(Continued)*	
	Clinical assessments of disease activity were performed on the same day as sigmoidoscopy
Comparisons	3 endoscopic scoring indices were evaluated
Outcomes	Inter-rater reliability (see Table 3)
Notes	Endoscopic scoring index evaluated: Baron Score, Modified Baron Score, Mayo Endoscopic Subscore Study reported in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	The clinician and endoscopist were blinded
Independent Observation?	Unclear	Not adequately described

SCCAI: Simple Clinical Colitis Activity Index
UC: ulcerative colitis
EAI: Endoscopic Activity Index
UCDAI: Ulcerative Colitis Disease Activity Index
UCEIS: Ulcerative Colitis Endoscopic Index of Severity
CD: Crohn's disease
UCCIS: Ulcerative Colitis Colonoscopic Index of Severity

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

Study	Reason for exclusion
Blonski 2011	This study sought to identify factors predictive of endoscopic and clinical disease course No validation was performed
Hameed 2001	This study evaluated whether clinical presentation correlates with endoscopic findings The Baron Score was used to assess endoscopic disease activity It is unclear whether a scoring instrument was used to assess clinical disease activity Study published in abstract form only
Kato 2011	This retrospective analysis aimed to determine whether there is discrepancy between sigmoidoscopy and colonoscopy in the examination of patients with UC using the Mayo score No validation was performed
Neumann 2012	This is a review article that discusses findings from Samuel 2013
Ohkusa 2006	This study does not report on endoscopic scoring index validation testing results
Powell-Tuck 1982	No estimates of correlation reported
Travis 2009	This study describes the development of the UCEIS There was no validation of the UCEIS performed
Travis 2011	This study describes the development of the UCEIS

Study	Reason for exclusion
	There was no validation of the UCEIS performed While inter- and intra-observer variability was calculated for the Baron Score during the model development phase, correlation estimates are given for individual items, not the overall Baron Score

UCEIS: Ulcerative Colitis Endoscopic Index of Severity

Characteristics of studies awaiting assessment *[ordered by study ID]*

Iacucci 2017

Methods	Study describes the development and validation of a new electronic virtual chromoendoscopy score
Data	Not yet assessed
Comparisons	Not yet assessed
Outcomes	Not yet assessed
Notes	Full text article in press

Kim 2016

Methods	Retrospective validation study involving 154 biopsy specimens from 82 patients with UC
Data	Biospy specimens were reviewed by 2 blinded pathologists
Comparisons	Geboes Score (histology)
Outcomes	Not yet assessed
Notes	Endoscopic scoring index evaluated: Mayo Clinic Endoscopic Subscore

Lee 2016

Methods	This study aimed to test validity and reliability of the UCEIS in a Korean clinical setting. 36 videos of sigmoidoscopy in patients with UC were stratified according to disease activity using Mayo score
Data	To be assessed
Comparisons	To be assessed
Outcomes	Not yet assessed
Notes	Endoscopic scoring index evaluated: UCEIS

Songur 2009

Methods	Prospective validation study comparing the EAI to a histologic measurement tool
Data	96 UC patients
Comparisons	Histologic Activity Index
Outcomes	Construct validity
Notes	Endoscopic scoring index evaluated: EAI Waiting for full text; it is unclear what histologic activity index was used

UC: ulcerative colitis
UCEIS: Ulcerative Colitis Endoscopic Index of Severity
EAI: Endoscopic Activity Index

ADDITIONAL TABLES

Table 1. Partially validated endoscopic scoring indices

	Index	Reference	Validation study ID
1	Azzolini Classification	Azzolini 2005	Jun 2008
2	Baron Score	Baron 1964	Burger 2011; Hirai 2010; Jun 2008; Osada 2010; Thomas 2009; Walsh 2009
3	Blackstone Endoscopic Interpretation	Blackstone 1984	Osada 2010
4	CGSUC	Zou 2005	Jun 2008
5	Endoscopic Activity Index (EAI)	Naganuma 2010	de Lange 2004; Naganuma 2010
6	Jeroen Score	Jeroen 2002	Jun 2008
7	Magnifying Colonoscopy Grade	Nishio 2006	Nishio 2006
8	Matts Score	Matts 1961	Naganuma 2010; Osada 2010
9	Mayo Clinic Endoscopic Subscore	Schroeder 1987	Daperno 2011; Dhanda 2012; Osada 2010; Rubin 2012; Walsh 2009
10	Modified Mayo Clinic Endoscopic Subscore	Lobatón 2015	Levesque 2014
11	Modified Baron Score	Feagan 2005	Jun 2008; Levesque 2014; Walsh 2009
12	Osada Score (Modified 6-Point Activity Index)	Osada 2010	Osada 2010
13	Rachmilewitz Endocopic Score	Rachmilewitz 1989	Hirai 2010; Naganuma 2010; Schoepfer 2009
14	St. Mark's Index (Powell-Tuck Index)	Powell-Tuck 1982	Higgins 2005a
15	Ulcerative Colitis Colonoscopic Index of Severity (UCCIS)	Samuel 2013	Samuel 2013

Table 1. Partially validated endoscopic scoring indices *(Continued)*

16	Ulcerative Coltiis Disease Activity Index (endoscopic) (Sutherland Index)	Sutherland 1987	Higgins 2005a
17	Ulcerative Colitis Endoscopic Index of Severity (UCEIS)	Travis 2012	Levesque 2014; Travis 2013
18	Truelove and Witts Sigmoidoscopic Score	Truelove 1955	Jun 2008
19	Watson Grade	Kiesslich 2012	Kiesslich 2012

Table 2. Non-validated endoscopic scoring indices

	Index	Reference
1	Beattie Score	Beattie 1996
2	Binder Score	Binder 1970
3	Carbonnel Score	Carbonnel 1994
4	Danielsson-Löfberg Score	Danielsson 1987; Löfberg 1994
5	Dick Score	Dick 1964
6	Friedmann Score	Friedmann 1986
7	Froslie Endoscopic Score	Froslie 2007
8	Lemann Score	Lemann 1995
9	Levine Score	Levine 2002
10	Lindgren Score	Lindgren 2002
11	Maier Score	Maier 1988
12	McPhee Proctoscopic Grading Scale	McPhee 1987
13	Rutter Score	Rutter 2004
14	Saverymuttu Score	Saverymuttu 1986
15	Sigmoidoscopic Index	Hanauer 2004
16	Sigmoidoscopic Inflammation Grade Scale/Lemann Score	Lemann 1995
17	Truelove and Richards Sigmoidoscopic Appearance	Truelove 1956
18	van der Heide Index	van der Heide 1987

Table 3. Reliability

Study ID	Index	Inter-rater κ (between raters)	Inter-rater ICC (between raters)	Intra-rater κ (within rater)	Intra-rater ICC (within rater)	Internal Consistency
Daperno 2011	Mayo Clinic Endoscopic Subscore	pre-training: 0.445 post-training: 0.713				
Daperno 2014	Mayo Clinic Endoscopic Subscore	experts: 0.53 non-experts: 0.71				
de Lange 2004	EAI	experts: 0.97 (95% CI 0.92-1.00) non-experts: 0.79 (95% CI 0.71-0.49)				
Kiesslich 2012	Watson Grade	0.87				
Osada 2010	Modified 6-point Activity Index	experts: 0.65 trainees: 0.54		experts: 0.79 trainee: 0.64		
	Matts Score	experts: 0.76 trainees: 0.44		experts: 0.78 trainees: 0.41		
	The Mayo Endoscopic Subscore	experts: 0.74 trainees: 0.46		experts: 0.75 trainees: 0.48		
	Baron Score	experts: 0.61 trainees: 0.47		experts: 0.62 trainees: 0.46		
	Blackstone Score	experts: 0.57 trainees: 0.46		experts: 0.73 trainees: 0.51		
Samuel 2013	UCCIS		Vascular pattern rectum: 0.75 sigmoid: 0.81			

Table 3. Reliability *(Continued)*

			descending colon: 0.74 transverse colon: 0.86 ascending/cecum: 0.85 Granularity rectum: 0.70 sigmoid: 0.78 descending colon: 0.73 transverse colon: 0.88 ascending/cecum: 0.82 Ulceration rectum: 0.80 sigmoid: 0.75 descending colon: 0.72 transverse colon: 0.73 ascending/cecum: 0.73 Bleeding/Friability rectum: 0.68 sigmoid: 0.58 descending colon: 0.56 transverse colon: 0.73 ascending/cecum: 0.77 SAES rectum: 0.79 sigmoid: 0.78 descending colon: 0.71 transverse colon: 0.84 ascending/cecum: 0.85		
Travis 2013	UCEIS	0.50		0.72	0.863*

* Cronbach alpha analysis
SAES: segmental assessment of endoscopic severity

Table 4. Criterion Validity

Study ID	Index	Comparison	Correlation

Table 4. Criterion Validity *(Continued)*

Nishio 2006	Magnifying Colonscopy Grade	Mucosal IL-8 activity	ρ = NS (P < 0.001)
Samuel 2013	UCCIS	C-reactive protein	r = 0.56 (P < 0.001)
		albumin	r = -0.55 (P < 0.001)
		hemoglobin	r = -0.39 (P < 0.01)
		platelet count	r = 0.19 (P > 0.05)
Schoepfer 2009	Rachmilewitz Endoscopic Score	Fecal calprotectin	r = 0.834 (P < 0.001)
		C-reactive protein	r = 0.503 (P < 0.001)
		Blood leukocytes	r = 0.461 (P < 0.001)

Table 5. Construct Validity

Study ID	Index	Comparison	Correlation
Burger 2011	Baron Score	SCCAI	$\hat{k}$ = 0.27
		Truelove and Richards Index	$\hat{k}$ = 0.58
Dhanda 2012	Mayo Clinic Endoscopic Subscore	Riley Score	Week 4 r = 0.55
Higgins 2005a	St. Mark's Index	UCDAI	r = 0.881 (95% CI 0.814-0.925); ρ = 0.867
		SCCAI	r = 0.908 (95% CI 0.855-0.924); ρ = 0.866
		Seo Index	r = 0.803 (95% CI 0.699-0.873); ρ = 0.705
Hirai 2010	Baron Score	Rachmilewitz Score	Week 0 r = 0.39 (95% CI 0.18-0.57, P = 0.0004) Week 4 r = 0.56 (95% CI 0.36-0.71, P < 0.0001) Week 8 r = 0.76 (95% CI 0.60-0.85, P < 0.0001)
		UCDAI	Week 0 r = 0.49 (95% CI 0.29-0.64, P < 0.0001) Week 4 r = 0.72 (95% CI 0.57-0.82, P < 0.0001)

Table 5. Construct Validity *(Continued)*

		Week 8
		r = 0.85 (95% CI 0.74-0.91, P < 0.0001)
	Seo Index	Week 0
		r = 0.29 (95% CI 0.06-0.49, P = 0.01)
		Week 2
		r = 0.29 (95% CI 0.04-0.51, P = 0.02)
		Week 4
		r = 0.53 (95% CI 0.29-0.70, P < 0.0001)
	Lichtiger Index	Week 0
		r = 0.47 (95% CI 0.26-0.62, P < 0.0001)
		Week 4
		r = 0.56 (95% CI 0.35-0.71, P < 0.0001)
		Week 8
		r = 0.78 (95% CI 0.64-0.78, P < 0.0001)
Rachmilewitz Endoscopic Score	Rachmilewitz Score	Week 0
		r = 0.34 (95% CI 0.11-0.52, P = 0.0003)
		Week 2
		r = 0.66 (95% CI 0.48-0.78, P < 0.0001)
		Week 4
		r = 0.89 (95% CI 0.73-0.71, P < 0.0001)
	UCDAI	Week 0
		r = 0.44 (95% CI 0.23-0.60, P < 0.0001)
		Week 4
		r = 0.79 (95% CI 0.67-0.87, P < 0.0001)
		Week 8
		r = 0.89 (95% CI 0.82-0.94, P < 0.0001)
	Lichtiger Index	Week 0
		r = 0.35 (95% CI 0.13-0.54, P =0.002)
		Week 4
		r = 0.28 (95% CI 0.02-0.49, P = 0.003)
		Week 8
		r = 0.65 (95% CI 0.44 to 0.78, P < 0.0001)
	Seo Index	Week 0

Table 5. Construct Validity *(Continued)*

			r = 0.33 (95% CI 0.10-0.51, P = 0.005) Week 4 r = 0.67 (95% CI 0.50-0.79, P < 0.0001) Week 8 r = 0.80 (95% CI 0.67-0.88, P < 0.0001)
Jun 2008	CGSUC	Truelove and Witts Score	ρ = 0.750 (P < 0.001)
		Baron Score	ρ = 0.740 (P < 0.001)
		Modified Baron Score	ρ = 0.742 (P < 0.001)
		Jeroen Score	ρ = 0.799 (P < 0.001)
		Azzolini Score	ρ = 0.685 (P < 0.001)
	Truelove and Witts Score	CGSUC	ρ = 0.750 (P < 0.001)
		Baron Score	ρ = 0.814 (P < 0.001)
		Modified Baron Score	ρ = 0.760 (P < 0.001)
		Jeroen Score	ρ = 0.782 (P < 0.001)
		Azzolini Score	ρ = 0.756 (P < 0.001)
	Baron Score	CGSUC	ρ = 0.740 (P < 0.001)
		Truelove and Witts Score	ρ = 0.814 (P < 0.001)
		Modified Baron Score	ρ = 0.750 (P < 0.001)
		Jeroen Score	ρ = 0.828 (P < 0.001)
		Azzolini Score	ρ = 0.732 (P < 0.001)
	Modified Baron Score	CGSUC	ρ = 0.742 (P < 0.001)
		Baron Score	ρ = 0.760 (P < 0.001)
		Truelove and Witts Score	ρ = 0.750 (P < 0.001)
		Jeroen Score	ρ = 0.761 (P < 0.001)
		Azzolini Score	ρ = 0.693 (P < 0.001)
	Jeroen Score	CGSUC	ρ = 0.799 (P < 0.001)
		Baron Score	ρ = 0.782 (P < 0.001)
		Truelove and Witts Score	ρ = 0.828 (P < 0.001)
		Modified Baron Score	ρ = 0.761 (P < 0.001)

Table 5. Construct Validity *(Continued)*

		Azzolini Score	ρ = 0.788 (P < 0.001)
	Azzolini Score	CGSUC	ρ = 0.685 (P < 0.001)
		Truelove and Witts Score	ρ = 0.756 (P < 0.001)
		Baron Score	ρ = 0.732 (P < 0.001)
		Modified Baron Score	ρ = 0.693 (P < 0.001)
		Jeroen Score	ρ = 0.788 (P < 0.001)
Naganuma 2010	EAI	Lichtiger Index	r = 0.77 (P < 0.001)
		Matts Score	r = 0.91 (P < 0.001)
		Rachmilewitz Endoscopic Score	r = 0.87, (P < 0.001)
Nishio 2006	Magnifying Colonoscopy Grade	Riley Score	ρ = NS (P < 0.001)
Rubin 2012	Mayo Clinic Endoscopic Subscore	SCCAI	r = 0.525 (P < 0.0001)
		Rubin Histologic Score	r = 0.597 (P < 0.0001)
Samuel 2013	UCCIS	SCCAI	r = 0.62 (P < 0.0001)
		Rachmilewitz Score	r = 0.5 (P < 0.001)
		Patient-Defined Remission Score	r = 0.43 (P < 0.01)
Schoepfer 2009	Rachmilewitz Score (endoscopic)	Rachmilwitz Score (clinical)	r = 0.672 (P < 0.01)
Thomas 2009	Baron Score	Truelove and Richards Score	$\hat{k}$ = 0.58
		SCCAI	$\hat{k}$ = 0.27
Travis 2013	UCEIS	Visual Analogue Scale	median 0.93 across investigators (minimum 0.78, maximum 0.99) statistically significant P ≥ 0.05
Walsh 2009	Baron Score	Modified Baron Score	$\hat{k}$ = 0.89
	Baron Score	Mayo Endoscopic Subscore	$\hat{k}$ = 0.83

ρ = Spearman's rank correlation coefficient
Abbreviations: CGSUC, Chinese Grading Score for Ulcerative Colitis; EAI, Endoscopic Activity Index; IL, Interleukin; NS, Not Stated; SCCAI, Simple Clinical Colitis Activity Index

Table 6. Responsiveness

Study ID	Index	Treatment	Effect size (95% CI)	Guyatt's responsiveness statistic (95% CI)	Area under the ROC curve (95% CI)	Mean change (P value)
Levesque 2014	Mayo Clinic Endoscopic Subscore	Asacol	0.49 (0.28, 0.71)	0.32 (0.11, 0.53)	0.66 (0.55, 0.78)	
	Modified Baron Score		0.49 (0.28, 0.71)	0.33 (0.13, 0.54)	0.65 (0.54, 0.77)	
	UCEIS		0.58 (0.36, 0.81)	0.47 (0.25, 0.69)	0.68 (0.58, 0.79)	
Ikeya 2016	Mayo Clinic Endoscopic Subscore	Tacrolimus				2.9 (+/- 0.9) to 2.0 (+/- 1.0) (P < 0.001)
	UCEIS					6.2 (+/- 0.9) to 3.4 (+/- 2.1) (P < 0.001)

Table 7. The Methodological Quality of Endoscopic Index Measurement Properties as Described in the Original Development Articles (COSMIN Checklist)

	A	B	C	D	E	F	G	H	I	J	
Study ID	**IC**	**RB**	**ME**	**COV**	**FA**	**HT**	**CCV**	**CRV**	**RP**	**IT**	**GN**
Burger 2011	-	-	-	-	-	good	-	-	-	-	-
Daperno 2011	-	good	-	-	-	-	-	-	-	-	-
Daperno 2014	-	-	-	-	-	-	-	-	-	-	-
de Lange 2004	-	good	-	-	-	-	-	-	-	-	-
Dhanda 2012	-	-	-	-	-	excellent	-	-	-	-	-
Higgins 2005a	-	-	-	-	-	good	-	-	-	-	-
Hirai 2010	-	-	-	-	-	good	-	-	-	-	-
Ikeya 2016	-	-	-	-	-	-	-	-	fair	-	-
Jun 2008	-	-	-	-	-	good	-	-	-	-	-
Kiesslich 2012	-	good	-	-	-	-	-	-	-	-	-
Levesque 2014	-	-	-	-	-	-	-	-	excel-lent	-	-
Naganuma 2010	-	-	-	-	-	excellent	-	-	-	-	-
Nishio 2006	-	-	-	-	-	good	-	good	-	-	-
Osada 2010	-	excel-lent	-	-	-	-	-	-	-	-	-
Rubin 2012	-	good	-	-	-	good	-	-	-	-	-
Samuel 2013	-	excel-lent	-	-	-	excellent	-	ex-cel-lent	-	-	-

Table 7. The Methodological Quality of Endoscopic Index Measurement Properties as Described in the Original Development Articles (COSMIN Checklist) *(Continued)*

Schoepfer 2009	-	-	-	-	-	excellent	-	excellent	-	-	-
Thomas 2009	-	-	-	-	-	good	-	-	-	-	-
Travis 2013	-	good	-	-	-	good	-	-	-	-	-
Walsh 2009	-	-	-	-	-	excellent	-	-	-	-	-

IC - internal consistency; RB - reliability; ME - measurement error; COV - content validity; FA - factor analysis; HT - hypothesis testing; CCV - cross cultural validity; CRV - criterion validity; RP - responsiveness; IT - interpretability; GN - generalizability

Table 8. Summary of operating properties of histologic scoring indices for Crohn's disease

Scoring index	Validity			Reliability				Responsiveness	Feasibility
	Content validity	Criterion validity	Construct validity	Intra-rater	Inter-rater	Test-retest	Internal consistency		
Azzolini Classification	?	?	+	?	?	?	?	?	?
Baron Score	?	?	+	+	+	?	?	?	?
Blackstone Endoscopic Interpretation	?	?	?	+	+	?	?	?	?
CGSUC	?	?	+	?	?	?	?	?	?
Endoscopic Activity Index (EAI)	?	?	+	?	+	?	?	?	?
Jeroen Score	?	?	+	?	?	?	?	?	?
Magnifying Colonscopy Grade	?	+	+	?	?	?	?	?	?
Matts Score	?	?	+	+	+	?	?	?	?
Mayo Clinic Endoscopic Subscore	?	?	+	+	+	?	?	+	?

Table 8. Summary of operating properties of histologic scoring indices for Crohn's disease *(Continued)*

Modified Mayo Clinic Endoscopic Subscore	?	?	?	?	?	?	?	+	?
Modified Baron Score	?	?	+	?	+	?	?	+	?
Osada Score (Modified 6-Point Activity Index)	?	?	?	+	+	?	?	?	?
Rachmilewitz Endocopic Score	?	+	+	?	?	?	?	?	?
St. Mark's Index (Powell-Tuck Index)	?	?	+	?	?	?	?	?	?
Ulcerative Colitis Colonoscopic Index of Severity (UCCIS)	?	?	?	?	?	?	?	?	?
Ulcerative Coltiis Disease Activity Index (endoscopic) (Sutherland Index)	?	?	?	?	?	?	?	?	?
Ulcerative Colitis Endoscopic Index of Severity (UCEIS)	?	?	?	?	?	?	?	?	?
Truelove and Witts Sigmoidoscopic Score	?	?	?	?	?	?	?	?	?
Watson Grade	?	?	?	?	+	?	?	?	?

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

APPENDICES

Appendix 1. Search strategies

MEDLINE and Embase

1 colitis.ti.

2 inflammatory bowel disease.ti.

3 IBD.ti.

4 (baron or blackstone or "endoscopic activity index" or Matts or Matts' or Matt's or Mayo or Rachmilewitz or Mark's or "Ulcerative Colitis Colonoscopic Index of Severity" or "UCCIS" or "Ulcerative Colitis Disease Activity Index" or "UCDAI" or Sutherland or UCEIS or Truelove).ab.

5 (baron or blackstone or "endoscopic activity index" or Matts or Matts' or Matt's or Mayo or Rachmilewitz or Mark's or "Ulcerative Colitis Colonoscopic Index of Severity" or "UCCIS" or "Ulcerative Colitis Disease Activity Index" or "UCDAI" or Sutherland or UCEIS or Truelove).ti.

6 depth.ti.

7 depth.ab.

8 (mucosal adj2 heal*).mp.

9 (mucosal adj2 improv*).mp.

10 (endoscop* adj2 heal*).mp.

11 (endoscop* adj2 improv*).mp.

12 (endoscop* adj respon*).mp.

13 (endoscop* adj2 remission).mp.

14 "stable remission".mp.

15 "deep remission".mp.

16 endoscop*.ti.

17 colonoscop*.ti.

18 sigmoidoscop*.ti.

19 scor*.ti.

20 scale.ti.

21 index*.ti.

22 indice*.ti.

23 grad*.ti.

24 valid*.ti.

25 valid*.ab.

26 inter-rater.ti. or inter-rater.ab.

27 interrater.ti. or interrater.ab.

28 intra-rater.ti. or intra-rater.ab.

29 intrarater.ti. or intrarater.ab.

30 inter-obsever.ti. or inter-observer.ab.

31 interobserver.ti. or interobserver.ab.

32 intra-observer.ti. or intra-observer.ab.

33 intraobserver.ti. or intraobserver.ab.

34 agree*.ti. or agree*.ab.

35 correlat*.ti.

36 correlat*.ab.

37 feasib*.ti. or feasib*.ab.

38 assess*.ti. or assess*.ab.

39 measure*.ti. or measure*.ab.

40 compar*.ti. or compar*.ab.

41 variab*.ti. or variab*.ab.

42 or/1-5

43 or/6-18

44 or/19-42

45 or/42-44

46 42 and 43 and 44 and 45

CENTRAL

#1 colitis

#2 inflammatory bowel disease

#3 IBD

#4 baron or blackstone or "endoscopic activity index" or Matts or Matts' or Matt's or Mayo or Rachmilewitz or Mark's or "Ulcerative Colitis Colonoscopic Index of Severity" or "UCCIS" or "Ulcerative Colitis Disease Activity Index" or "UCDAI" or Sutherland or UCEIS or Truelove

#5 depth

#6 mucosal heal*

#7 mucosal improv*

#8 endoscop* heal*

#9 endoscop* improv*

#10 endoscop* respon*

#11 endoscop* remission

#12 stable remission

#13 deep remission

#14 endoscop

#15 colonoscop

#16 sigmoidoscop

#17 scor

#18 scale

#19 index

#20 indice*

#21 grad*

#22 valid*

#23 valid*

#24 inter-rater

#25 interrater

#26 intra-rater

#27 intrarater

#28 inter-observer

#29 interobserver

#30 intra-observer

#31 intraobserver

#32 agree*

#33 correlat*

#34 feasib*

#35 assess*

#36 measure*

#37 compar*

#38 variab*

#39 #1 or #2 or #3 or #4

#40 #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16

#41 #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38

#42 #30 and #40 and #41

CONTRIBUTIONS OF AUTHORS

Development of concept: Mark Samaan, Mahmoud H Mosli, Claire E Parker, Sigrid A Nelson, John K MacDonald, Brian G Feagan, GY Zou, Vipul Jairath, Reena Khanna; drafting of manuscript: Nadia Mohammed Vashist, Claire E Parker; critical revision of the manuscript: Nadia Mohammed Vashist, Claire E Parker, Mark Samaan, Mahmoud H Mosli, Claire E Parker, Sigrid A Nelson, John K MacDonald, Brian G Feagan, GY Zou, Reena Khanna, Vipul Jairath.

DECLARATIONS OF INTEREST

Nadia Mohammed Vashist: None known

Mark Samaan: None known

Mahmoud H Mosl: None known

Claire E Parker: None known

John K MacDonald: None known

Sigrid A Nelson: None known

GY Zou: None known

Brian G Feagan has received Scientific Advisory Board fees from Abbott/AbbVie, Allergan, Amgen, Astra Zeneca, Atlantic Pharma, Avaxia Biologics Inc., Boehringer-Ingelheim, Bristol-Myers Squibb, Celgene, Centocor Inc., Elan/Biogen, Ferring, Galapagos, Genentech/Roche, JnJ/Janssen, Merck, Nestles, Novartis, Novonordisk, Pfizer, Prometheus Laboratories, Protagonist, Salix Pharma, Takeda, Teva, TiGenix, Tillotts Pharma AG, and UCB Pharma; consulting fees from Abbott/AbbVie, Ablynx, Akebia Therapeutics, Allergan, Amgen, Applied Molecular Transport Inc., Aptevo Therapeutics, Astra Zeneca, Atlantic Pharma, Avir Pharma, Baxter Healthcare Corp., Biogen Idec, Boehringer-Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan/Biogen, EnGene, Ferring Pharma, Roche/Genentech, Galapagos, GiCare Pharma, Gilead, Given Imaging Inc., GSK, Inception IBD Inc, Ironwood Pharma, Janssen Biotech (Centocor), JnJ/Janssen, Kyowa Kakko Kirin Co Ltd., Lexicon, Lilly, Lycera BioTech, Merck, Mesoblast Pharma, Millennium, Nektar, Nestles, Nextbiotix, Novonordisk, Pfizer, Prometheus Therapeutics and Diagnostics, Progenity, Protagonist, Receptos, Roche/Genentech, Salix Pharma, Serano, Shire, Sigmoid Pharma, Synergy Pharma Inc., Takeda, Teva Pharma, TiGenix, Tillotts, UCB Pharma, Vertex Pharma, Vivelix Pharma, VHsquared Ltd., Warner-Chilcott, Wyeth, Zealand, Zyngenia; grants/grants pending from AbbVie Inc., Amgen Inc., AstraZeneca/MedImmune Ltd., Atlantic Pharmaceuticals Ltd., Boehringer-Ingelheim, Celgene Corporation, Celltech, Genentech Inc/Hoffmann-La Roche Ltd., Gilead Sciences Inc., GlaxoSmithKline (GSK), Janssen Research & Development LLC., Pfizer Inc., Receptos Inc. / Celgene International, Sanofi, Santarus Inc., Takeda Development Center Americas Inc., Tillotts Pharma AG, UCB; and lecture fees from Abbott/AbbVie, JnJ/Janssen, Lilly, Takeda, Tillotts, UCB Pharma.

Reena Khanna has received honoraria from AbbVie, Jansen, Pfizer, Shire, Takeda, and Robarts Clinical Trials for consultancy. All of these activities are outside the submitted work.

Vipul Jairath has received scientific advisory board fees from Abbvie, Sandoz, Ferring, Pfizer and Janssen; speakers fees from Takeda, Ferring, Shire Janssen, Pfizer; travel support for conference attendance from Vifor pharmaceuticals. All of these activities are outside the submitted work.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

The methods for assessing the risk of bias in included studies was modified from the protocol. It was planned that risk of bias was to be assessed using blinded design, independent observation, performance bias and detection bias. Since this is a review of scoring indices rather than interventions, the last two items are not applicable. We chose to assess blinded design and independent observation combined with the use of a system based on the COSMIN tool to further assess risk bias.

The method for interpreting correlation coefficients was modified from the protocol. In the protocol we indicated that we would use the Landis and Koch criteria for the interpretation of correlation coefficients that were generated to assess observer agreement (Landis 1977). For the interpretation of correlation coefficients calculated to assess the direction and strength of a relationship between two variables (e.g. UCEIS and CRP), we decided to use the Cohen criteria (Cohen 1992).

INDEX TERMS

Medical Subject Headings (MeSH)

*Colonoscopy; Colitis, Ulcerative [*diagnosis] [pathology]; Reproducibility of Results; Severity of Illness Index; Sigmoidoscopy

MeSH check words

Humans