Cochrane
Library

Cochrane Database of Systematic Reviews

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

Mosli MH, Parker CE, Nelson SA, Baker KA, MacDonald JK, Zou GY, Feagan BG, Khanna R, Levesque BG, Jairath V

Mosli MH, Parker CE, Nelson SA, Baker KA, MacDonald JK, Zou GY, Feagan BG, Khanna R, Levesque BG, Jairath V.
Histologic scoring indices for evaluation of disease activity in ulcerative colitis.
Cochrane Database of Systematic Reviews 2017, Issue 5. Art. No.: CD011256.
DOI: [10.1002/14651858.CD011256.pub2](https://doi.org/10.1002/14651858.CD011256.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
BACKGROUND	3
OBJECTIVES	3
METHODS	3
RESULTS	5
Figure 1.	6
Figure 2.	8
DISCUSSION	10
AUTHORS' CONCLUSIONS	11
ACKNOWLEDGEMENTS	11
REFERENCES	12
CHARACTERISTICS OF STUDIES	18
ADDITIONAL TABLES	25
APPENDICES	34
WHAT'S NEW	35
CONTRIBUTIONS OF AUTHORS	35
DECLARATIONS OF INTEREST	35
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	35
NOTES	36
INDEX TERMS	36

[Methodology Review]

Histologic scoring indices for evaluation of disease activity in ulcerative colitis

Mahmoud H Mosli^{1,2}, Claire E Parker³, Sigrid A Nelson⁴, Kenneth A Baker³, John K MacDonald^{5,6}, GY Zou^{3,7}, Brian G Feagan^{3,5,6,7}, Reena Khanna^{3,6}, Barrett G Levesque⁴, Vipul Jairath^{3,6,7}

¹King Abdulaziz University Hospital, King Abdulaziz University, Jeddah, Saudi Arabia. ²Robarts Clinical Trials, Robarts Research Institute, London, Canada. ³Robarts Clinical Trials, London, Canada. ⁴Robarts Clinical Trials, San Diego, CA, USA. ⁵Cochrane IBD Group, Robarts Clinical Trials, London, Canada. ⁶Department of Medicine, University of Western Ontario, London, Canada. ⁷Department of Epidemiology and Biostatistics, University of Western Ontario, London, Canada

Contact: Vipul Jairath, Department of Medicine, University of Western Ontario, London, ON, Canada. vipul.jairath@robartsinc.com.

Editorial group: Cochrane IBD Group.

Publication status and date: Edited (no change to conclusions), published in Issue 7, 2017.

Citation: Mosli MH, Parker CE, Nelson SA, Baker KA, MacDonald JK, Zou GY, Feagan BG, Khanna R, Levesque BG, Jairath V. Histologic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No.: CD011256. DOI: [10.1002/14651858.CD011256.pub2](https://doi.org/10.1002/14651858.CD011256.pub2).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Disease activity can be determined using clinical, endoscopic or histologic criteria in patients with ulcerative colitis (UC). Persistent disease activity is associated with poor outcomes. Histologic disease activity has been shown to be associated with relapse, colectomy and colorectal cancer. The ability to objectively evaluate microscopic disease activity using histology is important for both clinical practice and clinical trials. However, the operating properties of the currently available histologic indices remain unclear.

Objectives

A systematic review was undertaken to identify and evaluate the development and operating characteristics of histologic disease activity indices used to assess disease activity in people with ulcerative colitis.

Search methods

We searched MEDLINE, EMBASE, PubMed, CENTRAL and the Cochrane IBD Review Group Specialized Trials Register from inception to 2 December 2016 for applicable studies. There were no language or document type restrictions.

Selection criteria

Any study design (e.g. randomized controlled trials, cohort studies, case series) that evaluated a histologic index in patients with UC were considered for inclusion. Eligible patients were adults (≥ 18 years), diagnosed with UC using conventional clinical, radiographic, endoscopic and histologic criteria.

Data collection and analysis

Two authors (MHM and CEP) independently reviewed the titles and abstracts of the studies identified from the literature search. A standardized form was used to assess eligibility of trials for inclusion and for data extraction.

Two authors (MHM and CEP) independently extracted and recorded data, which included the number of patients enrolled, number of patients per treatment arm, patient characteristics including age and gender distribution, and the name of the histologic index. Outcomes (i.e. intra-rater reliability, inter-rater reliability, internal consistency, content validity, criterion validity, construct validity, responsiveness, and feasibility) were recorded for each trial.

Main results

In total, 126 reports describing 30 scoring indices were identified through the screening process. Eleven of the 30 scoring indices have undergone some form of index validation. Intra-rater reliability was assessed for eight scoring indices. Inter-rater reliability was evaluated for all 11 of the scoring indices. Three of the indices underwent content validation. Two of the included scoring indices assessed criterion validity. Six of the included scoring indices explored content validity. Two of the included scoring indices were tested for responsiveness.

Authors' conclusions

The Nancy Index and the Robarts Histopathology Index have undergone the most validation in that four operating properties including reliability, content validity, construct validity (hypothesis testing) and criterion validity have been tested. However, none of the currently available histologic scoring indices have been fully validated. In order to determine the optimal endpoint for histologic healing in UC, more research is required. The optimal index would need to be fully validated.

PLAIN LANGUAGE SUMMARY

Histologic measurement tools for evaluating disease in ulcerative colitis patients

What is ulcerative colitis?

Ulcerative colitis is a life-long (chronic) inflammatory bowel disease that causes inflammation and ulceration (sores) in the large intestine (colon). Patients with ulcerative colitis often experience diarrhoea, bloody stools, weight loss and abdominal pain. When patients experience symptoms, the disease is considered "active", whereas when symptoms are not present the disease is considered to be "in remission".

What is a histologic scoring index?

A histologic tissue sample (biopsy) can be taken from a patient's colon during colonoscopy. A colonoscopy is a non-surgical procedure used to view the large intestine. Once the tissue sample is taken, it is placed (mounted) on a glass slide and looked at using a microscope. A histologic scoring index is a system used to assess the patient's disease severity using the tissue sample.

What did the researchers investigate?

It is important that histologic scoring indices are valid (i.e. they accurately measure what they are supposed to measure). The researchers identified histologic scoring indices that have been validated.

What did the researchers find?

The researchers found that 11 out of the 30 histologic scoring indices that exist have been partially validated. The Nancy Index and the Robarts Histopathology Index have undergone the most validation compared to the other nine indices. However, none of the currently available histologic scoring indices have been fully validated. In order to determine the ideal index to measure histologic healing in UC, more research is required. The ideal index would need to be fully validated.

BACKGROUND

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) of unknown etiology that occurs in genetically predisposed individuals. The spectrum of disease in UC ranges from mild diarrhoea to fulminant colitis requiring hospitalisation and possibly emergent colectomy. Patients with UC typically present with bloody diarrhoea and abdominal cramps. Patients with chronically uncontrolled disease can develop long-term complications such as colorectal cancer or need for colectomy (Abraham 2009). In order to avoid such complications, pharmacologic treatments, including oral or rectal aminosalicylates, or both, corticosteroids, immunosuppressives, tumour necrosis factor-alpha antagonists, and anti-integrin therapy are initiated with the goal of inducing and maintaining clinical and endoscopic remission (Cohen 2002; Baumgart 2007). In clinical trials, investigators rely on composite scores to assess disease activity, which frequently involves determining the presence or absence of colonic inflammation by standard colonoscopy (Truelove 1955). Although microscopic inflammation or histologic activity has been linked to an overall poor outcome, it has not been adopted by investigators as a routine outcome in clinical trials, mostly due to the lack of a gold-standard, to incremental costs and to uncertainty regarding its practical value (Pineton de Chambrun 2010; Ardizzone 2011).

Histology plays an important role in diagnosing UC and can serve as a tool to determine response to therapy. Multiple histologic scoring systems have been developed to quantify colonic micro-inflammation seen in tissue samples obtained during colonoscopy in a categorical or numerical way. Such tools determine the degree of acute or chronic inflammatory cell infiltrates, the presence or absence of architectural distortion of colonic crypts, and the integrity of the colonic epithelium.

The first histologic index used in clinical trials for UC was the Truelove and Richards index (Truelove 1956). This was followed by the Matts Score which was described in 1961 and first applied to 126 serial biopsies showing that there was a direct relationship between endoscopic and histologic activity (Matts 1961). Similarly, the Watts Score demonstrated that a preserved mucosal vascular pattern is almost always indicative of microscopically inactive disease (Watts 1966).

The Initial Riley Score was described in 1988 as an evaluative instrument to determine mucosal healing in a randomized controlled trial (RCT) that compared delayed release mesalamine to enterically coated sulfasalazine and placebo as maintenance therapy for clinically quiescent UC. In this study, two blinded pathologists independently graded inflammation according to five levels which subjectively categorize tissue samples based on the degree of chronic inflammation and tissue destruction (Riley 1988). Riley 1991 subsequently described the widely used Riley Score in a clinical trial designed to predict relapse in clinically and endoscopically quiescent UC patients. This score incorporates six histologic features commonly used to determine disease activity which include: acute inflammatory cell infiltrate (neutrophils in the lamina propria), crypt abscesses, mucin depletion, surface epithelial integrity, chronic inflammatory cell infiltrate (round cells in the lamina propria), and crypt architectural irregularities. Each feature was graded on a four-point scale as none, mild, moderate, or severe (Riley 1991). This score was modified by Feagan

2005 and used as an outcome in a large multicenter randomised placebo-controlled trial evaluating the use of vedolizumab for the treatment of active UC. The Modified Riley Score removed features of chronicity that were thought to be resistant to responsiveness (Feagan 2005).

The Geboes Score, a commonly used histologic index for UC, was developed in 2000 using a multivariate regression model which resulted in an index composed of seven categories (Geboes 2000).

The Chicago Index, also known as the Rubin Index, has not been widely used in clinical practice or clinical trials as it has only been reported in abstract form. This six-point histologic index was used in a case-control study in which two blinded pathologists graded UC inflammation without clinical knowledge and as a result disease activity was determined to be an independent risk factor for colonic neoplasia. This result was confirmed by a subsequent analysis between histologic activity and increased risk of future colectomy and hospitalizations (Rubin 2007). The Riley Index has also been shown to correlate directly with the risk of neoplasia, a feature that was also seen with the Harpaz Index, also known as the Mount Sinai Index or Fiel Index (Fiel 2003).

Why it is important to do this review

There are few data available on the operating properties of existing histologic scoring indices despite widespread use in clinical trials. This review will evaluate the relative merits of histologic indices that have undergone validation testing in order to underscore where further research is needed.

OBJECTIVES

The primary objective is to systematically review the current literature describing the development and operating characteristics of histologic disease activity indices in UC.

METHODS

Criteria for considering studies for this review

Types of studies

Any study design (e.g. randomised controlled trials, cohort studies, case series) that evaluates a histologic index in patients with UC was considered for inclusion. Study subjects included adult patients (≥ 18 years), diagnosed with UC using conventional clinical, radiographic, histologic and endoscopic criteria.

Types of data

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

Types of methods

The methods used to construct and validate the histologic index (e.g. reliability, validity, responsiveness and feasibility) were examined in detail and described for each eligible study. We also reported the number of pathologists who scored the histologic index in each study and whether these pathologists were blinded or were aware of other rater's scores.

Types of outcome measures

Reliability: Reliability was assessed by recording reports of intra-rater and inter-rater reliability, test-retest reliability, or internal consistency. Intra-rater and inter-rater reliability are assessed by determining the inter-class correlation coefficient (ICC) or kappa statistic for repeat assessments made by the same rater, and for assessments made by different raters. The Landis and Koch criteria was used to interpret the ICC and kappa values. An ICC of < 0.2 was considered 'slight', 0.21-0.40 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)

Validity: Each study was assessed to determine whether validity was measured, broadly defined as evidence that variations in UC activity causally produce variations in the index measurement outcomes. Studies were reviewed for whether content validity, criterion validity, and construct validity for histologic index scores in specific clinical situations were reported.

Studies attempting to demonstrate content validity are successful if the components of the histologic index are sufficient to measure disease activity in UC. Generally, content validation is qualitatively assessed. For example, an expert panel may be asked to give an opinion on face validity, or a systematic review of the literature may be conducted to support the development of an index.

Criterion validity is considered to be established if the index is considered to be an adequate reflection of true UC disease activity, as assessed against a gold standard of measurement. Unfortunately, there is no single gold standard for assessing histologic activity in UC, which limited, but did not prevent, this kind of assessment. Statistical parameters reported for agreement between the histologic index and objective biomarkers were assessed (i.e. sensitivity, specificity, receiver operating characteristic (ROC) curve, area under the curve (AUC), mean difference, weighted kappa, Spearman's r squared, and the ICC). Data from studies of predictive criterion validity, which compare whether the score predicts true UC activity or sequelae in the future (such as surgery, or disability) were also recorded.

Studies that reported on the construct validity of the histologic indices, which takes into account the lack of a gold standard for disease activity and assesses whether histologic indices are consistent with other hypotheses of true disease activity, were included in the current review. For example, correlations between the histologic index and clinical and endoscopic indices were recorded.

Responsiveness: The ability of the index to detect change following a period of known histologic change (e.g. after a treatment of known efficacy is administered) serves as an assessment of responsiveness. Responsiveness can be quantified by examining the correlation between mean change scores between indices and indicators of effect size or its functions (Zou 2005), or the use of ROC curves to describe how well various score changes can distinguish between improved and unimproved patients (Deyo 1991).

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

Search methods for identification of studies

Electronic searches

We searched MEDLINE (Ovid), EMBASE (Ovid), PubMed, the Cochrane Library (CENTRAL), and the Cochrane IBD Group Specialized Trials Register from inception to December 2, 2016 for applicable studies. No language or document type restrictions were applied. The search strategies are listed in [Appendix 1](#).

Searching other resources

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

1. Digestive Disease Week (DDW);
2. United European Gastroenterology Week (UEGW); and
3. European Crohn's and Colitis Organization (ECCO) annual conference.

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 (MHM and CEP) independently reviewed the titles and abstracts of the studies identified from the literature search. The full text of potentially relevant citations were reviewed for inclusion and the study investigators were contacted to clarify any unclear data. Any disagreements were resolved by discussion and consensus with a third author (BGL).

A standardized form was used to assess eligibility of trials for inclusion in the study based on the inclusion criteria outlined above.

Data extraction and management

A standardized form was used to extract information from selected studies. Two authors (MHM, CEP) 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: age and gender distribution;
- c) Histologic Index used; and,
- d) Outcomes: measures of intra-rater reliability; inter-rater reliability; responsiveness; validity; feasibility; construct validity; 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
- Independent observation by endoscopists

Blinding to clinical information such as symptoms, physical examination or laboratory information is important to the objective assessment of histologic data. Furthermore, independent observation is essential to ensure that we are confident in the inter-rater reliability coefficients.

We also assessed the methodological quality of the included studies using the COSMIN (CONsensus-based STANDards for the selection of health Measurement INSTRumets) checklist. The checklist consists of 10 properties: internal consistency, reliability, measurement error, content validity, structural validity (factor analysis), hypothesis testing, cross-cultural validity, criterion validity, responsiveness to change and interpretability. Each property is rated on a four-point scale (1 = poor, 2 = fair, 3 = good, or 4 = excellent). The overall score for the assessment of an individual measurement property is obtained by taking the lowest score for any of the items in the property box (i.e. if any item in the property box is scored as "poor" then the overall score for that property is 'poor').

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

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

RESULTS

Description of studies

Results of the search

In total, 6036 results were retrieved by the search strategies. After the exclusion of 2615 duplicates, 3421 citations were assessed for eligibility. Of these citations, 3295 were deemed non-applicable and 126 reports of 30 histologic scoring indices were identified. Nineteen of the histologic scoring indices (described in 44 reports) were excluded due to a lack of validation testing. Eleven of the histologic scoring indices (described in 82 reports), which have been partially validated were included in the systematic review (see [Figure 1](#)).

Figure 1. Study flow diagram.

Included studies

The eleven histologic scoring indices (Feagan 2005; Fiel 2003; Geboes 2000; Gomes 1986; Jauregui-Amezaga 2016; Marchal-Bressenot 2017; Mosli 2017; Riley 1991; Rubin 2007; Theede 2015; Truelove 1956), that have been partially validated are described in the [Characteristics of included studies](#) tables and [Table 1](#).

Setting

Validation testing of four of the scoring indices (the Geboes Score, Modified Riley Score, Nancy Index and Robarts Histopathology Index) was based on retrospectively collected data. In [Mosli 2017](#), five pathologists assessed 154 biopsy slides, previously collected during an RCT of MLN02 (Feagan 2005), three times using the Geboes and Modified Riley Scores. Subsequently, four pathologists were asked to read 50 biopsy slides (taken from the same RCT) three times using the Geboes and Modified Riley Scores with standardized scoring conventions applied. During the responsiveness phase of [Mosli 2017](#), a single central reader assessed 154 pairs of slides (each pair consisting of a baseline biopsy and a week 4 or 6, post-treatment biopsy) using the Geboes Score, Modified Riley Score and newly created Robarts Histopathology Index.

The validation of the Nancy Index employed retrospectively collected cohort data. During the reliability testing phase, three pathologists scored 100 biopsies using the Geboes Score and the newly created Nancy Index ([Marchal-Bressenot 2017](#)). To test for responsiveness, 30 pairs of biopsy slides (that showed a histologic change) were assessed by a single expert pathologist according to the Geboes Score and Nancy Index.

Validation testing of seven of the scoring indices (Geboes Score, Gomes Index, Simplified Geboes Score, Riley Score, Chicago Index, Modified Harpaz Index and Truelove and Richards Index) was done using prospectively collected data. The development and initial reliability testing of the Geboes Score was conducted with prospective RCT data. In this study, 99 biopsies (68 obtained from inflamed mucosa and 31 from endoscopically non-inflamed mucosa) were assessed on two occasions by three pathologists ([Geboes 2000](#)).

The remaining six scoring indices were partially validated using prospectively collected observational data. Criterion validation was performed by a single pathologist for the Gomes Index using biopsies from 28 UC patients undergoing routine colonoscopy. In [Jauregui-Amezaga 2016](#), the Geboes Score and Simplified Geboes Score were tested for reliability by two trained pathologists using biopsy specimens from 92 patients requiring colonoscopy at a

tertiary referral center. The Riley Score was evaluated for reliability by two pathologists using biopsies taken from 82 patients with a confirmed diagnosis of UC who were asymptomatic and in endoscopic remission ([Riley 1991](#)). In order to explore criterion validity, the Riley Score and Mayo Endoscopic Subscore were calculated by a single pathologist using 263 biopsies from 131 UC patients. [Rubin 2007](#) performed construct validation by having pathologists (number not reported) assess biopsies of 86 UC patients requiring standard colonoscopy or sigmoidoscopy with the Chicago Index. Clinical and endoscopic disease activity were assessed using the Simple Colitis Clinical Activity Index and the Mayo Endoscopic Subscore, respectively. In [Theede 2015](#), biopsies from 120 UC patients with inactive or active disease requiring sigmoidoscopy were collected, and two pathologists scored specimens according to the Modified Harpaz Index. Construct validity was also assessed using the Mayo Clinic Endoscopic Subscore and the Ulcerative Colitis Endoscopic Index of Severity. Finally, the Truelove and Richards Index was used by two pathologists to assess 91 biopsies from UC patients who had presented at a tertiary referral center requiring sigmoidoscopy ([Truelove 1956](#)). Clinical disease activity using the Simple Clinical Colitis Activity Index and endoscopic disease activity using the Baron Score were also evaluated.

It is unclear whether the Harpaz Index was partially validated using prospective or retrospectively collected data, since this study is available in abstract form only ([Fiel 2003](#)).

Excluded studies

Nineteen histologic scoring indices ([Baars 2012](#); [D'Argenio 2001](#); [Floren 1987](#); [Friedman 1985](#); [Gramlich 2007](#); [Hanauer 1993](#); [Iacucci 2015](#); [Keren 1984](#); [Korelitz 1976](#); [Matts 1961](#); [Nishiyama 2014](#); [Odze 1993](#); [Powell-Tuck 1982](#); [Riley 1988](#); [Rutter 2004](#); [Sandborn 1993](#); [Saverymutter 1986](#); [Watts 1966](#); [Wiernicka 2015](#)) have not undergone any validation procedures, and were therefore excluded ([Characteristics of excluded studies](#) and [Table 2](#)).

Risk of bias in included studies

Blinding

In seven of the eleven validated scoring indices ([Feagan 2005](#); [Geboes 2000](#); [Gomes 1986](#); [Marchal-Bressenot 2017](#); [Mosli 2017](#); [Theede 2015](#); [Truelove 1956](#)), the individual(s) who performed the histologic assessment were blinded to other relevant patient data (e.g. clinical and endoscopic data). The other four indices that underwent validation testing were reported in abstract form only, and did not include sufficient information on blinding to allow a judgment ([Figure 2](#)).

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

Independent Observation

A total of six of the eleven validated scoring indices indicated that the histologists who performed the assessments did so in an independent fashion (Feagan 2005; Geboes 2000; Marchal-Bressenot 2017; Mosli 2017; Riley 1991; Truelove 1956). Three of the included scoring indices did not adequately describe whether independent assessment took place (Jauregui-Amezaga 2016; Rubin 2007; Theede 2015). This risk of bias item was not relevant in the case of Gomes 1986, as only criterion validity (> one reader not required) was assessed for this scoring index (Figure 2).

Effect of methods

Reliability

Eight of the included scoring indices (the Modified Riley Score, Harpaz Index, Geboes Score, Modified Geboes Score, Nancy Index, Roberts Histopathology Index, Riley Score, Modified Harpaz Index) have undergone reliability testing (Feagan 2005; Fiel 2003; Geboes 2000; Jauregui-Amezaga 2016; Marchal-Bressenot 2017; Mosli 2017; Riley 1991; Theede 2015).

Intra-rater reliability (i.e. agreement with self over multiple assessments of the same biopsy slide) was calculated for four of these indices: the Modified Riley Score, the Geboes Score, the Nancy Index and the Roberts Histopathology Index (Feagan 2005; Geboes 2000; Marchal-Bressenot 2017; Mosli 2017). According to the Landis and Koch criteria, correlation estimates ranged from 'substantial' to 'almost perfect'. Mosli 2017 evaluated intra-rater reliability of the Modified Riley Score and Geboes Score on two occasions: once before standardized scoring conventions were applied, and once after. Initially, the ICCs were 0.71 (95% CI 0.63 to 0.80) for the Modified Riley Score and 0.82 (95% CI 0.73 to 0.88) for the Geboes Score. After the conventions were applied, the ICC for the Modified Riley Score increased to 0.85 (95% CI 0.77 to 0.91) and 0.88 (95% CI 0.79 to 0.93) for the Geboes Score. Mosli 2017 also evaluated the intra-rater reliability of the Roberts Histopathology Index, and determined the intra-rater ICC to be 0.82 (95% CI 0.74 to 0.86). The intra-rater reliability of the Nancy Index was examined in Marchal-Bressenot 2017, and the ICC was found to be 0.88 (95% CI 0.82 to 0.92) (Table 3).

All eight scoring indices that underwent reliability testing were assessed for inter-rater reliability (i.e. agreement between different individuals). Estimates of correlation ranged from 'moderate' to 'almost perfect'. The Modified Riley Score was initially found to have

an ICC of 0.48 (95% CI 0.35 to 0.66), but after scoring conventions were applied in Mosli 2017, agreement improved to 0.80 (95% CI 0.69 to 0.87). For the Harpaz Index, Fiel 2003 reported a kappa statistic (κ) of 0.90 (95% CI not reported). Geboes 2000 evaluated inter-rater agreement of the Geboes Index using three raters, and reported $\kappa = 0.59, 0.62$ and 0.70 (95% CIs not reported). Likewise, a reliability study conducted by Jauregui-Amezaga 2016 found the Geboes Score to have ICCs of 0.60 and 0.50 (95% CIs not reported). Prior to introducing scoring conventions, Mosli 2017 determined the inter-rater ICC to be 0.82 (95% CI 0.73 to 0.88) for the Geboes Score, however, after the conventions were applied, agreement was 0.79 (95% CI 0.63 to 0.87). Inter-rater reliability, as measured by ICCs, was found to be 0.86 (95% CI 0.81 to 0.99) for the Nancy Index and 0.92 (95% CI 0.88 to 0.94) for the Roberts Histopathology Index (Marchal-Bressenot 2017; Mosli 2017). For the Riley Index, the estimate of agreement was 0.94 (95% CI 0.90 to 0.98), as measured by κ (Riley 1991). Inter-rater reliability of the Modified Harpaz Index was expressed as having varied by 4.35% between the two pathologists who scored the biopsies (Theede 2015).

Validity

Content validity

Three of the scoring indices (the Geboes Index, Nancy Index and Roberts Histopathology Index) have undergone content validation. In Geboes 2000, item selection and index development was based on a review of the existing literature. With respect to the Nancy Index, the inter-rater and intra-rater reliability of individual items derived from pre-existing scores (the Geboes Score, Riley Score, Gramlich Index, Gupta Index, and Global Visual Evaluation) was calculated (Marchal-Bressenot 2017). Items with high reliability were incorporated into the new index, with novel items identified through literature searching and expert opinion panels also included. For the development of the Roberts Histopathology Index, items from pre-existing scores (the Geboes Score, Modified Riley Score and Visual Analogue Scale) were included if intra-rater and inter-rater reliability was rated as high (Mosli 2017). A consensus process took place after item selection in order to standardize definitions and scoring rules for items with relatively higher rates of disagreement (Table 4).

Criterion validity

Three of the included scoring indices (the Gomes Index, The Harpaz Index, and Riley Score) have undergone criterion validation (Gomes

1986; Riley 1991; Theede 2015) (Table 5). The correlation estimates for this operating property range from 'slight' to 'moderate'.

Albumen concentration

The correlation coefficient (r) for the relationship between the Gomes Index and albumen concentration was 0.03 (no P value reported).

C-reactive protein (CRP)

The estimate of correlation for the Gomes Index and CRP was $r = 0.01$ (no P value reported), while the correlation between the Riley Score and CRP was $r = 0.31$ ($P < 0.0001$).

Erythrocyte sedimentation rate (ESR)

For the Gomes Index, the estimate of correlation with ESR was $r = 0.13$ (no P value reported).

Fecal calprotectin

The correlation between the Riley Score and fecal calprotectin was $r = 0.46$ ($P < 0.0001$).

Higher concentrations of fecal calprotectin were observed in patients with mild histological activity compared to those with histological remission (236.5 mg/kg versus 56 mg/kg, $P = 0.02$). A cutoff level of 171 mg/kg predicted histological healing, with positive predictive value of 0.75 and negative predictive value of 0.90.

Fecal lactoferrin

The estimate of correlation between the Riley Score and fecal lactoferrin was $r = 0.41$ ($P < 0.0001$).

Platelets

The estimate of correlation between platelets and the Gomes Index was $r = 0.13$ (no P value reported).

PMN-elastase

PMN-elastase was estimated to have a correlation of $r = 0.45$ ($P < 0.0001$) with the Riley Score.

White blood cell (WBC) count

The estimated correlation between the Gomes Index and WBC count and the Riley Score and WBC count was $r = 0.13$ (no P value reported) and $r = 0.20$ ($P < 0.02$), respectively.

Construct validity

A total of six scoring indices have undergone construct validation (Geboes 2000; Marchal-Bressenot 2017; Mosli 2017; Rubin 2007; Theede 2015; Truelove 1956). The correlation estimates between the histologic scoring indices and other measures of disease activity (i.e. clinical and endoscopic disease activity indices) ranged from 'moderate' to 'almost perfect'. Geboes 2000 reported a correlation of $r = 0.48$ ($P < 0.0001$) between the Geboes Score and the Mayo Clinic Endoscopic Subscore. The Nancy Index was compared to two indices: the Global Visual Evaluation and the Geboes Score. Correlations of $r = 0.88$ (95% CI 0.82 to 0.91), 0.82 (95% CI 0.74 to 0.87) and 0.87 (95% CI 0.82 to 0.91) were reported for the former

and $r = 0.90$ (95% CI 0.85 to 0.93), 0.84 (95% CI 0.77 to 0.89) and 0.94 (95% CI 0.91 to 0.96) for the latter (Marchal-Bressenot 2017). The Roberts Histopathology Index was compared to the Visual Analogue Scale, Modified Baron Score, Ulcerative Colitis Clinical Score and Inflammatory Bowel Disease Questionnaire, with respective correlations of $r = 0.91$, 0.61, 0.62 and -0.48 observed (Mosli 2017). The Chicago Index was compared to a clinical index (the Ulcerative Colitis Endoscopic Index of Severity) and an endoscopic index (the Mayo Endoscopic Subscore). Correlations of $r = 0.51$ ($P < 0.0001$) and $r = 0.60$ ($P < 0.0001$) were calculated for the former and latter (Rubin 2007). Two endoscopic scores were compared to the Modified Harpaz Index: the Ulcerative Colitis Endoscopic Index of Severity and the Mayo Clinic Endoscopic Subscore. Estimates of correlation (as measured by Kendall's tau (τ)) were $\tau = 0.63$ ($P < 0.0001$) and $\tau = 0.67$ ($P < 0.0001$), respectively. Finally, the estimate of correlation between the Truelove and Richards Index and the Simple Colitis Clinical Activity Index was $\kappa = 0.47$ and the Truelove and Richards Index and Baron Score was $\kappa = 0.58$ (Truelove 1956) (Table 6).

Responsiveness

The Nancy Index and the Roberts Histopathology Index are the only two histologic indices that have been subject to responsiveness testing (Marchal-Bressenot 2017; Mosli 2017). In Marchal-Bressenot 2017, responsiveness of the Nancy Index was assessed by retrospectively evaluating paired biopsies taken from 30 patients in which consecutive endoscopies were performed and histologic disease activity had changed (as defined by > 1 point on the Geboes conversion 9 scale). The median time between the two biopsies was 451 days, and one central reader scored all 60 biopsies. The mean change and standard deviation (SD) between the two biopsy scores was calculated using the Nancy Index, in addition to the Geboes Score and Global Visual Evaluation. The relationship between the change score of the Nancy Index and the Geboes Score and the Nancy Index and the Global Visual Evaluation was assessed using Pearson's correlation co-efficient and the 95% CI. The mean change in Nancy Index, Geboes Score and Global Visual Evaluation was -2.53 (SD 1.10), -15.86 (SD 7.26) and -4.83 (SD 2.13), respectively. The estimate of correlation between the Nancy Index and the Geboes Score was $r = 0.91$ (95% CI 0.81 to 0.96) and the Nancy Index and Global Visual Evaluation was $r = 0.89$ (95% CI 0.77 to 0.94).

In Mosli 2017, responsiveness of the Roberts Histopathology Index was evaluated using data from an RCT of a treatment of known efficacy (MLN02). One central reader evaluated 154 pairs of biopsies, taken before treatment and post-treatment (week 4 or 6). Change score correlations and the standardized effect size and Guyatt's responsiveness statistic (using: (1) patients assigned to MLN02 considered 'changed' and 'unchanged'; (2) an absolute change in endoscopic disease activity, as assessed by > 1 point on the Modified Baron Scale; (3) an absolute change > 2 points in the Mayo Clinic Score rectal bleeding score combined with the stool frequency score) were calculated for the Roberts Histopathology Index, the Geboes Score and the Modified Riley Score. The estimate of correlation between the change in Roberts Histopathology Index and change in Geboes Score, and the change in Roberts Histopathology index and change in Modified Riley Score, was estimated to be 0.75 (95% CI 0.67 to 0.82) and 0.84 (95% CI 0.79 to 0.88), respectively. With respect to (1) patients assigned to MLN02 considered 'changed' and 'unchanged'; (2) an absolute change in endoscopic disease activity, as assessed by > 1 point on the Modified Baron Scale; (3) an absolute change > 2 points in the

Mayo Clinic Score rectal bleeding score combined with the stool frequency score, the standardized effect size was 1.05 (95% CI 0.79 to 1.3), 0.81 (95% CI 0.58 to 1.05) and 1.05 (95% CI 0.78 to 1.31), respectively. Guyatt's responsiveness statistic for (1) patients assigned to MLN02 considered 'changed' and 'unchanged'; (2) an absolute change in endoscopic disease activity, as assessed by > 1 point on the Modified Baron Scale; (3) an absolute change > 2 points in the Mayo Clinic Score rectal bleeding score combined with the stool frequency score was 0.88 (95% CI 0.64 to 1.12), 0.73 (95% CI 0.50 to 0.96) and 0.84 (95% CI 0.59 to 1.09), respectively (Table 7).

Feasibility

None of the included studies assessed feasibility.

Methodological quality

The methodological quality of the included studies was assessed using the COSMIN tool (Table 8; Table 9). Of the eight studies that evaluated reliability, five studies (Feagan 2005; Geboes 2000; Marchal-Bressenot 2017; Mosli 2017; Theede 2015) were rated as 'excellent', one (Riley 1991) was rated as 'good' and one (Jauregui-Amezaga 2016) was rated as 'fair'.

With regard to content validity, all three of the studies (Geboes 2000; Marchal-Bressenot 2017; Mosli 2017) that assessed this property were rated as 'excellent'. The six studies that assessed construct validity focused on hypothesis testing. Of these studies, five were rated as 'excellent' (Geboes 2000; Marchal-Bressenot 2017; Mosli 2017; Theede 2015; Truelove 1956), and one was rated as 'good' (Rubin 2007). Of the two studies that assessed criterion validity, one was rated 'excellent' (Gomes 1986), and one was rated as 'good' (Riley 1991). Two studies explored responsiveness: Marchal-Bressenot 2017 (the Nancy Index) and Mosli 2017 (the Roberts Histopathology Index). Both were rated as 'excellent'. None of the eleven studies assessed internal consistency, measurement error, structural validity, cross-cultural validity and interpretability and therefore these measurement properties could not be assessed using COSMIN.

DISCUSSION

Summary of main results

A total of 126 reports describing 30 scoring indices were identified by the search strategy and screening process. Eleven of these scoring indices have been partially validated and 19 have not undergone any form of validation testing. Correlation estimates of intra-rater reliability for eight of the scoring indices range from 'substantial' to 'almost perfect'. Inter-rater reliability has been assessed for all 11 of the partially validated indices, with correlation estimates ranging from 'moderate' to 'almost perfect'. Three of the included scoring indices, the Geboes Score, Nancy Index and Roberts Histopathology Index, used literature review and expert opinion for index development and therefore integrated content validation. Two of the included indices, the Gomes Index and the Riley Index, assessed criterion validity by calculating correlation estimates between the index in question and various biomarkers (CRP, ESR, WBC count, platelets, albumen concentration, fecal lactoferrin, fecal calprotectin and PMN-elastase). A total of six of the included scoring indices explored construct validity by comparing the index in question to other measures of disease activity (histologic, clinical or endoscopic). Two of the included

studies, the Nancy Index and the Roberts Histopathology Index, measured responsiveness. Responsiveness testing of the Nancy Index consisted of an expert pathologist scoring two sets of 30 biopsy specimens that had undergone histologic change (as defined by the Geboes Score) and calculating change scores and correlation estimates for the Nancy Index, Global Visual Evaluation and Geboes Score. The Roberts Histopathology Index was evaluated for responsiveness by having an expert pathologist score two sets of 154 biopsy specimens (baseline and post-treatment), from patients who had received MLN02 as part of an RCT. The pathologist also scored the biopsy specimens using the Geboes Score and Modified Riley Score. Correlation estimates between the change scores were calculated, in addition to the standard effect size and Guyatt's responsiveness statistic for (1) patients assigned to MLN02 considered 'changed' and 'unchanged'; (2) an absolute change in endoscopic disease activity, as assessed by > 1 point on the Modified Baron Scale; (3) an absolute change > 2 points in the Mayo Clinic Score rectal bleeding score combined with the stool frequency score.

Overall completeness and applicability of evidence

The Nancy Index and the Roberts Histopathology Index have undergone the most validation in that four (reliability, content validity, construct validity (hypothesis testing) and criterion validity) operating properties have been tested. However, none of the currently available histologic scoring indices have been fully validated.

Quality of the evidence

The COSMIN tool was used to assess the methodological quality of the included studies (Table 8). The eleven included studies received scores ranging from 'fair' to 'excellent' with respect to the 10 operating properties assessed by this tool.

Potential biases in the review process

Sample quality is an important potential confounder that is infrequently mentioned or assessed in validation studies of histologic scoring indices. In Geboes 2000, the researchers found that out of the 99 biopsies studied, 31, 36 and 22 of the samples were rated as good, substandard and poor quality, respectively. As a result, 13 of the 22 poor quality samples were omitted from this study. While it is possible to remove poor quality specimens from a validation study, it was not always clear whether this quality check and exclusion of poor quality specimens was performed in the studies included in this systematic review.

Further research is needed to determine the best way to collect and process biopsy samples in order to facilitate both validation studies and the use of histologic disease activity indices in clinical practice and clinical trials.

Agreements and disagreements with other studies or reviews

A systematic review conducted by Byrant et al. identified four UC histologic scoring indices that have been partially validated: the Truelove and Richards Index, the Riley Score, the Geboes Score and the Harpaz Index (Bryant 2014). The current systematic review identified an additional seven partially validated histologic indices: the Modified Riley Score, the Gomes Index, the Simplified Geboes

Score, the Nancy Index, the Robarts Histopathology Index, the Chicago Index and the Modified Harpaz Index.

AUTHORS' CONCLUSIONS

Implication for methodological research

While several of the histologic indices identified have undergone several aspects of validation testing, none of the measures

identified by this systematic review are fully validated. In order to determine the optimal endpoint for histologic healing in UC, more research is required.

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

Feagan 2005 {published data only}

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.

Fiel 2003 {published data only}

Bouguen G, Levesque BG, Pola S, Evans E, Sandborn WJ. Feasibility of endoscopic assessment and treating to target to achieve mucosal healing in ulcerative colitis. *Inflammatory Bowel Diseases* 2014;**20**(2):231-9.

* Fiel M, Qin L, Suriawinita A, Xu R, Qui L, Bitar M, et al. Histologic grading of disease activity in chronic IBD: inter- and intra-observer variation among pathologists with different levels of experience. *Modern Pathology* 2003;**83**(1):118A.

Gasia MF, Gui X, Panaccione R, Kaplan G, Ghosh S, Lacucci M. Clinical outcome of ulcerative colitis patients with complete mucosal healing but with subtle abnormalities detected by novel iSCAN endoscopic. *Gastroenterology* 2015;**148**(4):S826.

Gasia MF, Panaccione R, Kaplan G, Ghosh S, Gui X, Lacucci M. Clinical outcome of ulcerative colitis with mucosal healing demonstrated by white light endoscopy but with subtle abnormalities detected by high definition iSCAN endoscopy. *Journal of Crohn's and Colitis* 2015;**9**:S164.

Gupta RB, Harpaz N, Itzkowitz S, Hossain S, Matula S, Kornbluth A, et al. Histologic inflammation is a risk factor for progression to colorectal neoplasia in ulcerative colitis: a cohort study. *Gastroenterology* 2007;**133**(4):1099-105.

Hefti MM, Chessin DB, Harpaz N, Steinhagen RM, Ullman TA. Severity of inflammation as a predictor of colectomy in patients with chronic ulcerative colitis. *Diseases of the Colon and Rectum* 2009;**52**(2):193-7.

Iacucci M, Gasia MF, Hassan C, Panaccione R, Kaplan GG, Ghosh S, et al. Complete mucosal healing defined by endoscopic Mayo subscore still demonstrates abnormalities by novel high definition colonoscopy and refined histological gradings. *Endoscopy* 2015;**47**(8):726-34.

Iacucci M, Gasia MF, Gui X, Panaccione R, Kaplan G, Love J, et al. iSCAN-High definition colonoscopy correlates with white light endoscopy and histology assessment in mucosal healing for ulcerative colitis. *Journal of Crohn's and Colitis* 2013;**7**(Supplement 1):S53.

Geboes 2000 {published data only}

Botina A, Shchukina O, Kondrashina E, Kharitidis A, Markova E. Histological examination of colonic mucosa from patients with clinical remission of ulcerative colitis. *Virchows Archiv: an International Journal of Pathology* 2015;**467**:S192.

Dargavel C, Kevans D, Kabakchiev B, Riddell RH, Kirsch R, Silverberg MS. Predictive value of basal plasmacytosis on

clinical relapse in quiescent ulcerative colitis. *Gastroenterology* 2015;**148**(4):S115.

David R, De Hertogh G, Jerrold T, Hong W, Ben A, Geert D. Histologic activity after induction therapy with multimatrix mesalamine is associated with 12-month remission status in ulcerative colitis. *Inflammatory Bowel Diseases* 2014;**20**:S69.

Di Nardo G, Oliva S, Aloï M, Conte F, Viola F, Nuti F, et al. Effectiveness of a rectal infusion of *Lactobacillus Reuteri* ATCC 55730 in children with distal active ulcerative colitis. *Digestive and Liver Disease* 2009;**41**:S233.

Farkas K, Reisz Z, Sejben A, Tiszlavicz L, Szucs M, Szepes, et al. Correlation of histological activity and basal plasmacytosis with mucosal healing in ulcerative colitis patients. *Journal of Crohn's and Colitis* 2015;**9**:S183.

Fernandez-Blanco I, Taxonera C, Cara C, Fernandez-Diaz G. Endoscopic and histologic remission correlation with biomarkers in UC patients treated with adalimumab. *Gastroenterology* 2015;**1**:S791.

Fron C, Belleannee G, Chabrun E, Poullenot F, Subtil C, Zerbib F, et al. Mucosal and histologic response to cyclosporin in ulcerative colitis: a cohort study. *Gastroenterology* 2015;**1**:S863.

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

Jauregui-Amezaga A, De Hertogh G, Bessissow T, Lemmens B, Lobaton T, Ferrante M, et al. A simplified histological Geboes score for ulcerative colitis. *Gastroenterology* 2015;**148**(4):S439-40.

Jauregui-Amezaga A, De Hertogh G, Bessissow T, Lemmens B, Lobaton T, Ferrante M, et al. Development of a simplified histological Geboes score for ulcerative colitis. *Journal of Crohn's and Colitis* 2015;**9**:S22-3.

Jauregui-Amezaga A, De Hertogh G, Bessissow T, Lemmens B, Lobaton T, Ferrante M, et al. A simplified histological Geboes score for ulcerative colitis. *Gastroenterology* 2015;**1**:S439-40.

Kim DB, Lee KM, Lee JM, Chung YY, Sung HJ, Paik, C. N, et al. Correlation between histologic activity and clinical, endoscopic, and serologic activities in patients with ulcerative colitis. *Journal of Gastroenterology and Hepatology* 2014;**29**:127-8.

Knyazev O, Kagramanova A, Churikova A, Konoplyannikov A, Khomeriki S, Parfenov A, et al. The use of mesenchymal stromal cells in order to achieve deep (biological) remission of ulcerative colitis. *Journal of Crohn's & Colitis* 2015;**9**:S367-8.

Kondrashina E, Schukina O, Kharitidis A, Botina A, Markova E. Evaluation of histological parameters in patients with clinical remission of ulcerative colitis. *Journal of Crohn's & Colitis* 2015;**9**:S214.

Lemmens B, Arijs I, Van Assche GA, Rutgeerts PJ, Geboes K, Vermeire S, et al. Correlation between the endoscopic and

histological score in assessing the activity of ulcerative colitis. *Gastroenterology* 2011;**140**(5):S421.

Lobaton 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.

Mayer L, Sandborn WJ, Stepanov Y, Geboes K, Hardi R, Yellin M, et al. Anti-IP-10 antibody (BMS-936557) for ulcerative colitis: a phase II randomised study. *Gut* 2014;**63**(3):442-50.

Mooiweer E, Severs M, Schipper ME, Fidder HH, Siersema PD, Laheij R, et al. Low fecal calprotectin predicts sustained clinical remission in inflammatory bowel disease patients: a plea for deep remission. *Journal of Crohn's and Colitis* 2015;**9**(1):50-5.

O'Byrne S, Keir ME, Cabanski CR, Zhao R, Colombel JF, Panes J, et al. Rectal bleeding accurately reflects level of mucosal inflammation in patients with ulcerative colitis. *Gastroenterology* 2015;**148**(4):S442.

Popp C, Staniceanu F, Micu G, Nichita L, Cioplea M, Mateescu R, et al. Evaluation of histologic features with potential prognostic value in ulcerative colitis. *Romanian Journal of Internal Medicine* 2014;**52**(4):256-62.

Rosenberg L, Lawlor Go, Zenlea T, Goldsmith JD, Gifford A, Falchuk KR, et al. Predictors of endoscopic inflammation in patients with ulcerative colitis in clinical remission. *Inflammatory Bowel Diseases* 2013;**19**(4):779-84.

Rosenberg L, Nanda KS, Zenlea T, Gifford A, Lawlor GO, Falchuk KR, et al. Histologic markers of inflammation in patients with ulcerative colitis in clinical remission. *Clinical Gastroenterology and Hepatology* 2013;**11**(8):991-6.

Staniceanu F, Popp C, Micu G, Neacsu C, Mageriu V, Stoica A, et al. Prognostic value of some histologic parameters in ulcerative colitis-is there hope for a reliable marker?. *Virchows Archiv* 2014;**465**(Supp 1):S232.

Travis S, Yap LM, Hawkey C, Warren B, Lazarov M, Fong T, et al. RDP58 is a novel and potentially effective oral therapy for ulcerative colitis. *Inflammatory Bowel Diseases* 2005;**11**(8):713-9.

Tursi A, Elisei W, Picchio M, Forti G, Penna A, Inchingolo CD, et al. Histological inflammation in ulcerative colitis in deep remission under treatment with infliximab. *Clinics and Research in Hepatology and Gastroenterology* 2015;**39**(1):107-13.

Wiernicka A, Szymanska S, Cielecka-Kuszyk J, Dadalski M, Kierkus J. Histological healing after infliximab induction therapy in children with ulcerative colitis. *World Journal of Gastroenterology* 2015;**21**(37):10654-61.

Zittan E, Kelly O, Kirsch R, Burns J, Stempak JM, Nguyen GC, et al. Low fecal calprotectin correlates with histological remission and mucosal healing in ulcerative colitis and colonic Crohn's disease. *Gastroenterology* 2015;**148**(4 Supp 1):S441.

Gomes 1986 {published data only}

Bennink R, Peeters M, D'Haens G, Rutgeerts P, Mortelmans L. Tc-99m HMPAO white blood cell scintigraphy in the assessment of the extent and severity of an acute exacerbation of ulcerative colitis. *Clinical Nuclear Medicine* 2001;**26**(2):99-104.

Bennink RJ, Peeters M, Rutgeerts P, Mortelmans L. Evaluation of early treatment response and predicting the need for colectomy in active ulcerative colitis with 99mTc-HMPAO white blood cell scintigraphy. *Journal of Nuclear Medicine* 2004;**45**(10):1698-704.

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

Jauregui-Amezaga 2016 {published data only}

Jauregui-Amezaga A, Das Y, Lobatón T, Lemmens B, Bessissow T, Ferrante M, et al. Validation of the Simplified Geboes Score for Ulcerative Colitis. *Journal of Crohn's and Colitis* 2016;**10**:S144.

Marchal-Bressenot 2017 {published data only}

Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, et al. Development and validation of the Nancy histological index for UC. *Gut* 2017;**66**(1):43-9.

Mosli 2017 {published data only}

Mosli MH, Feagan BG, Zou G, Geboes K, Langner C, Peterson MR, et al. Agreement of central readers in the evaluation of histopathological disease activity in ulcerative colitis. *Gastroenterology* 2014;**5**(1):S-225.

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

Mosli MH, Feagan BG, Zou G, Sandborn WJ, D'Haens G, Khanna R, et al. Reproducibility of histological assessments of disease activity in UC. *Gut* 2015;**64**(11):1765-73.

Riley 1991 {published data only}

Ando T, Watanabe O, Furuta R, Maeda O, Nishio Y, Nishiwaki T, et al. Predictors of a response to cyclosporine or leukocyte removal therapy in patients with refractory ulcerative colitis. *Digestive Endoscopy* 2005;**17**(2):153-8.

Fukunaga K, Ohda Y, Hida N, Iimuro M, Yokoyama Y, Kamikozuru K, et al. Placebo controlled evaluation of Xilei San, a herbal preparation in patients with intractable ulcerative proctitis. *Journal of Gastroenterology and Hepatology* 2012;**27**(12):1808-15.

Kruis W, Kiudelis G, Racz I, Gorelov IA, Pokrotnieks J, Horynski M, et al. Once daily versus three times daily mesalazine granules in active ulcerative colitis: a double-blind, double-dummy, randomised, non-inferiority trial. *Gut* 2009;**58**(2):233-40.

Langhorst J, Rueffer A, Dobos GJ, Boone JH. Monitoring of histologically defined healing in ulcerative colitis: performance of fecal lactoferrin, calprotectin, PMN-elastase and CRP and WBC compared to the Riley score. *Gastroenterology* 2015;**148**(4):S448.

Lemmens B, Arijis I, Van Assche G, Sagaert X, Geboes K, Ferrante M, et al. Correlation between the endoscopic and histologic score in assessing the activity of ulcerative colitis. *Inflammatory Bowel Diseases* 2013;**19**(6):1194-201.

Lemmens B, Arijis I, Van Assche GA, Rutgeerts PJ, Geboes K, Vermeire S, et al. Correlation between the endoscopic and histological score in assessing the activity of ulcerative colitis. *Gastroenterology* 2011;**1**:S421-2.

Matsumoto T, Kuroki F, Mizuno M, Nakamura S, Iida M, et al. Application of magnifying chromoscopy for the assessment of severity in patients with mild to moderate ulcerative colitis. *Gastrointestinal Endoscopy* 1997;**46**(5):400-5.

Nomura O, Osada T, Shibuya T, Sakamoto N, Ishikawa D, Asaoka D, et al. Predictive ability of relapse in patients with resolved ulcerative colitis using autofluorescence imaging endoscopy. *Gastrointestinal Endoscopy* 2014;**1**:AB465.

* Riley SA, Mani V, Goodman MJ, Dutt S, Herd ME. Microscopic activity in ulcerative colitis: what does it mean?. *Gut* 1991;**32**(2):174-8.

Riley SA, Mani V, Goodman MJ, Herd ME, Dutt S, Turnberg LA. Comparison of delayed-release 5-aminosalicylic acid (mesalazine) and sulfasalazine as maintenance treatment for patients with ulcerative colitis. *Gastroenterology* 1988;**94**(6):1383-9.

Rutka M, Milassin A, Szepes Z, Szucs M, Nyari T, Balint A, et al. Is mucosal healing more common than clinical remission in ulcerative colitis?—Is it the truth or only a myth coming from the studies?. *Scandinavian Journal of Gastroenterology* 2015;**50**(8):985-990.

Rubin 2007 {published data only}

Mikolajczyk AE, Watson S, Ackerman MT, Goeppinger SR, Hart J, Turner JR, et al. Assessment of the degree of variation of histologic inflammation in ulcerative colitis. *Gastroenterology* 2014;**5**(146):S232.

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* 2012;**107**:S694.

Rubin DT, Cohen RD, Sandborn WJ, Lichtenstein GR, Axler J, Riddell R, et al. Budesonide MMX 9 mg for inducing remission in patients with mild-to-moderate ulcerative colitis not adequately controlled with oral 5-ASAs. *Journal of Crohn's & Colitis* 2015;**9**(Suppl 1):S7.

* Rubin DT, Huo D, Hetzel JT, Bunnag AP, Sedrak M, Hart JA, et al. Increased degree of histological inflammation predicts colectomy and hospitalisation in patients with ulcerative colitis. *Gastroenterology* 2007;**132**(4):A19.

Rubin DT, Huo D, Kinnucan JA, Sedrak MS, McCullom NE, Bunnag AP, et al. Inflammation is an independent risk factor for colonic neoplasia in patients with ulcerative colitis: a case-control study. *Clinical Gastroenterology and Hepatology* 2013;**11**(12):1601-8.

Theede 2015 {published data only}

Theede K, Holck S, Ibsen P, Ladelund S, Nordgaard-Lassen I, Nielsen AM. Level of fecal calprotectin correlates With endoscopic and histologic inflammation and identifies patients with mucosal healing in ulcerative colitis. *Clinical Gastroenterology and Hepatology* 2015;**13**(11):1929-36.

Truelove 1956 {published data only}

Ardizzone S, Petrillo M, Antonacci CM, Bianchi Porro G. Sucralfate and hydrocortisone enemas in the treatment of active ulcerative proctitis—a randomized single-blind comparative study. *Alimentary Pharmacology and Therapeutics* 1996;**10**(6):957-60.

Ardizzone S, Petrillo M, Imbesi V, Cerutti R, Bollani S, Bianchi Porro G. Is maintenance therapy always necessary for patients with ulcerative colitis in remission?. *Alimentary Pharmacology and Therapeutics* 1999;**13**(3):373-9.

Bayles T, Sninsky C. Budesonide enema is an effective alternative to hydrocortisone enema in active distal ulcerative colitis. *Gastroenterology* 1995;**108**(4 Suppl 2):S778.

Campieri M, Cottone M, Miglio F, Manenti F, Astegiano M, D'Arienzo A, et al. Beclomethasone dipropionate enemas versus prednisolone sodium phosphate enemas in the treatment of distal ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 1998;**12**(4):361-6.

Campieri M, Gionchetti P, Belluzzi A, Brignola C, Tampieri M, Lannone P, et al. Optimum dosage of 5-aminosalicylic acid as rectal enemas in active ulcerative colitis (UC). *Gastroenterology* 1989;**96**(5 Part 2):A72.

Campieri M, Lanfranchi GA, Bertoni F, Brignola C, Bazzocchi G, Minguzzi M, et al. 4-aminosalicylic acid (4-ASA) and 5-aminosalicylic acid (5-ASA) in topical treatment of ulcerative colitis patients. *Gastroenterology* 1984;**86**(5 Part 2):A1039.

Chen QK, Yuan SZ, Wen ZF, Zhong YQ, Li CJ, Wu HS, et al. Characteristics and therapeutic efficacy of sulfasalazine in patients with mildly and moderately active ulcerative colitis. *World Journal of Gastroenterology* 2005;**11**(16):2462-6.

D'Albasio G, Paoluzi P, Campieri M, Porro GB, Pera A, Prantera C, et al. Maintenance treatment of ulcerative proctitis with mesalazine suppositories: a double-blind placebo-controlled trial. *American Journal of Gastroenterology* 1998;**93**(5):799-803.

Dong LF, Ouyang Q. The efficacy of azathiopurine in the treatment of ulcerative colitis. *Journal of Gastroenterology and Hepatology* 2011;**26**:134.

Feurle GE, Theuer D, Velasco S, Barry BA, Wordehoff D, Sommer A, et al. Olsalazine Versus Placebo in the Treatment of Mild to Moderate Ulcerative Colitis: A Randomized Double-Blind Trial. *Gastroenterology* 1988;**94**, No. 5, Part 2:A126.

Gionchetti P, Rizzello F, Venturi A, Ferretti M, Brignola C, Miglioli M, et al. Comparison of oral with rectal mesalazine in the treatment of ulcerative proctitis. *Diseases of the Colon and Rectum* 1998;**41**(1):93-7.

Green JT, Thomas GAO, Rhodes J, Williams GT, Evans BK, Russell MAH, et al. Nicotine enemas for active ulcerative colitis - A pilot study. *Alimentary Pharmacology and Therapeutics* 1997;**11**(5):859-63.

Ireland A, Mason CH, Jewell DP. Controlled trial comparing olsalazine and sulphasalazine for the maintenance treatment of ulcerative colitis. *Gut* 1988;**29**(6):835-7.

Kruis W, Schreiber S, Theuer D, Brandes JW, Schütz E, Howaldt S. Low dose balsalazide (1.5 g twice daily) and mesalazine (0.5 g three times daily) maintained remission of ulcerative colitis but high dose balsalazide (3.0 g twice daily) was superior in preventing relapses. *Gut* 2001;**49**(6):783-9.

Liang, H. L.Ouyang, Q. A clinical trial of combined use of rosiglitazone and 5-aminosalicylate for ulcerative colitis. *World Journal of Gastroenterology* 2008;**14**(1):114-9.

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

Mikhailova TL, Sishkova E, Poniewierka E, Zhidkov KP, Bakulin IG, Kupcinskis, L. Randomised clinical trial: the efficacy and safety of propionyl-L-carnitine therapy in patients with ulcerative colitis receiving stable oral treatment. *Alimentary Pharmacology and Therapeutics* 2011;**34**(9):1088-97.

Rizzello F, Gionchetti P, D'Arienzo A, Manguso F, Matteo G, Annese, V. Oral beclomethasone dipropionate in the treatment of active ulcerative colitis: a double-blind placebo-controlled study. *Alimentary Pharmacology and Therapeutics* 2002;**16**(6):1109-16.

Sinha A, Nightingdale JM, West KP. A gradual dose reduction is needed after 2 weeks of oral prednisolone treatment for ulcerative colitis. *Gastroenterology* 2000;**118**(4 Suppl 2):A782.

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* 2009;**58**:A101-2.

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

References to studies excluded from this review

Baars 2012 {published data only}

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

D'Argenio 2001 {published data only}

* D'Argenio G, Cosenza V, Riegler G, Della Valle N, Deritis F, Mazzacca G. Serum transglutaminase correlates with endoscopic and histopathologic grading in patients with ulcerative colitis. *Digestive Diseases and Sciences* 2001;**46**(3):649-57.

Girlich C, Schacherer D, Jung EM, Klebl F, Huber E. Comparison between quantitative assessment of bowel wall vascularization by contrast-enhanced ultrasound and results of histopathological scoring in ulcerative colitis. *International Journal of Colorectal Disease* 2012;**27**(2):193-8.

Scheppach W, Sommer H, Kirchner T, Paganelli GM, Bartram P, Christl S, et al. Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. *Gastroenterology* 1992;**103**(1):51-6.

Floren 1987 {published data only}

Floren CH, Benoni C, Willen R. Histologic and colonoscopic assessment of disease extension in ulcerative colitis. *Scandinavian Journal of Gastroenterology* 1987;**22**(4):459-62.

Friedman 1985 {published data only}

Friedman L, Richter J, Kirkham S. 5-aminosalicylic acid (5-ASA) enemas in refractory distal ulcerative colitis (UC): a randomized, controlled trial. *Gastroenterology* 1985;**88**:A1388.

Gramlich 2007 {published data only}

Bressenot A, Salleron J, Bastien C, Danese S, Boulagnon-Rombi C, Peyrin-Biroulet L. Comparing histological activity indexes in UC. *Gut* 2015;**64**(9):1412-8.

* Gramlich T, Petras RE. Pathology of inflammatory bowel disease. *Seminars in Pediatric Surgery* 2007;**16**(3):154-63.

Hanauer 1993 {published data only}

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.

Iacucci 2015 {published data only}

Gui X, Fort Gasia M, Ghosh S, Iacucci, M. A refined IBD histopathologic grading system (ECAP scheme)-more sensitive and informative in assessment of mucosal healing. *Laboratory Investigation* 2014;**94**:176-A.

* Iacucci M, Fort Gasia M, Hassan C, Panaccione R, Kaplan GG, Ghosh S, et al. Complete mucosal healing defined by endoscopic Mayo subscore still demonstrates abnormalities by novel high definition colonoscopy and refined histological gradings. *Endoscopy* 2015;**47**(8):726-34.

Keren 1984 {published data only}

Keren DF, Appelman HD, Dobbins WO III, Wells JJ, Whisenant B, Foley J, et al. Correlation of histopathologic evidence of disease activity with the presence of immunoglobulin containing cells in the colons of patients with inflammatory bowel disease. *Human Pathology* 1984;**15**(8):757-63.

Korelitz 1976 {published data only}

Korelitz BI, Sommers SC. Responses to drug therapy in ulcerative colitis. *American Journal of Digestive Diseases* 1976;**21**(6):441-7.

Matts 1961 {published data only}

Kato K, Mizuno S, Umesaki Y, Ishii Y, Sugitani M, Imaoka A, et al. Randomized placebo-controlled trial assessing the effect

of bifidobacteria-fermented milk on active ulcerative colitis. *Alimentary Pharmacology & Therapeutics* 2004;**20**(10):1133-41.

Kimura K, Kanai T, Bessho R, Hosoe N, Kobayashi T, Takayama T. In vivo visualization and evaluation of colorectal inflammation in ulcerative colitis by a newly integrated endocytoscopy. *Gastroenterology* 2011;**1**:S422.

Maeda Y, Kudo SE, Ohtsuka K, Wakamura K, Mori Y, Yamauchi A. First experience of endocytoscopic narrow-band imaging for evaluation of inflammatory activity in ulcerative colitis. *Gastrointestinal Endoscopy* 2013;**1**:AB442.

* Matts SG. The value of rectal biopsy in the diagnosis of ulcerative colitis. *Quarterly Journal of Medicine* 1961;**30**(4):393-407.

Moriichi K, Fujiya M, Utsumi T, Sakatani A, Tanaka K, Dokoshi T. Quantification of autofluorescence imaging is useful for objectively assessing the severity of ulcerative colitis. *Gastrointestinal Endoscopy* 2015;**81**(5):AB384-5.

Naganuma M, Yahagi N, Bessho R, Fujimoto A, Hosoe N, Kashiwagi K, et al. Evaluation of severity using endoscopic dual red imaging targeting submucosal vessel in patients with ulcerative colitis. *Gastroenterology* 2015;**148**(4):S579.

Nishiyama S, Tanaka S, Oka S, Shigita K, Asayama N, Sagami S, et al. Correlation between endocytoscopy and histopathological activity in the remission stage of ulcerative colitis: A pilot study. *Gastrointestinal Endoscopy* 2014;**79**(5):AB570-1.

Ohkusa T, Nomura T, Terai T, Miwa H, Kobayashi O, Hojo M, et al. Effectiveness of antibiotic combination therapy in patients with active ulcerative colitis: a randomized, controlled pilot trial with long-term follow-up. *Scandinavian Journal of Gastroenterology* 2005;**40**(11):1334-42.

Oliva S, Di Nardo G, Hassan C, Spada C, Aloï M, Ferrari F, et al. Second-generation colon capsule endoscopy vs colonoscopy in pediatric ulcerative colitis: A pilot study. *Endoscopy* 2014;**46**(6):485-92.

Tozawa K, Fukui H, Yamasaki T, Kondo T, Kono T, Toyoshima F, et al. Serum REG Ialpha is a marker to predict histological and clinical activity in ulcerative colitis. *Gastroenterology* 2015;**1**:S450.

Nishiyama 2014 {published data only}

Nishiyama S, Tanaka S, Oka S, Shigita K, Asayama N, Sagami S. Correlation between endocytoscopy and histopathological activity in the remission stage of ulcerative colitis: A pilot study. *Gastrointestinal Endoscopy* 2014;**79**(5):AB570-1.

Odze 1993 {published data only}

Odze R, Antonioli D, Peppercorn M, Goldman H. Effect of topical 5-aminosalicylic acid (5-ASA) therapy on rectal mucosal biopsy morphology in chronic ulcerative colitis. *American Journal of Surgical Pathology* 1993;**17**(9):869-75.

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.

Riley 1988 {published data only}

Riley SA, Mani V, Goodman MJ, Herd ME, Dutt S, Turnberg LA. Comparison of delayed-release 5-aminosalicylic acid (mesalazine) and sulfasalazine as maintenance treatment for patients with ulcerative colitis. *Gastroenterology* 1988;**94**(6):1383-9.

Rutter 2004 {published data only}

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.

Sandborn 1993 {published data only}

Pagnini C, Menasci F, Festa S, Rizzatti G, Corleto VD, Delle Fave M, et al. Application of clinical indexes in ulcerative colitis patients in regular follow-up visit: correlation with endoscopic 'mucosal healing' and implication for management. Preliminary results. *European Review for Medical and Pharmacological Sciences* 2015;**19**(19):3674-81.

Sandborn WJ, Tremaine WJ, Offord KP, Lawson GM, Petersen BT, Batts KP, et al. Transdermal nicotine for mildly to moderately active ulcerative colitis. A randomized, double-blind, placebo-controlled trial. *Annals of Internal Medicine* 1997;**126**(5):364-71.

* Sandborn WJ, Tremaine WJ, Schroeder KW, Steiner BL, Batts KP, Lawson GM. Cyclosporine enemas for treatment-resistant, mildly to moderately active, left-sided ulcerative colitis. *American Journal of Gastroenterology* 1993;**88**(5):640-5.

Saverymuttu 1986 {published data only}

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

Bunn SK, Bisset WM, Main MJ, Gray ES, Olson S, Golden BE. Fecal calprotectin: validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. *Journal of Pediatric Gastroenterology and Nutrition* 2001;**33**(1):14-22.

Celasco G, Papa A, Jones R, Moro L, Bozzella R, Surace M, et al. Clinical trial: oral colon-release parnaparin sodium tablets (CB-01-05 MMX) for active left-sided ulcerative colitis. *Alimentary pharmacology & therapeutics* 2010;**31**(3):375-86.

D'Haens GR, Kovacs A, Vergauwe P, Nagy F, Molnar T, Bouhnik Y, et al. Clinical trial: Preliminary efficacy and safety study of a new Budesonide-MMX(R) 9 mg extended-release tablets in patients with active left-sided ulcerative colitis. *Journal of Crohn's and Colitis* 2010;**4**(2):153-60.

- Dotan I, Hallak A, Arber N, Santo M, Alexandrowitz A, Knaani Y, et al. Low-dose low-molecular weight heparin (enoxaparin) is effective as adjuvant treatment in active ulcerative colitis an open trial. *Digestive Diseases and Sciences* 2001;**46**(10):2239-44.
- Langmead L, Feakins RM, Goldthorpe S, Holt H, Tsironi E, Silva A, et al. Randomized, double-blind, placebo-controlled trial of oral aloe vera gel for active ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 2004;**19**(7):739-47.
- Madsen SM, Schlichting P, Davidsen B, Nielsen OH, Federspiel B, Munkholm P. An open-labeled, randomized study comparing systemic interferon-alpha-2A and prednisolone enemas in the treatment of left-sided ulcerative colitis. *American Journal of Gastroenterology* 2001;**96**(6):1807-15.
- Mahmud N, McDonald GSA, Kelleher D, Weir DG. Microalbuminuria correlates with intestinal histopathological grading in patients with inflammatory bowel disease. *Gut* 1996;**38**(1):99-103.
- * Saverymuttu SH, Camilleri M, Rees H, Lavender JP, Hodgson HJ, Chadwick 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 Pt 1):1121-8.
- Watts 1966** {published data only}
- Almer S, Franzen L, Peters AM, Tjadermo M, Ekberg S, Granerus G, et al. Do technetium-99m hexamethylpropylene amine oxime-labeled leukocytes truly reflect the mucosal inflammation in patients with ulcerative colitis?. *Scandinavian Journal of Gastroenterology* 1992;**27**(12):1031-8.
- * Watts JM, Thompson H, Goligher JC. Sigmoidoscopy and cytology in the detection of microscopic disease of the rectal mucosa in ulcerative colitis. *Gut* 1966;**7**:288-94.
- Wiernicka 2015** {published data only}
- Szymanska S, Wiernicka A, Cielecka-Kuszyk J, Szymanska E, Dadalski M, Kierkus J. Influence of induction therapy with infliximab on histological changes in children with ulcerative colitis preliminary data. *Journal of Crohn's and Colitis* 2014;**8**:S249-S250.
- * Wiernicka A, Szymanska S, Cielecka-Kuszyk J, Dadalski M, Kierkus J. Histological healing after infliximab induction therapy in children with ulcerative colitis. *World Journal of Gastroenterology* 2015;**21**(37):10654-61.
- Additional references**
- Abraham 2009**
- Abraham C, Cho JH. Inflammatory bowel disease. *New England Journal of Medicine* 2009;**361**(2066):78.
- Ardizzone 2011**
- Ardizzone S, Cassinotti A, Duca P, Mazzali C, Penati C, Manes G, et al. Mucosal healing predicts late outcomes after the first course of corticosteroids for newly diagnosed ulcerative colitis. *Clinical Gastroenterology and Hepatology* 2011;**9**(6):483-9.
- Baumgart 2007**
- Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007;**369**(9573):1641-57.
- Bryant 2014**
- Bryant RV, Winer S, Travis SP, Riddell RH. Systematic review: histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative. *Journal of Crohns and Colitis* 2014;**8**(12):1582-97.
- Cohen 1988**
- Cohen J. Statistical power analysis for the behavioral sciences. 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
- Cohen 2002**
- Cohen RD. Evolving medical therapies for ulcerative colitis. *Current Gastroenterology Reports* 2002;**4**(6):497-505.
- 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):S142-58.
- Guyatt 1987**
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987;**40**(2):171-8.
- Landis 1977**
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**(1):159-74.
- Pineton de Chambrun 2010**
- Pineton de Chambrun G, Peyrin-Biroulet L, Lemann M, Colombel JF. Clinical implications of mucosal healing for the management of IBD. *Nature Reviews. Gastroenterology and Hepatology* 2010;**7**(1):15-29.
- Truelove 1955**
- Truelove SC, Witts LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *British Medical Journal* 1955;**2**(4947):1041-8.
- Zou 2005**
- Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. *Quality of Life Research* 2005;**14**(6):1545-52.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Feagan 2005

Methods	Retrospective data from a multicenter, randomized placebo controlled trial
Data	<p>N = 181</p> <p>Patient characteristics in the trial that introduced the score: adult patients with active disease defined as an Ulcerative Colitis Clinical Score of 5-9 points, with a score ≥ 1 on stool frequency or rectal bleeding, and a Modified Baron Score of ≥ 2 on sigmoidoscopic examination, with the disease ≥ 25 cm from the anal verge</p> <p>See Mosli 2017 for details on the study that assessed reliability</p>
Comparisons	No comparison was made with other histological indices
Outcomes	<p>Reliability testing</p> <p>See Table 3 for results</p>
Notes	<p>Scoring system assessed: Modified Riley Score</p> <p>Risk of bias assessment based on Mosli 2017 since this was the study in which reliability testing was conducted</p>

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Histologists who calculated the Modified Riley Scores were blinded to clinical and other lab data
Independent observation?	Yes	Histologists performed their assessments independently

Fiel 2003

Methods	Inter-observer and intra-observer reliability study of the Harpaz/Mount Sinai Index
Data	Kappa = 0.9 for intra-rater as well as inter-rater reliability
Comparisons	No comparison was made with other histological indices
Outcomes	<p>Reliability testing</p> <p>See Table 3 for results</p>
Notes	<p>Scoring system assessed: Harpaz/Mount Sinai Index</p> <p>Abstract publication</p> <p>Authors were contacted and supplied limited data; this review will be updated if new data from the study become available</p>

Risk of bias

Fiel 2003 (Continued)

Item	Authors' judgement	Description
Blinding?	Unclear	Abstract publication: details not reported
Independent observation?	Unclear	Abstract publication: details not reported

Geboes 2000

Methods	Index development and validation based on prospective, randomized controlled trial data	
Data	<p>Index was created based on literature review and expert opinion</p> <p>3 pathologists examined 99 biopsy slides prospectively obtained on 2 occasions from actively inflamed (n = 68) and quiescent (n = 31) colonic mucosa in patients with distal UC</p> <p>Inter-rater reliability was assessed twice; immediately after index development and after index modification</p> <p>In a second validation study, 263 biopsies from 131 UC patients that had previously been evaluated using the Mayo Clinic Endoscopic Subscore were scored using the Geboes and Riley systems</p>	
Comparisons	<p>Mayo Score (endoscopic activity)</p> <p>Riley Score (histologic activity)</p>	
Outcomes	<p>Reliability testing, content validation, construct validation</p> <p>See Table 3, Table 4 and Table 6 for results</p>	
Notes	<p>Scoring system assessed: Geboes Score</p> <p>Two phases: index development and index validation (reliability)</p> <p>Separate study assessed construct validity</p> <p>Inter-rater reliability measured with kappa in Geboes 2000; inter-rater reliability measured with ICC in Mosli 2017</p>	

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Readings were performed without knowledge of the endoscopic status
Independent observation?	Yes	Readings were performed independently

Gomes 1986

Methods	Index development and validation based on prospectively collected randomized controlled trial data	
Data	<p>Clinical, endoscopic and histologic data was collected from 28 patients with UC undergoing routine colonoscopy</p> <p>Evaluation performed by a single pathologist</p>	

Gomes 1986 (Continued)

Comparisons	C-reactive protein; erythrocyte sedimentation rate; white blood cell count; platelets; albumen concentration
Outcomes	Criterion validation See Table 5 for results
Notes	Scoring system assessed: Gomes Index

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Pathologist unaware of colonic appearance and clinical information
Independent observation?	Unclear	Not applicable (reliability testing not performed)

Jauregui-Amezaga 2016

Methods	Data collected prospectively from a tertiary referral centre
Data	Clinical, endoscopic, and histologic data were collected from 92 patients with UC Histologic activity was assessed with the Geboes Score and the Simplified Geboes Score by 2 trained readers (A and B) and an expert gastrointestinal pathologist (reader C)
Comparisons	No comparison was made with other histological indices
Outcomes	Reliability testing See Table 3 for results
Notes	Scoring system assessed: Simplified Geboes Score Reported in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Unclear whether the histologists who calculated the Geboes and Simplified Geboes Scores were blinded to clinical and other lab data
Independent observation?	Unclear	Not adequately described

Marchal-Bressenot 2017

Methods	Index development and validation study based on retrospective cohort data
Data	Index was created by measuring inter-rater and intra-rater ICCs of index items from pre-existing histologic indices and from expert opinion/literature review Data collected retrospectively from a cohort of patients (60 patients in the development phase; 100 patients in the validation phase)

Marchal-Bressenot 2017 (Continued)

Comparisons	Geboes Index Global Visual Evaluation
Outcomes	Reliability testing, content validation, construct validation, responsiveness testing See Table 3 , Table 4 , Table 6 and Table 7 for results
Notes	Scoring system assessed: Nancy Index Two phases: index development and validation First published as an e-pub in 2015

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	Histologists who calculated the Geboes, Riley, and Global Visual Evaluation Scores were blinded to clinical and other lab data
Independent observation?	Yes	Histologists performed their assessments independently

Mosli 2017

Methods	Index development and validation study based on retrospective randomized controlled trial data
Data	Index was created by measuring inter-rater and intra-rater ICCs of index items from pre-existing histologic indices For responsiveness and reliability testing, data from a randomized controlled trial of MLN02 (154/181 histology slides) were used (Feagan 2005) Reliability: (1) five pathologists assessed 49 biopsies three times, two weeks apart; (2) four pathologists assessed 50 slides on three occasions Responsiveness : one central reader pathologist assessed 154 paired slides (baseline and week 4 or week 6 post-treatment with MLN02)
Comparisons	Geboes Index Modified Riley Score Visual Analogue Scale Inflammatory Bowel Disease Questionnaire
Outcomes	Reliability testing, content validation, construct validation, responsiveness testing See Table 3 , Table 4 , Table 6 , and Table 7 for results
Notes	Scoring system assessed: Roberts Histopathology Index Two phases: index development and index validation First published as an e-pub in 2015

Risk of bias

Mosli 2017 (Continued)

Item	Authors' judgement	Description
Blinding?	Yes	Histologists who calculated the Modified Riley and Geboes Scores and Visual Analogue Scale were blinded to clinical and other lab data
Independent observation?	Yes	Histologists performed their assessments independently

Riley 1991

Methods	Data collected prospectively from a tertiary referral centre Inter-rater reliability and criterion validity testing of the Riley score
Data	82 patients in symptomatic and sigmoidoscopic remission were recruited A biopsy was taken at study entry and independently graded by two histopathologists In a separate study involving 150 UC patients, 253 endoscopies and 233 biopsies were performed to compare various biomarkers to histologic assessments
Comparisons	No comparison was made with other histological indices
Outcomes	Reliability testing; criterion validation See Table 3 and Table 5 for results
Notes	Scoring system assessed: Riley Score

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Unclear whether the histologists had knowledge of the patients
Independent observation?	Yes	Histopathologists graded independently

Rubin 2007

Methods	Validation study based on data collected prospectively from a tertiary referral centre
Data	86 UC patients undergoing endoscopy were included; endoscopic images and biopsies were obtained Endoscopic activity was assessed by the endoscopist Histologic activity was assessed by expert pathologists
Comparisons	Simple Colitis Clinical Activity Index (clinical) Mayo Clinic Endoscopic Subscore (endoscopic)
Outcomes	Construct validation See Table 6 for results
Notes	Scoring system assessed: Rubin/Chicago/Histologic Inflammation Activity Scale

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

Rubin 2007 (Continued)

Construct validation data published in abstract form only

Risk of bias

Item	Authors' judgement	Description
Blinding?	Unclear	Unclear whether the histologists who calculated the Rubin scores were blinded to clinical and other lab data
Independent observation?	Unclear	Not adequately described

Theede 2015

Methods	Prospective, cross-sectional validation study; data collected prospectively from a tertiary referral centre	
Data	120 patients with active or inactive UC who required endoscopy were included	
Comparisons	Mayo Clinic Endoscopic Subscore (endoscopic) Ulcerative Colitis Endoscopic Index of Severity (endoscopic)	
Outcomes	Inter-rater reliability, construct validation See Table 3 and Table 6 for results	
Notes	Scoring system assessed: Modified Harpaz Index In the case of diverse readings, which rarely occurred, consensus was reached by conference	

Risk of bias

Item	Authors' judgement	Description
Blinding?	Yes	The gastrointestinal pathologists who scored were unaware of any clinical information
Independent observation?	Unclear	Not adequately described

Truelove 1956

Methods	The original score was based on a prospective study of 111 serial biopsy specimens from 42 UC patients and 24 controls The validation study was based on data collected prospectively from a tertiary referral centre	
Data	Patients were clinically assessed at a tertiary referral centre by 4 gastroenterologists before endoscopy and mucosal biopsy were performed 2 pathologists graded the biopsies of 91 patients with active UC	
Comparisons	Simple Clinical Colitis Activity Index (clinical) Baron Score (endoscopic)	

Truelove 1956 *(Continued)*

Outcomes	Construct validation	
	See Table 6 for results	
Notes	Scoring system assessed: Truelove and Richards Index	
<i>Risk of bias</i>		
Item	Authors' judgement	Description
Blinding?	Yes	Assessors were blinded to clinical and endoscopic data
Independent observation?	Yes	All assessments were made independently

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

Study	Reason for exclusion
Baars 2012	Not a validation study
D'Argenio 2001	Not a validation study
Floren 1987	Not a validation study
Friedman 1985	Not a validation study
Gramlich 2007	Not a validation study
Hanauer 1993	Not a validation study
Iacucci 2015	Not a validation study
Keren 1984	Not a validation study
Korelitz 1976	Not a validation study
Mattis 1961	Not a validation study
Nishiyama 2014	Not a validation study
Odze 1993	Not a validation study
Powell-Tuck 1982	Not a validation study
Riley 1988	Not a validation study
Rutter 2004	Not a validation study
Sandborn 1993	Not a validation study
Saverymattu 1986	Not a validation study
Watts 1966	Not a validation study

Study	Reason for exclusion
Wiernicka 2015	Not a validation study

ADDITIONAL TABLES

Table 1. Indices that have been fully or partially validated

	Reference	Index
1	Feagan 2005	Modified Riley Score
2	Fiel 2003	Harpaz/Mount Sinai Index
3	Geboes 2000	Geboes Score
4	Gomes 1986	Gomes Index
5	Jauregui-Amezaga 2016	Simplified Geboes Score
6	Marchal-Bressenot 2017	Nancy Index
7	Mosli 2017	Robarts Histopathology Score
9	Riley 1991	Riley Score
9	Rubin 2007	Chicago/Rubin/Histologic Inflammation Activity Scale
10	Theede 2015	Modified Harpaz Index
11	Truelove 1956	Truelove and Richards Index

Table 2. Indices that have not been fully or partially validated

	Index	Reference
1	Baars Index	Baars 2012
2	British Society of Gastroenterology Protocol	Wiernicka 2015
3	D'Argenio/Scheppach Index	D'Argenio 2001
4	ECAP (Extent, Chronicity, Activity Plus additional findings) System	Iacucci 2015
5	Endocytoscopy System (ECS)	Nishiyama 2014
6	Floren Index	Floren 1987
7	Friedman Index	Friedman 1985

Table 2. Indices that have not been fully or partially validated *(Continued)*

8	Gramlich Score	Gramlich 2007
9	Hanauer Index	Hanauer 1993
10	Initial Riley Score	Riley 1988
11	Keren Score	Keren 1984
12	Korelitz Index	Korelitz 1976
13	Matts Score	Matts 1961
14	Odze Index	Odze 1993
15	Powell-Tuck Score	Powell-Tuck 1982
16	Rutter Score	Rutter 2004
17	Sandborn Index	Sandborn 1993
18	Saverymuttu Index	Saverymuttu 1986
19	Watts Score	Watts 1966

Table 3. Reliability

Study ID	Index	Inter-rater kappa (between raters)	Inter-rater ICC (between raters)	Intra-rater ICC (within rater)
Feagan 2005	Modified Riley Score		See Mosli 2017	See Mosli 2017
Fiel 2003	Harpaz/Mount Sinai Index	0.90		
Geboes 2000	Geboes Score	Three readers: 0.62, 0.70, 0.59; also see Jauregui-Amezaga 2016	See Mosli 2017	See Mosli 2017
Jauregui-Amezaga 2016	Simplified Geboes Score		Reader A vs C: 0.7 Reader B vs C: 0.7	
	Geboes Score		Reader A vs C: 0.6 Reader B vs C: 0.5	
Marchal-Bressenot 2017	Nancy index		0.86 (95% CI 0.81 to 0.99)	0.88 (95% CI 0.82 to 0.92)

Table 3. Reliability (Continued)

Mosli 2017	Robarts Histopathology Index	0.92 (95% CI 0.88 to 0.94)	0.82 (95% CI 0.74 to 0.86)
	Geboes Score	0.56 (95% CI 0.39 to 0.67)	0.82 (95% CI 0.73 to 0.88)
	Modified Riley Score	0.48 (95% CI 0.35 to 0.66)	0.71 (95% CI 0.63 to 0.80)
	Geboes Score**	0.79 (95% CI 0.63 to 0.87)	0.88 (95% CI 0.79 to 0.93)
	Modified Riley Score**	0.80 (95% CI 0.69 to 0.87)	0.85 (95% CI 0.77 to 0.91)
Riley 1991	Riley Score	0.94 (95% CI 0.90 to 0.98)	
Theede 2015	Modified Harpaz Index	4.35%*	

Abbreviations: ICC, intraclass correlation coefficient

*Inter-rater variation of 4.35% between the two pathologists evaluating the biopsies

**After scoring conventions applied

Table 4. Content Validity

Study ID	Index	Methods
Geboes 2000	Geboes Score	Items were included in the new index based a literature review
Marchal-Bressenot 2017	Nancy Index	Intra-rater and inter-rater reliability for index items of pre-existing histologic scores (the Geboes Score, The Riley Score, the Gramlich Index, the Gupta Index and the Global Visual Evaluation) were measured using ICCs. Items with high reliability were used in the Nancy Index Other items were included in the new index based on expert opinion and literature review
Mosli 2017	Robarts Histopathology Index	Intra-rater and inter-rater reliability for index items of pre-existing histologic scores (the Geboes Score, Modified Riley Score and a Visual Analogue Scale) were measured using intraclass correlation coefficients. Items with high reliability were used in the Robarts Histopathology Index. Consensus process was conducted after the item selection phase to standardize definitions/scoring rules for high disagreement items

Table 5. Criterion Validity

Study ID	Index	Outcome	Correlation coefficient (r)	Positive predictive value
				Negative predictive value
Gomes 1986	Gomes Index	Albumen concentration	0.028	
		C-reactive protein	0.01	

Table 5. Criterion Validity (Continued)

		Erythrocyte sedimentation rate	0.13	
		Platelets	0.13	
		White blood cell count	0.131	
Riley 1991	Riley Score	C-reactive protein	0.307 (p <0.0001)	
		Fecal lactoferrin	0.408 (p <0.0001)	
		Fecal calprotectin	0.458 (p <0.0001)	
		PMN-elastase	0.447 (p <0.0001)	
		White blood cell count	0.203 (p <0.02)	
Theede 2015	Modified Harpaz Index	Fecal calprotectin		A cutoff level of 171 mg/kg predicted histological healing, with positive predictive value of 0.75 and negative predictive value of 0.90.

Table 6. Construct Validity

Study ID	Index	Comparison	Correlation coefficient (r)	Kappa
Geboes 2000	Geboes Score	Mayo Clinic Endoscopic Sub-score	0.482* (p < 0.0001)	
Marchal-Bressenot 2017	Nancy Index	Global Visual Evaluation	Reader 1: 0.876 (95% CI 0.819-0.914)	
			Reader 2: 0.819 (95% 0.739 to 0.873)	
			Reader 3: 0.874 (95% CI 0.816 to 0.913)	
		Geboes Score	Reader 1: 0.899 (95% CI 0.853 to 0.931)	
			Reader 2: 0.843 (95% CI 0.773 to 0.891)	
			Reader 3: 0.939 (95% CI 0.909 to 0.958)	
Mosli 2017	Robarts Histopathology Index	Visual Analogue Scale	0.91 (predicted 0.90)	
		Modified Baron Score	0.61 (predicted 0.60)	
		Ulcerative Colitis Clinical Score	0.62 (predicted 0.40)	
		Inflammatory Bowel Disease Questionnaire	-0.48 (predicted -0.30)	

Table 6. Construct Validity (Continued)

Rubin 2007	Rubin/Chicago/Histological Activity Index	Simple Colitis Clinical Activity Index	0.508 (p < 0.0001)
		Mayo Clinic Endoscopic Sub-score	0.597 (p < 0.0001)
Theede 2015	Modified Harpaz Index	Ulcerative Colitis Endoscopic Index of Severity	0.63* (p < 0.0001)
		Mayo Clinic Endoscopic Sub-score	0.67* (p < 0.0001)
Truelove 1956	Truelove and Richards Index	Simple Colitis Clinical Activity Index	0.47
		Baron Score	0.58

*As measured by Kendall's tau

Table 7. Responsiveness

Study ID	Index	Treatment	Responsiveness Measure	Correlation
Marchal-Bressenot 2017	Nancy Index	No treatment of known efficacy used	Biopsy specimen pairs from 30 UC patients were retrospectively reviewed	a) Nancy Index and Global Visual Evaluation: 0.886 (95% CI 0.766 to 0.943)
			Median time between the two biopsies was 451 days (range: 41-1169 days)	
			a) Nancy Index	b) Nancy Index and Geboes Score: 0.910 (95% CI 0.813 to 0.955)
			Mean change (standard deviation (SD)): -2.53 (1.10)	
			b) Global Visual Evaluation	
			Mean change (SD): -4.83 (2.13)	
Mosli 2017	Robarts Histopathology Index	MLN02 (baseline and week 6 biopsies for 154/181 patients)	Biopsy specimen pairs from 154 UC patients (baseline and 4-6 weeks post-treatment) were retrospectively reviewed	a) Change in Robarts Histopathology Index and Geboes Score: 0.75 (95% CI 0.67 to 0.82)
			Correlation estimates for change in Robarts Histopathology Index, Geboes Score and Modified Riley Score	b) Change in Robarts Histopathology Index and Modified Riley Score: 0.84 (95% CI 0.79 to 0.88)
			Standardised effect size (SES) calculated using:	a) 1.05 (95% CI 0.79 to 1.3)*
			a) Treatment allocation with patients assigned to MLN02 considered 'changed' and	b) 0.81 (95% CI 0.58 to 1.05)*

Table 7. Responsiveness (Continued)

those assigned to placebo considered 'un-changed'	c) 1.05 (95% CI 0.78 to 1.31)*
b) Absolute change in Modified Baron Score of greater than 1 point	
c) Absolute change in sum of Mayo Clinic Score rectal bleeding and stool frequency subscores of at least 2 points	
<hr/>	
Guyatt's responsiveness statistic (GRS; Guyatt 1987) calculated using:	a) 0.88 (95% CI 0.64 to 1.12)*
a) treatment allocation with patients assigned to MLN02 considered 'changed' and those assigned to placebo considered 'un-changed'	b) 0.73 (95% CI 0.50 to 0.96)*
	c) 0.84 (95% CI 0.59 to 1.09)*
b) Absolute change in Modified Baron Score of greater than 1 point	
c) Absolute change in sum of Mayo Clinic Score rectal bleeding and stool frequency subscores of at least 2 points	
<hr/>	

*Effect sizes of 0.2, 0.5 and 0.8 were considered to represent low, moderate and large degrees of responsiveness, respectively ([Cohen 1988](#)).

Table 8. Summary of Operating Properties of Histologic Scoring Indices for Ulcerative Colitis

Study ID	Scoring index	Validity			Reliability				Respon- siveness	Feasibili- ty
		Content validity	Criterion validity	Construct validity	In- tra-rater	In- ter-rater	Test- retest	Internal consis- tency		
Feagan 2005	Modified Riley Score	?	?	?	?	+	?	?	?	?
Fiel 2003	Harpaz/Mount Sinai In- dex	?	?	?	?	+	?	?	?	?
Geboes 2000	Geboes Score	+	?	+	?	+	?	?	?	?
Gomes 1986	Gomes Index	?	+	?	?	?	?	?	?	?
Jauregui-Ameza- ga 2016	Simplified Geboes Score	?	?	?	?	+	?	?	?	?
Mar- chal-Bressenot 2017	Nancy Index	+	?	+	+	+	?	?	+	?
Mosli 2017	Robarts Histopathology Score	+	?	+	+	+	?	?	+	?
Riley 1991	Riley Score	?	+	?	?	+	?	?	?	?
Rubin 2007	Chicago/Rubin/Histolog- ic Inflammation Activity Scale	?	?	+	?	?	?	?	?	?
Theede 2015	Modified Harpaz Index	?	?	+	?	+	?	?	?	?
Truelove 1956	Truelove and Richards Index	?	?	+	?	?	?	?	?	?

+ positive rating

? no information or indeterminate rating

- Negative rating

Table 9. The Methodological Quality of Histologic Index Measurement Properties (COSMIN Checklist)

	Study ID	Scoring index	Internal consistency	Reliability	Measurement error	Content validity	Structural validity (CSV)	Hypothesis testing (CSV)	Cross-cultural validity (CSV)	Criterion validity	Responsiveness	Interpretability
1	Feagan 2005	Modified Riley Score	?	excellent	?	?	?	?	?	?	?	?
2	Fiel 2003	Harpaz/Mount Sinai Index	?	fair	?	?	?	?	?	?	?	?
3	Geboes 2000	Geboes Score	?	excellent	?	excellent	?	excellent	?	?	?	?
4	Gomes 1986	Gomes Index	?	?	?	?	?	?	?	excellent	?	?
5	Jau-regui-Ameza-ga 2016	Simplified Geboes Score	?	fair	?	?	?	?	?	?	?	?
6	Mar-chal-Bressenot 2017	Nancy Index	?	excellent	?	excellent	?	excellent	?	?	excellent	?
7	Mosli 2017	Robarts Histopathology Score	?	excellent	?	excellent	?	excellent	?	?	excellent	?
8	Riley 1991	Riley Score	?	good	?	?	?	?	?	good	?	?
9	Rubin 2007	Chicago/Rubin/Histologic Inflammation Activity Scale	?	?	?	?	?	good	?	?	?	?
10	Theede 2015	Modified Harpaz Index	?	excellent	?	?	?	excellent	?	?	?	?
11	Truelove 1956	Truelove and Richards Index	?	?	?	?	?	excellent	?	?	?	?

CSV = construct validity
? no information or indeterminate rating

APPENDICES

Appendix 1. Search Strategies

MEDLINE - Search Strategy

1. ulcerative colitis.mp. or exp Colitis, Ulcerative/
2. histol*.mp. or exp Pathology, Clinical/
3. exp Immunohistochemistry/ or immunohisto*.mp.
4. exp Pathology/ or patholog*.mp.
5. exp Biopsy/ or biops*.mp.
6. 2 or 3 or 4 or 5
7. (index or index* or indice or indice* or scale* or scali* or score or score* or scori* or riley or Gebo*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]
8. 1 and 6 and 7

EMBASE - Search Strategy

1. histol*.mp. or exp histology/
2. exp immunohistochemistry/ or immunohisto*.mp.
3. exp pathology/ or patholog*.mp.
4. exp biopsy/ or biops*.mp.
5. 1 or 2 or 3 or 4
6. ulcerative colitis.mp. or exp ulcerative colitis/
7. (index or index* or indice or indice* or scale* or scali* or score or score* or scori* or riley or Gebo*).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
8. 5 and 6 and 7

PubMed - Search Strategy

1. Search (colitis and ulcerat*)
2. Search (histolog* OR patholog* OR immunohisto* OR biops*)
3. Search (index OR index* OR indice OR indice* OR scale* OR scali* OR score OR score* OR scori* OR riley OR Gebo*)
4. Search (#1 AND #2 AND #3)

Cochrane Library (CENTRAL) - Search Strategy

1. (colitis and ulcerat*)
2. histol* or pathol* or immunohisto* or biops*
3. index or index* or indice or indice* or scale or scale* or scali* or score or score* or scori* or riley or Gebo*
4. #1 and #2 and #3

IBD/FBD Specialized Register - Search Strategy

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

WHAT'S NEW

Date	Event	Description
7 July 2017	Amended	Correction of minor errors in table

CONTRIBUTIONS OF AUTHORS

All authors were involved in the planning, execution and drafting of this systematic review.

DECLARATIONS OF INTEREST

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

Mahmoud H Mosli: None known.

Claire E Parker: None known.

Sigrid A Nelson: None known

Kenneth A Baker: None known.

John K MacDonald: None known.

GY Zou: None known.

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

Reena Khanna: Dr. Khanna has received consulting fees from AbbVie, Janssen, Pfizer, Shire, Takeda. All of these financial activities are outside the submitted work.

Barrett G Levesque: Dr Levesque has received fees from Prometheus Labs, Takeda, Abbvie, Nestle Health Sciences, Gilead, and Roche for consultancy; and payment for lectures from Warner Chilcott, Salix, and UCB Pharma. All of these financial activities are outside the submitted work.

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

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

The methods for assessing risk of bias in included studies were modified from the protocol. Originally, we planned on using four items to assess risk of bias: blinded design, independent observation, performance bias and detection bias. However, since this is a review of scoring indices rather than interventions, the last two items are not applicable. We chose to evaluate blinded design and independent observation, and to further assess risk of bias using a system based on the COSMIN tool.

In the protocol, we did not state what method we would use to interpret the ICCs. We have chosen to employ the Landis and Koch method ([Landis 1977](#)), and this is now explained in the text.

Select wording in the Background section was also modified for clarity.

NOTES

The literature search was conducted on 2 December 2016. While e-publications introducing the Nancy Index and Robarts Histopathology Index were available at this time, these manuscripts were not published in print until January 2017. The references from January 2017 were added to the flow diagram under the 'identified through other sources' section, and used as the study IDs.

INDEX TERMS

Medical Subject Headings (MeSH)

*Severity of Illness Index; Blood Sedimentation; C-Reactive Protein [analysis]; Colitis, Ulcerative [*pathology]; Feces [chemistry]; Lactoferrin [analysis]; Leukocyte Count; Leukocyte L1 Antigen Complex [analysis]; Observer Variation; Pancreatic Elastase [analysis]; Reproducibility of Results

MeSH check words

Humans