



Comparing disease activity indices in ulcerative colitis

A.J. Walsh^a, A. Ghosh^b, A.O. Brain^g, O. Buchel^c, D. Burger^d, S. Thomas^e, L. White^g, G.S. Collins^f, S. Keshav^g, S.P.L. Travis^{g,*}

^a Currently at Department of Gastroenterology, St Vincent's Hospital, Sydney, Australia

^b Medical Sciences Division, University of Oxford, Oxford, UK

^c Currently at Department of Surgery, Universitas Hospital, University of the Free State, Bloemfontein, South Africa

^d Currently at Department of Gastroenterology, Princess Alexandra Hospital, Brisbane, Australia

^e Currently at Department of Biomedical Sciences, University of Sheffield, Sheffield, UK

^f Centre for Statistics in Medicine, University of Oxford, Oxford, UK

^g Translational Gastroenterology Unit, John Radcliffe Hospital, Oxford, UK

Received 1 August 2013; received in revised form 28 August 2013; accepted 9 September 2013

KEYWORDS

Ulcerative colitis;
Activity index;
Endoscopy;
Mayo Clinic index;
Clinical trial design

Abstract

Background: Comparisons between disease activity indices for ulcerative colitis (UC) are few. This study evaluates three indices, to determine the potential impact of inter-observer variation on clinical trial recruitment or outcome as well as their clinical relevance.

Methods: One hundred patients with UC were prospectively evaluated, each by four specialists, followed by videosigmoidoscopy, which was later scored by each specialist. The Simple Clinical Colitis Activity (SCCAI), Mayo Clinic and Seo indices were compared by assigning a disease activity category from published thresholds for remission, mild, moderate and severe activity. Inter-observer variation was evaluated using Kappa statistics and its effect for each patient on recruitment and outcome measures for representative clinical trials calculated. Clinical relevance was assessed by comparing an independently assigned clinical category, taking all information into account as if in clinic, with the disease activity assigned by the indices.

Results: Inter-observer agreement for SCCAI ($\kappa = 0.75$, 95% CI 0.70–0.81), Mayo Clinic ($\kappa = 0.72$, 95% CI 0.67–0.78) and Seo ($\kappa = 0.89$, 95% CI 0.83–0.95) indices was good or very good as was the agreement for rectal bleeding ($\kappa = 0.77$) and stool frequency ($\kappa = 0.90$). Endoscopy in the Mayo Clinic index had the greatest variation ($\kappa = 0.38$). Inter-observer variation alone would have excluded up to 1 in 5 patients from recruitment or remission criteria in representative trials. Categorisation by the SCCAI, Mayo Clinic and Seo indices agreed with the independently assigned clinical category in 61%, 67% and 47% of cases respectively.

* Corresponding author at: Translational Gastroenterology Unit, John Radcliffe Hospital, Oxford OX3 9DU, UK. Tel.: +44 1865 228753; fax: +44 1865 228763.

E-mail address: simon.travis@ndm.ox.ac.uk (S.P.L. Travis).

1873-9946/\$ - see front matter. Crown Copyright © 2013 Published by Elsevier B.V. on behalf of European Crohn's and Colitis Organisation. All rights reserved.

<http://dx.doi.org/10.1016/j.crohns.2013.09.010>

Conclusions: Trial recruitment and outcome measures are affected by inter-observer variation in UC activity indices, and endoscopic scoring was the component most susceptible to variation. Crown Copyright © 2013 Published by Elsevier B.V. on behalf of European Crohn's and Colitis Organisation. All rights reserved.

1. Introduction

Instruments for assessing disease activity in ulcerative colitis (UC) are needed to define inclusion criteria and outcomes of clinical trials with precision, although they are less often used in routine clinical practice. This is despite, or even because of, a multiplicity of disease activity indices for UC, with almost a dozen developed for use in clinical trials.¹ The Pediatric Ulcerative Colitis index (PUCAI), for use in children, is the only clinical index that has been validated for symptom severity.² None in inter-observer variation has been examined and there has been only one systematic study evaluating the activity indices in UC,¹ which made no empirical comparisons between different indices. Inter-observer variation is known to be substantial in endoscopic assessment.³ The effect of variation in clinical assessment on trial recruitment or reported outcomes has not been studied.

There are many problems with current disease activity indices for UC, not least the absence of formal evaluation. Most are modifications of pre-existing indices, which therefore use similar terms, few of which are consistently defined, but omit symptoms of importance to patients, such as urgency or faecal continence. Since thresholds for remission, active disease and response to treatment vary,^{1,4} it is difficult to compare therapeutic trials and this is another reason that indices for UC are rarely used in clinical practice.

In clinical practice, disease activity in UC is assessed to a greater or lesser degree by clinical symptoms, endoscopic appearance, histopathology, biomarkers and quality of life. Consequently composite indices have been developed, such as the Mayo Clinic index, that include clinical symptoms, endoscopy, aspects of quality of life and the physician's global assessment (PGA).⁵ In an attempt to bring objectivity to the assessment of disease activity, the Seo index⁶ combines biomarkers with clinical symptoms. On the other hand, it may be better to separate symptoms from endoscopy,³ histopathology,⁷ biomarkers and quality of life.⁸ The Simple Clinical Colitis Activity Index⁹ is based on clinical symptoms alone. It is easier to validate separate indices^{3,7,8} so composite thresholds might then be set for recruitment or outcomes in clinical trials.

Recruitment to clinical trials requires that minimum and maximum disease activities are specified in the inclusion and exclusion criteria. Furthermore, the outcomes of treatment are typically defined as the number of patients with a defined change in disease activity index, or meeting a threshold criterion, typically for remission. Clearly therefore, inter-observer variation in determining clinical index scores would affect recruitment and outcomes. The implications of such variation have already been demonstrated in a trial of mesalazine for UC, where the outcome was contingent on inter-observer variation in endoscopy alone.¹⁰

The aim of this study was to evaluate the inter-observer variation in a subset of UC indices in the same set of patients.

A clinical index (Simple Clinical Colitis index⁹), a composite clinical and endoscopy index (Mayo Clinic index⁵) and a composite clinical and biomarker index (Seo index⁶) were selected. This allowed an assessment of the impact that inter-observer variation in the assessment of activity might have on recruitment or remission outcomes defined in representative clinical trials. It additionally helped determine which items of which index have the most variability and assessed the potential clinical relevance of the three indices.

2. Methods

2.1. Patients

Patients with UC of varying disease activities and extent of disease who requested a review appointment at the Inflammatory Bowel Disease clinic at the John Radcliffe Hospital, Oxford, were invited to participate. UC had previously been diagnosed by conventional criteria.¹¹ Patients with Crohn's disease or colitis yet to be classified were excluded.

2.2. Clinic logistics and endoscopy

One hundred patients were prospectively evaluated in 16 consecutive, specially designed clinics between August 2007 and April 2008. Each patient consented (Oxford LREC 536407Q1605/58ORH) to be seen and examined by the same four specialists in inflammatory bowel disease (AB, AW, SK, ST), to have a blood test, and to undergo videosigmoidoscopy on the same day. Each patient completed a record of symptoms (Supplementary file A) before being seen and examined by each specialist in random order. Each specialist recorded the clinical symptoms on a standard form that captured the terms for each index, blinded to other results (Supplementary file B). The last specialist to see the patient was responsible for the treatment decisions. The patient then proceeded to videosigmoidoscopy on the same day, performed by a fifth specialist (OB), according to a standard protocol.¹² Videos were anonymised and then scored at a later date (Baron,¹³ Modified Baron,¹³ Mayo Score Flexible Proctosigmoidoscopy Assessment,⁵ Sutherland Mucosal Appearance Assessment¹⁴) (Supplementary file C), in random order, by the first four specialists who were asked to score the worst affected area, blinded to all clinical details. Videosigmoidoscopy was unavailable for 4 patients (pregnancy 1, patient left prior to sigmoidoscopy 2, recording equipment failure 1).

2.3. Indices

Three indices were selected for comparison. Six other indices were recorded (Truelove and Witts' index,¹⁵ Powell-Tuck/St Mark's index,¹⁶ Sutherland index/Ulcerative Colitis Disease

Table 1 Simple Clinical Colitis Activity index (SCCAI) descriptors, thresholds and kappa.

Descriptor	Level	Kappa (95% CI)
Bowel frequency (day)		0.83 (0.78 to 0.88)
1–3	0	
4–6	1	
7–9	2	
>9	3	
Bowel frequency (night)		0.89 (0.81 to 0.97)
1–3	0	
4–6	1	
Urgency of defaecation		0.76 (0.71 to 0.81)
Nil	0	
Hurry	1	
Immediately	2	
Incontinence	3	
Blood in stool		0.77 (0.72 to 0.83)
No	0	
Trace	1	
Occasionally frank	2	
Usually frank	3	
General well being		0.88 (0.83 to 0.94)
Very well	0	
Slightly below par	1	
Poor	2	
Very poor	3	
Terrible	4	
Extracolonic features	1 (per manifestation)	0.25 (0.18 to 0.32)
Thresholds (Jowett et al. ²⁴ ; Higgins et al. ²⁸ ; D'Haens et al. ¹)		
Remission	≤2	
Mild	3 to 5	
Moderate	6 to 11	
Severe	≥12	
Total score	(Range 0–19)	0.75 (0.70 to 0.81)

Components of the Simple Clinical Colitis Activity Index (SCCAI),⁹ indicating thresholds used to categorise into remission, mild, moderate and severe disease activity with corresponding reference papers and interobserver variation in individual components and overall score as measured by the kappa statistic.

Activity index,¹⁴ Lichtiger/Modified Truelove and Witts' index,¹⁷ Rachmilewitz/Colitis Activity index,¹⁸ Ulcerative Colitis Clinical Score¹⁹) but owing to marked similarity between many of the indices and the statistical challenge of comparing multiple indices with different scales, three were chosen to represent a purely clinical index, a composite clinical and endoscopic index and a composite clinical and biomarker index. Terminology is used as consistently as possible: *index* refers to an instrument for assessing disease activity; *item* refers to a component within that index with a *level* of severity often recorded using a Likert scale; and *score* is used to describe the overall measure of the index. *Disease activity category* is the assessment of activity (remission, mild, moderate, severe) assigned by the score. *Clinical category* is used to describe the overall assessment of activity taking into account clinical, endoscopic, histopathology and biomarker information, beyond that assessed by any individual index. This category formed the basis of the assessment of clinical relevance.

The SCCAI^{2,9} (Table 1) is a clinical index which has been compared prospectively with a composite index including sigmoidoscopy and biomarkers in 86 adult patients, showing close correlation ($r = 0.959$, $p < 0.0001$).^{2,20} Since the SCCAI is a purely clinical index, patients can complete it independently. Scores range from 0 to 19 points (no activity to most severe).

The Mayo Clinic index^{5,21} (Table 2) is a composite clinical, endoscopic, quality of life and global assessment index, widely used in clinical trials. Scores range from 0 to 12 points (no activity, to most severe). Sub-scores (combining rectal bleeding, stool frequency and Physician's Global Assessment, or the Endoscopy sub-score) are also used in clinical trials to record a partial Mayo Clinic score. Although the original index has not been evaluated against an independent measure of disease activity, it has gained credence by widespread use in clinical trials.¹

The Seo index^{4,6} (Table 3), also not validated, was derived using multivariable regression analysis of prospective data on

Table 2 Mayo Clinic index descriptors, thresholds and kappa.

Descriptor	Level	Kappa (95% CI)
Stool frequency		0.73 (0.68 to 0.77)
Normal no. stools for this patient	0	
1–2 stools more than normal	1	
3– 4 stools more than normal	2	
5 or more stools more than normal	3	
Rectal bleeding		0.77 (0.72 to 0.83)
No blood seen	0	
Streaks of blood with stool less than half the time	1	
Obvious blood with stool most of the time	2	
Blood alone passed	3	
Findings of flexible proctosigmoidoscopy		0.38 (0.33 to 0.43)
Normal or inactive disease	0	
Mild disease	1	
Moderate disease	2	
Severe disease	3	
Physician's global assessment		0.56 (0.51 to 0.61)
Normal	0	
Mild disease	1	
Moderate disease	2	
Severe disease	3	
Thresholds (Rutgeerts et al. ²⁵ ; D'Haens et al. ¹)		
Remission	≤2 with no subscore >1	
Mild	3–5	
Moderate	6 to 10	
Severe	≥10	
Total score	(Range 0–12)	0.72 (0.67 to 0.78)

Components of the Mayo Clinic index,⁵ indicating thresholds used to categorise into remission, mild, moderate and severe disease activity with corresponding reference papers and interobserver variation in individual components and overall score as measured by the kappa statistic.

18 clinical, laboratory and sigmoidoscopy variables, collected from 72 patients during 85 clinical events. Scores range from 100 to 300 points (no activity, to most severe). (Seo score = $60 \times \text{blood in stool} + 13 \times \text{bowel movements} + 0.5 \times \text{ESR} - 4 \times \text{haemoglobin (g/dL)} - 15 \times \text{albumin} + 200$).

Data from the patient and specialist report forms were entered into a database, with each level for each item documented and the total score calculated for each index. Values were independently checked for consistency (GC).

2.4. Comparison between indices

To compare the three indices allowing for three different ordinal scales (0–19 for SCCAI, 0–12 for Mayo Clinic and 100–300 for Seo indices), the total score from each index was assigned a disease activity category (remission, mild, moderate, severe activity) according to published thresholds¹ (Tables 1–3). This enabled inter-observer variation between the 4 specialists to be calculated (see statistical methods) without being influenced by the span of the scale.

2.5. Potential effect of inter-observer variation on clinical trial recruitment or outcome

To determine the impact of inter-observer variation on trial inclusion or remission criteria defined in representative

trials, a reference clinical trial for each index was selected (Table 4). Such is the variation in definitions used by different studies, that the selection of the study (ST & SK) was based on the use of the specific index as a primary endpoint, usually in the papers used to derive categories of activity (above). Clinical, endoscopic and biomarker data where appropriate (but not age, medication or other criteria) on each patient from each specialist were matched against inclusion and remission criteria for disease activity in the respective trials. The proportions of patients meeting trial recruitment criteria or remission outcome defined in the reference trial as a consequence of inter-observer agreement were then calculated.

2.6. Potential clinical relevance

To determine how scores from each index compared with an overall assessment of disease activity as might be made in clinic, an experienced clinician (ST) was given access to all data regarding symptoms, examination, blood results, sigmoidoscopy assessment and histopathology. Using this information, patients were assigned a clinical category (remission, mild, moderate, severe). To minimize recall bias, assignment of the clinical category was performed several months after the patient visit, and the clinician was blinded to the identity of the patient. The clinical category was then compared with the disease activity category assigned by each of the three indices.

Table 3 Seo index descriptors, thresholds and kappa.

Descriptor	Level	Weight	Kappa (95% CI)
Bloody stool		60	0.90 (0.82 to 0.98)
Little or none	0		
Present	1		
Bowel movements		13	0.84 (0.78 to 0.90)
≤4	1		
5–7	2		
≥8	3		
Erythrocyte sedimentation rate (mm/h)	–	0.5	–
Haemoglobin (g/dL)	–	–4	–
Albumin (g/dL)	–	–15	–
Constant	–	200	–
Thresholds (Järnerot et al. 2005 ²⁶ ; Higgins et al. 2005 ²⁸)			
Remission	<120		
Mild	120 to 149		
Moderate	150 to 220		
Severe	>220		
Total score	–	–	0.89 (0.83 to 0.95)

Components of the Seo index,⁶ indicating weighting, thresholds used to categorise into remission, mild, moderate and severe disease activity with corresponding reference papers and interobserver variation in individual components and overall score as measured by the kappa statistic.

3. Statistical analysis

Raw agreement was calculated as the proportion of exact agreement between specialists. Agreement between the 4 specialists and the disease activity category for each index was evaluated using the Kappa statistic. For each individual item in the indices, agreement between the four specialists was assessed both as a percentage of all patients and by Fleiss's Kappa statistic for categorical data, taking the average pair-wise agreement between any two specialists. Kappa values are interpreted as: <0.2 poor agreement; 0.21 to 0.4 fair agreement; 0.41 to 0.6 moderate agreement; 0.61 to 0.8 good agreement and 0.81 to 1.0 very good agreement.²² Agreement between the disease activity category for each index assigned by the 4 specialists and the clinical category was evaluated as a percentage for all patients.

No formal sample size calculation was carried out because the study is exploratory and descriptive, with no formal

hypothesis testing. A sample size of 100 consecutive patients rated by 4 observers was assumed sufficiently large enough to provide reasonable estimates of confidence intervals.

4. Results

4.1. Patients

Of 100 patients, 50 were male, median age 49 years (range 19 to 82 years), median duration of disease 107 months (range <1 to 575 months), maximum disease extent: proctitis (E1) 27%, left-sided colitis (E2) 39%, pancolitis (E3) 34%.²³

4.2. Inter-observer variation for the three indices

Inter-observer variation between indices for disease activity category showed kappa values for the SCCAI, Mayo Clinic and

Table 4 Percentage agreement of Simple Clinical Colitis Index (SCCAI), Mayo Clinic index and Seo index inclusion and remission criteria.

Index	Inclusion			Remission		
	Reference	Criteria	Total agreement	Reference	Criteria	Total agreement
SCCAI	Jowett et al. ²⁴	≥5	85%	Jowett et al. ²⁴	≤2	89%
Mayo Clinic index	Rutgeerts et al. ²⁵	≥6 and mayo endo ≥2	80%	Rutgeerts et al. ²⁵	≤2 with no subscore >1	83%
Seo	Järnerot et al. ²⁶	>150	94%	Higgins et al. ²⁷	<120	95%

Percentage agreement regarding inclusion and remission criteria between all 4 clinicians for the Simple Clinical Colitis Activity Index (SCCAI), Mayo Clinic index and Seo index as specified by anchor papers.

Seo indices to be 0.75 (95% CI 0.70–0.81), 0.72 (95% CI 0.67–0.78) and 0.89 (95% CI 0.83–0.95), respectively (Tables 1–3). By way of reference, kappa between raters for endoscopy alone using the Modified Baron index was 0.44 (with 31% complete agreement).

4.3. Inter-observer variation between the items of each index

4.3.1. SCCAI

The greatest agreement between raters was for nocturnal bowel frequency (91% of all patients; $\kappa = 0.89$), with more variation for daytime bowel frequency (80%; $\kappa = 0.83$) and blood in stool (65%; $\kappa = 0.77$). The only item that significantly increased the average kappa value when removed from the index was 'extracolonic features' (average item κ increased from 0.73 to 0.83) (Table 1).

4.3.2. Mayo Clinic index

The greatest agreement between raters was for rectal bleeding (75%, $\kappa = 0.77$). Physician's global assessment (47%; $\kappa = 0.56$) and endoscopic subscore (21%; $\kappa = 0.38$) were appreciably less consistent. The impact of removing the endoscopic subscore from the Mayo Clinic index increased the average item κ from 0.61 to 0.69 (Table 2).

4.3.3. Seo index

The greatest agreement between specialists was for biomarker values (100%, $\kappa = 1.0$). There was 91% agreement for rectal bleeding ($\kappa = 0.90$) and 85% for stool frequency ($\kappa = 0.84$). Removal of either of these items did not significantly improve the average kappa value (Table 3).

4.4. Potential effect of inter-observer variation on clinical trial recruitment (Table 4)

4.4.1. SCCAI

There was 85% agreement between specialists for identifying patients with relapse defined by a study evaluating the definition of relapse using the SCCAI.^{3,24} This implies that if the agreement between a group of four investigators was required, 1 in 6 patients would potentially be excluded from enrolment according to this threshold.

4.4.2. Mayo Clinic index

There was 80% agreement between specialists for identifying patients who would have met the inclusion criteria for disease activity in the ACT1 & ACT 2 studies.^{7,25} This implies that 1 in 5 patients would potentially be excluded from enrolment due to inter-observer variation between 4 investigators.

4.4.3. Seo index

There was 94% agreement between specialists for patients meeting inclusion criteria for disease activity in a study evaluating infliximab in acute colitis.^{7,8,25,26} This implies that 1 in 16 patients would potentially be excluded due to inter-observer variation.

Table 5 Agreement of Simple Clinical Colitis Index (SCCAI), Mayo Clinic index and Seo index with clinical standard.

Index	% agreement with clinical standard
SCCAI	61 (range 56–64)
Mayo Clinic index	67 (range 61–71)
Seo	47 (range 39–56)

Percentage agreement of the Simple Clinical Colitis Index (SCCAI), Mayo Clinic index and Seo index with the clinical standard assigned by an experienced, blinded clinician given access to all data beyond that available from the indices, including symptoms, examination findings, blood results, flexible sigmoidoscopy and histopathology.

4.5. Potential effect of inter-observer variation on clinical trial remission outcome (Table 4)

4.5.1. SCCAI

There was 89% agreement between specialists with regard to remission defined as SCCAI <2.^{5,24} This means that if agreement between four investigators was required, 1 in 9 patients might be excluded from an endpoint due to inter-observer variation.

4.5.2. Mayo Clinic index

There was 83% agreement for remission as defined in the ACT 1 & 2 studies.^{9,25} This means that if agreement between four investigators was required, 1 in 5 patients might be excluded from an endpoint defining remission by these criteria, as a consequence of inter-observer variation.

4.5.3. Seo index

There was 95% agreement for remission defined in a study evaluating patient-defined endpoints.²⁷ This means that 1 in 20 patients would be excluded from such an endpoint due to inter-observer variation.

4.6. Potential clinical relevance

Agreement between disease activity categories assigned for each index and the clinical category assigned by a clinician with access to all information beyond that recorded by any individual index, were 61% (range 56–64) for the SCCAI, 67% (range 61–71) for the Mayo Clinic index and 47% (range 45 to 49) for the Seo index (Table 5). By way of reference, the percentage agreement between the clinical category and endoscopy (Modified Baron index) was 45% (range 39–56).

5. Discussion

This study intended to determine whether commonly used disease activity indices for UC are reliable and clinically relevant and how inter-observer variation might affect trial recruitment or remission outcomes defined in representative clinical trials. There is no acknowledged metric for comparing indices, which vary in their scale of scores and the data collected. The three indices were chosen to represent three types of index: a clinical index (SCCAI), a composite clinical and endoscopic index (Mayo Clinic index)

and a composite clinical and biomarker index (Seo). A widely used endoscopic index (modified Baron index) was used by way of reference.

Inter-observer agreement for all three indices was good or very good, all performing better than the Modified Baron Score which only had moderate agreement between the same four raters. It is not surprising that the Seo index was the most reliable, since the biomarkers (blood results) were constant between each rater. Valid comparison of inter-observer variation between indices requires a common denominator, so because the scales for each index differed (with maximum scores ranging from 12 to 300), each index was assigned a disease activity category (remission, mild, moderate, severe) according to published thresholds. This adds another layer of complexity since the thresholds (let alone the indices themselves) have not been validated.

With regard to items within the indices, agreement for nocturnal stool frequency and rectal bleeding performed well or very well across the three indices. It may seem surprising that there was any variation, since patients were seen by the four specialists in rapid succession, but this reflects variation in clinical consultation and differences in the definition of each level. By way of example, the total number of stools per day is used in the SCCAI in contrast to the increase in the number of stools per day in the Mayo Clinic index. Documentation of extraintestinal features was highly variable between specialists. This is a weakness of the SCCAI in addition to the limited relevance of extra-intestinal features in evaluating activity, since susceptibility is influenced by genetics rather than activity alone. The endoscopic component of the Mayo Clinic index performed surprisingly poorly (21% agreement, $\kappa = 0.38$), although consistent with variation in endoscopic assessment by other specialists.³ Indeed, the partial Mayo Clinic index without endoscopy performed better (average item $\kappa = 0.69$) than the index itself (average item $\kappa = 0.61$). This supports the view that the partial Mayo Clinic index is reliable, but also suggests that there may be merit in separating endoscopic scoring from symptom-based or quality of life indices. The introduction of the UCEIS may reduce the variation in endoscopic scoring, allowing endoscopy to remain inclusion criteria and an endpoint for UC trials, especially with the advent of central reading of videos.²⁹

The study shows that inter-observer variation in disease activity indices has a potential impact on trial recruitment or remission outcomes. For the Mayo Clinic index, around 1 in 5 patients would have been excluded from recruitment to the reference trial (disease activity criteria for ACT 1 & 2) had agreement been required between all 4 specialists. The outcome of remission defined in the trial would have been similarly affected, solely as a consequence of inter-observer variation between 4 specialists. This has implications for calculating the power of a study and also for the conduct of a study. Given that endoscopy is the component subject to most inter-observer variation, central reading of videosigmoidoscopy is increasingly the norm.¹⁰ Whilst central reading does not abolish subjectivity, consistency is improved and this can alter the outcome of clinical trials.¹⁰ It is worth noting that steps to reduce inter-observer variation may matter more to small (Phase 2) studies than to larger studies, depending on the anticipated size of effect.

Nevertheless, good inter-observer agreement is not the only the goal of an index. It needs to be clinically relevant

and best represent the severity of a patient's disease. Our data suggest that the Mayo Clinic index most closely reflects clinical assessment in outpatients, probably because it includes the item 'Physician's Global Assessment'. This is in effect of the 'clinical category' assigned by an experienced clinician taking all information into account. The Seo index based on biomarkers and subject to least inter-observer variation, performed poorly compared to the other two indices, with only 47% agreement between the index and the assigned clinical category. The Seo index also performs less well in terms of feasibility, discriminative ability, test-retest reliability and responsiveness.²

All three indices studied show good or very good inter-observer agreement, largely because they ask similar clinical questions. Even so, there was a sufficient inter-observer variation when using the indices to categorise disease activity to have a potential impact on recruitment to clinical trials or attaining a defined outcome. Endoscopy is most subject to inter-observer variation. Consequently composite indices that include endoscopy are intrinsically more variable. On a practical level, the Mayo Clinic index and the SCCAI appear to correlate best with an independently assigned clinical category. There is no ideal index, but international agreement on whether to adopt a single index, or to create a composite outcome measure from validated indices of clinical, endoscopic, histologic and quality of life variables would facilitate comparisons between clinical trials in UC.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.chrons.2013.09.010>.

References

1. D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology* 2007;132:763–86.
2. Turner D, Griffiths AM, Mack D, Otley AR, Seow CH, Steinhart AH, et al. Assessing disease activity in ulcerative colitis. *Inflamm Bowel Dis* 2010;16:651–6.
3. Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JC, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2012;61:535–42.
4. Travis SPL, Higgins PDR, Orchard T, Van Der Woude CJ, Panaccione R, Bitton A, et al. Review article: defining remission in ulcerative colitis. *Aliment Pharmacol Ther* 2011;34:113–24.
5. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med* 1987;317:1625–9.
6. Seo M, Okada M, Yao T, Ueki M, Arima S, Okumura M. An index of disease activity in patients with ulcerative colitis. *Am J Gastroenterol* 1992;87:971–6.
7. Geboes K, Riddell R, Öst A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;47:404–9.
8. Irvine EJ, Feagan B, Rochon J, Archambault A, Fedorak RN, Groll A, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease Canadian Crohn's Relapse Prevention Trial Study Group. *Gastroenterology* 1994;106:287–96.
9. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998;43:29–32.

10. Feagan BG, Sandborn WJ, D'Haens G, Pola S, McDonald JWD, Rutgeerts P, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;**145**:149–157.e2.
11. Dignass A, Eliakim R, Magro F, Maaser C, Chowers Y, Geboes K, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. *J Crohn's Colitis* 2012;**6**:965–90.
12. Sandborn WJ, Regula J, Feagan BG, Belousova E, Jovic N, Lukas M, et al. Delayed-release oral mesalamine 4.8 g/day (800-mg tablet) is effective for patients with moderately active ulcerative colitis. *Gastroenterology* 2009;**137**:1934–1943.e3.
13. Baron JH, Connell AM, Lennard-Jones JE. Variation between observers in describing mucosal appearances in proctocolitis. *Br Med J* 1964;**1**:89.
14. Sutherland LR, Martin F, Greer S, Robinson M, Greenberger N, Saibil F, et al. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. *Gastroenterology* 1987;**92**:1894–8.
15. Truelove SC, Witts LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *Br Med J* 1955;**2**:1041–8.
16. Powell-Tuck J, Bown RL, Lennard-Jones JE. A comparison of oral prednisolone given as single or multiple daily doses for active proctocolitis; 1978.
17. Lichtiger S, Present DH. Preliminary report: cyclosporin in treatment of severe active ulcerative colitis. *Lancet* 1990;**336**:16–9.
18. Rachmilewitz D. Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. *BMJ* 1989;**298**:82–6.
19. Feagan BG, Greenberg GR, Wild G, Fedorak RN, Paré P, McDonald JW, et al. Treatment of ulcerative colitis with a humanized antibody to the alpha4beta7 integrin. *N Engl J Med* 2005;**352**:2499–507.
20. Turner D, Seow CH, Greenberg GR, Griffiths AM, Silverberg MS, Steinhart AH. A systematic prospective comparison of noninvasive disease activity indices in ulcerative colitis. *YJCGH* 2009;**7**:1081–8.
21. Van Assche G, D'Haens G, Noman M, Vermeire S, Hiele M, Asnong K, et al. Randomized, double-blind comparison of 4 mg/kg versus 2 mg/kg intravenous cyclosporine in severe ulcerative colitis. *Gastroenterology* 2003;**125**:1025–31.
22. Landis JR, Koch GG. JSTOR: biometrics. *Biometrics* 1977;**33**:363–74 (No. 2 (Jun., 1977)).
23. Satsangi J. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;**55**:749–53.
24. Jowett SL, Seal CJ, Phillips E, Gregory W, Barton JR, Welfare MR. Defining relapse of ulcerative colitis using a symptom-based activity index. *Scand J Gastroenterol* 2003;**38**:164–71.
25. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;**353**:2462–76.
26. Järnerot G, Hertervig E, Friis-Liby I, Blomquist L, Karlén P, Grännö C, et al. Infliximab as rescue therapy in severe to moderately severe ulcerative colitis: a randomized, placebo-controlled study. *Gastroenterology* 2005;**128**:1805–11.
27. Higgins PDR. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. *Gut* 2005;**54**:782–8.
28. Higgins PDR, Schwartz M, Mapili J, Zimmermann EM. Is endoscopy necessary for the measurement of disease activity in ulcerative colitis? *Am J Gastroenterol* 2005;**100**:355–61.
29. Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, et al. Reliability and initial validation of the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gastroenterology* 2013, <http://dx.doi.org/10.1053/j.gastro.2013> pii: S0016-5085(13)01075-5. (in press epub Jul 24).