

The quantitative validation of non-endoscopic disease activity indices in ulcerative colitis

P. D. R. HIGGINS*, J. LEUNG†, M. SCHWARTZ‡, J. MAPILIS, P. A. WREN¶ & E. M. ZIMMERMANN*

*Department of Internal Medicine, University of Michigan, Ann Arbor, MI; †Department of Internal Medicine, University of Iowa, Iowa; ‡Department of Internal Medicine, University of Chicago, Chicago; §Department of Family Medicine, Wayne State University, Detroit; ¶Department of Health Behavior and Health Education, University of Michigan, School of Public Health, Ann Arbor, MI, USA

Correspondence to: Dr P. Higgins, Box 0682, 1150 West Medical Center Drive, Ann Arbor, MI 48109, USA.
E-mail: phiggins@umich.edu

Publication data

Submitted 28 August 2006
First decision 15 September 2006
Resubmitted 14 November 2006
Accepted 14 November 2006

SUMMARY

Background

Ulcerative colitis disease activity indices have not been formally validated.

Aim

To analyse quantitatively the psychometric and performance validity of two non-endoscopic indices for ulcerative colitis, the Simple Clinical Colitis Activity Index and the Seo Index.

Methods

In 66 patients with ulcerative colitis, the measurement of disease activity was repeated with the two non-endoscopic indices, St Mark's Index, and the Inflammatory Bowel Disease Questionnaire. Psychometric validity was evaluated by measuring the content, construct, criterion-convergent and criterion-predictive validity on a 0–1 scale. Performance validity was evaluated by measuring the reproducibility and responsiveness on a 0–1 scale.

Results

The Simple Clinical Colitis Activity Index had good to excellent psychometric and performance validity, while the Seo Index had moderate to excellent psychometric validity and moderate to good performance validity. The Simple Clinical Colitis Activity Index had weaknesses in content validity and in responsiveness. The Seo Index had weaknesses in content validity, construct validity and responsiveness.

Conclusions

These two non-endoscopic indices for ulcerative colitis have good psychometric and performance validity, and are now the most rigorously validated disease activity indices for ulcerative colitis. The Simple Clinical Colitis Activity Index appears to have better overall validity. Quantitative evaluation identifies weaknesses in disease activity indices, and can lead to better disease activity indices for ulcerative colitis.

Aliment Pharmacol Ther 25, 333–342

INTRODUCTION AND BACKGROUND

The measurement of disease activity in ulcerative colitis is critical in determining whether new therapies are effective, but there is no gold standard for measuring disease activity in ulcerative colitis. Many empirically derived indices have been developed, usually including measures of bowel movement frequency, stool blood and clinical assessment.^{1–4} Some indices have added biomarkers in the blood (haemoglobin, albumin or erythrocyte sedimentation rate)^{5, 6} or endoscopic assessment.^{7, 8} None of these indices has been rigorously validated. The validity of a measurement can be measured in several ways, and these approaches can be divided into two categories: psychometric validity and performance validity. Psychometric validity is important to demonstrate that the instrument measures the correct symptom domains or disease state, but once that bar has been reached, it is not particularly valuable for quantitatively comparing two different instruments that are both psychometrically valid. Performance validity measures the ability of the instrument to accurately and reproducibly measure a disease state, and to differentiate between different levels of severity. Performance validity allows the quantitative evaluation and comparison of different instruments, evaluating how well they meet the measurement requirements of a clinical trial.

There are very limited data on the four components of psychometric validity (content validity, construct validity, criterion-convergent validity and criterion-predictive validity) in the existing disease activity indices in ulcerative colitis. Content validity is determined by the presence of all the important components (also known as domains or factors) in the measurement of the disease state. Construct validity is determined by the correlation with another validated index based on the same disease construct. Criterion-convergent validity is determined by the correlation with commonly used indices in the same field. Criterion-predictive validity is determined by the ability of the index to predict clinically important outcomes. No studies have been conducted on the performance validity (reproducibility in stable disease activity and responsiveness when disease activity changes) of the existing disease activity indices in ulcerative colitis.

This leaves regulatory agencies in the difficult position of making decisions about the benefits and risks of highly potent, and potentially dangerous, biological therapies for ulcerative colitis without proven measures

of efficacy.⁹ Regulatory agencies have been forced to use empiric definitions from a non-validated endoscopic index of remission and response in making their decisions about new therapies.⁷ While psychometric validity is important to show that a scale truly measures the disease activity, performance validity is equally important in selecting an instrument to evaluate clinical trial results. The stability of a disease activity index in patients who have no change in the severity of their disease is essential in limiting the placebo response rate, and the responsiveness of the instrument is critical in being able to detect meaningful clinical improvement.

This also leaves clinicians without well-established useful measures of disease activity with which to judge the outcomes of clinical trials, and to compare the efficacy of different therapies for ulcerative colitis. Without these tools, many investigators modify existing indices or invent new ones that suit the therapy in question. The choice of an instrument to show a new therapy in the best light can introduce bias, and can make it very difficult for clinicians to make informed choices about how best to treat their patients with ulcerative colitis.

Recent data published by our group showed that two non-endoscopic indices, the Simple Clinical Colitis Activity Index (SCCAI)³ and the Seo Index⁵ (named after its originator, Mitsuru Seo) (Table 1), were able to predict clinically meaningful endpoints in patients with ulcerative colitis. However, these non-endoscopic and less costly indices are unlikely to be widely used unless there is evidence that they have both psychometric and performance validity. We hypothesized that these may be valid instruments for the measurement of ulcerative colitis. Repeat measurements of disease activity were obtained in patients who participated in our previous study to determine the psychometric and performance validity of these two indices in ulcerative colitis. This quantitative approach can be applied to any index of disease activity in ulcerative colitis, and can be used to compare the psychometric and performance validity of different indices.

MATERIALS AND METHODS

Subject population and study design

This was a longitudinal cohort study of patients with ulcerative colitis, with repeat measurements of their disease activity, as in a clinical trial. The patient population and procedures have been described

Table 1. Structure and previous validation of non-endoscopic indices for ulcerative colitis

Index	Simple Clinical Colitis Activity Index ³	Seo Index ⁵
Components	Six questions: Day stool frequency Night stool frequency Urgency Stool blood General well-being Extraintestinal	Two questions: Stool Frequency Stool blood Three laboratory tests: Erythrocyte Sedimentation Rate Haemoglobin Albumin
Psychometric validity		
Content	None	None
Construct	None	None
Criterion convergent	Correlated with St Mark's Index and Seo Index	Correlated with both Truelove and Witt's classification and endoscopic findings ^{22, 23}
Criterion predictive	Predicted patient clinical relapse ¹⁶	Predicted need for colectomy ²⁴
Performance validity		
Reproducibility	None	None
Responsiveness	None	None

previously.^{10, 11} Briefly, 74 consecutive patients who were undergoing endoscopy and had ulcerative colitis were identified. Seventy consented to participate, and four were excluded for either concomitant severe illnesses or protocol violations in data collection. Measurement of St Mark's Index,² the Ulcerative Colitis Disease Activity Index,⁸ the Inflammatory Bowel Disease Questionnaire (IBDQ),¹² the SCCAI³ and the Seo Index⁵ were performed on the day of endoscopy in these 66 patients. Fifty-six of these subjects were able to return 3–12 months later for repeat measurement of the SCCAI and Seo Indices. In addition to the indices, patients were asked if they were in remission at each visit. Remission was defined as the patient's answer to the question: 'Has your ulcerative colitis been in remission (not active) over the past week?'. At the return visit, subjects were also asked if their ulcerative colitis was better or worse on a seven-point scale (1 = much better; 2 = somewhat better, 3 = a little better, 4 = about the same, 5 = a little worse, 6 = somewhat worse, 7 = much worse [Figure 1]). Significant clinical improvement was defined as a score of 1–2 on this scale. Stable disease was defined as a score of 3–5 on this scale.

Statistical analysis

The psychometric validity was determined by analysing the four components of psychometric validity in a quantitative manner on a 0 (invalid) to 1 (perfectly

valid) scale. Content validity was evaluated by the fraction of the identified domains in ulcerative colitis disease activity (from our previously published factor analysis)¹¹ that were included in each index. For construct validity, we determined the Spearman correlation of the non-endoscopic indices with the bowel symptom subscore of the validated IBDQ.^{13, 14} As higher scores in the IBDQ correlate with less disease activity, we multiplied the correlation by -1 to obtain a positive number between 0 and 1. For criterion-convergent validity, we determined the Spearman correlation of the SCCAI and Seo Index with the original index for ulcerative colitis, St Mark's Index, administered at the same session. For criterion-predictive validity, we measured the area under the receiver-operating characteristic (ROC) curve (the AUROC) for the non-endoscopic indices to predict patient-defined remission, which has a 0 to 1 scale.

Performance validity was determined by measuring reproducibility and responsiveness in quantitatively on a 0 (invalid) -1 (perfectly valid) scale. For reproducibility, we used previously determined bounds of significant change in each index (± 2.5 points for the SCCAI, ± 15 points for the Seo Index) to identify bounds for significant change. We then determined what fraction of the subjects who reported no significant change (Likert improvement scale of 3–5) had a change in their disease activity scores that indicated no change in their disease activity. For responsiveness, we calculated the Spearman correlation between the

change in disease activity and the patients' score on our Likert improvement scale.

The validity scores are all reported on a 0 to 1 scale. An index that is completely invalid would have a validity score of 0, and a perfectly valid index should have a validity score of 1. For the purposes of interpretation, we defined *a priori* that a validity score of 0–0.19 indicated 'poor' validity, 0.2–0.39 'fair' validity, 0.4–0.59 'moderate' validity, 0.6–0.79 'good' validity, and 0.8–1.0 'excellent' validity. All statistical calculations were performed with Stata 9.0 (College Station, TX, USA). The study protocol was approved by the University of Michigan IRB-MED Institutional Review Board (NIH Assurance no. M-1184 on 13 November 2002, IRB no. 2002-0801).

RESULTS

The SCCAI and the Seo Index have incomplete content validity

Our previous factor analysis of four ulcerative colitis disease activity indices identified five domains in the

measurement of ulcerative colitis.¹¹ These include stool frequency, stool blood, abdominal symptoms, laboratory biomarkers and temperature (Table 1). We determined the content validity of the SCCAI and the Seo Index by assessing what fraction of these five domains was included in each index. Table 2 shows that both non-endoscopic indices contain the two most critical domains (stool frequency and stool blood), but that each lacks two domains. This produces a good content validity score of 0.6 on a 0–1 scale for both indices.

The SCCAI, but not the Seo Index, has good construct validity

The construct validity was calculated on a 0–1 scale by comparing the two non-endoscopic indices with the bowel symptom subscore of the validated IBDQ. This quality-of-life measure has been validated in a variety of settings,¹⁵ and the bowel symptom subscore is based on the same disease construct as the disease activity measures for ulcerative colitis. As these distributions are skewed, the Spearman (nonparametric)

1 Much better	2 Somewhat better	3 A little better	4 About the same	5 A little worse	6 Somewhat worse	7 Much worse
------------------	----------------------	----------------------	---------------------	---------------------	---------------------	-----------------

Figure 1. Likert Improvement Scale: Subjects at their return visit were asked to rate their ulcerative colitis symptoms in relation to their previous visit on a seven-point Likert improvement scale. For the purposes of data analysis, scores of 1 or 2 (white region) corresponding to much improved or somewhat improved were considered clinically significant improvement. Scores of 3, 4 or 5 (light grey region) were considered 'no significant change' in symptoms.

Inherent factor	% of Variance	Present in SCCAI?	Present in Seo Index?
Stool frequency	62	Yes, 0.2	Yes, 0.2
Stool blood	10	Yes, 0.2	Yes, 0.2
Abdominal symptoms	8	Yes, 0.2	No
Laboratory tests	5	No	Yes, 0.2
Temperature	Unique factor	No	No
Total content validity (fraction of factors)	–	0.6	0.6

Table 2. Content validity of non-endoscopic indices is incomplete

Our previous factor analysis identified four major factors and one unique item (temperature) in the measurement of disease activity in ulcerative colitis by standard indices (St Mark's Index, Ulcerative Colitis Disease Activity Index, SCCAI and Seo Index). For each of the five inherent factors, we designated in each cell whether this factor is included in each index, and added 0.2 (fraction of the whole) to the content validity score.

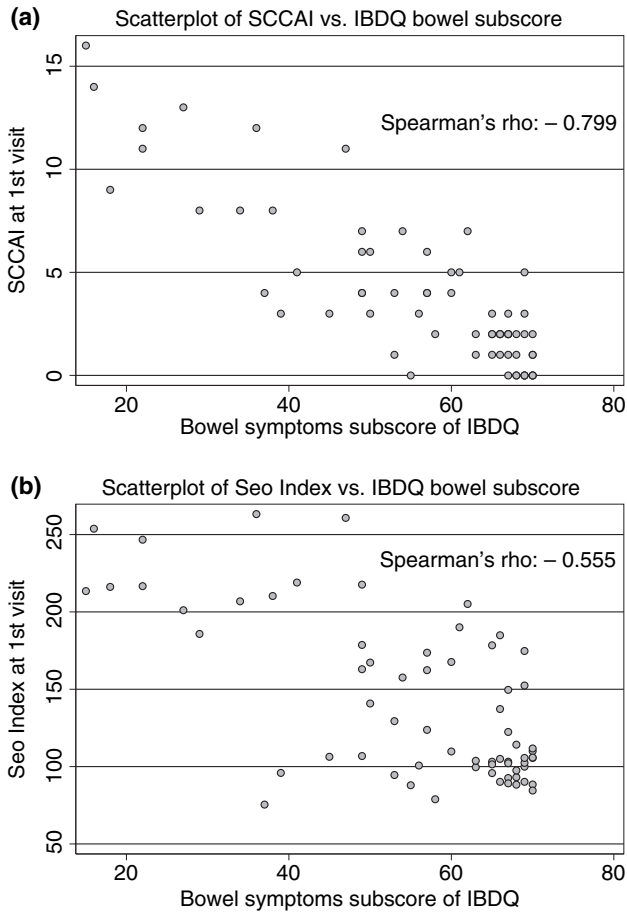


Figure 2. Construct validity: Scatter plots of the non-endoscopic indices vs. the IBDQ bowel subscore at the first visit. (a) The scatter plot of the SCCAI vs. the IBDQ bowel subscore is shown, with a Spearman correlation of -0.799 . (b) The scatter plot of the Seo Index vs. the IBDQ bowel subscore is shown, with a Spearman correlation of -0.555 . For the calculation of validity scores, both correlations were multiplied by -1 to obtain a positive value between 0 and 1.

correlation of each index with the bowel symptom subscore of the IBDQ was measured. In each case, the correlation was negative, as the IBDQ score rises with higher quality of life, and the disease activity scores become lower as disease activity improves. These correlations are illustrated in Figure 2. Spearman's rho for the SCCAI was -0.799 (95% CI: -0.690 to -0.872), and for the Seo Index, -0.555 (95% CI: -0.361 to -0.702). In order to convert these to a 0–1 scale, both correlations were multiplied by -1 to produce construct validity scores of 0.799 (good) and 0.555 (fair), respectively.

Both the SCCAI and Seo Index have good criterion-convergent validity

In order to measure the criterion-convergent validity, we compared the non-endoscopic indices with the most established endoscopic index, St Mark's Index. The Spearman correlation of the non-endoscopic indices was calculated, and was found to be 0.866 (95% CI: 0.789 to 0.916) (excellent) for the SCCAI, and 0.705 (95% CI: 0.558 to 0.809) (good) for the Seo Index. These correlations are illustrated in Figure 3. Each correlation is used directly as a value on a 0–1 scale for the criterion-convergent validity score.

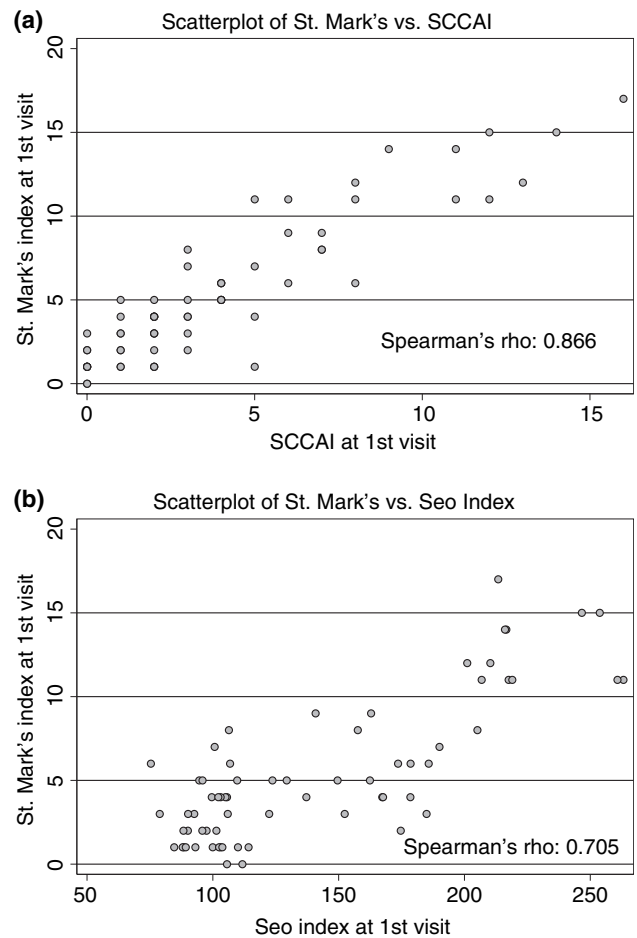


Figure 3. Criterion-convergent validity: scatter plots of the St Mark's Index vs. the non-endoscopic indices at the first visit. (a) The scatter plot of the St Mark's Index vs. the SCCAI is shown, with a Spearman correlation of 0.866. (b) The scatter plot of the St Mark's Index vs. the Seo Index is shown, with a Spearman correlation of 0.705.

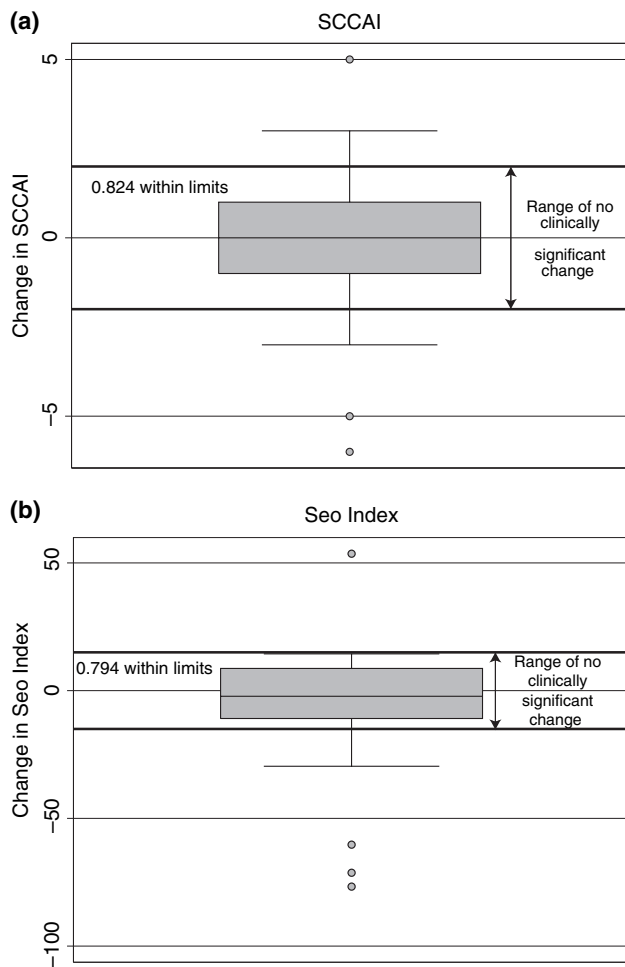


Figure 4. Reproducibility: box plots of non-endoscopic indices vs. previously defined range of no significant change. (a) The box plot of the SCCAI is shown with heavy lines demarcating the limits of no significant change (± 2). Outliers are shown as grey dots, and 0.824 of the measurements taken in patients who reported Likert improvement scores of no significant change (3, 4 or 5) fall within these limits. (b) A similar box plot of the Seo Index is shown with heavy lines demarcating the limits of no significant change (± 15). Outliers are shown as grey dots, and 0.794 of the measurements taken in patients who reported Likert improvement scores of no significant change (3, 4 or 5) fall within these limits.

Both the SCCAI and Seo Index have good criterion-predictive validity

The most important endpoints in clinical trials are the prediction of clinically significant improvement and of clinical remission. From a pragmatic point of view, patients are the ultimate arbiters of whether a treatment is effective, as they will choose to use or discon-

tinue a therapy based on whether they have achieved remission. Therefore we used the patient's own assessment of their status as the gold standard for clinical outcomes. We have previously assessed the sensitivity and specificity of these indices for patients with ulcerative colitis, using the gold standard of patient-determined remission. Both indices have been shown to be able to predict remission. In order to produce a single score on a 0–1 scale, the AUROC curve for clinical remission was calculated for both the SCCAI (0.911, excellent) and the Seo Index (0.920, excellent). These were used directly as the criterion-predictive validity scores.

Both the SCCAI and Seo Index have good reproducibility

In order to assess reproducibility, patients self-assessed the change in their disease status since their previous visit. For patients who had no significant change in their disease severity (a score of 3–5 on the Likert Improvement Scale), we determined whether the non-endoscopic indices would be able to identify these patients with stable disease. In our previous study,¹⁰ we were able to define cutoffs that identify significant improvement in disease activity in each index (change of >2.5 points in the SCCAI, change of >15 points in the Seo Index). In Figure 4 we illustrate with a box plot the distribution of the change in patient scores in the SCCAI and the Seo Index, and calculate what fraction of patients with stable disease have changes in their non-endoscopic indices that lie within this range of non-significant change. A fraction of 0.824 (excellent) of the patients with stable disease were correctly identified by the SCCAI cutoffs, and 0.794 (good) of the patients with stable disease were correctly identified by the Seo Index cutoffs. These were used directly as the reproducibility validity scores.

Both the SCCAI and Seo Index have moderate responsiveness

Disease activity indices in clinical trials must be able to accurately identify changes in disease activity, yet none of the existing indices has been tested for this attribute. We compared the change in the disease activity indices with the patients' self-assessment of change in disease status on the Likert Improvement Scale. This is presented in box plots in Figure 5. Notably, none of our patients rated themselves as

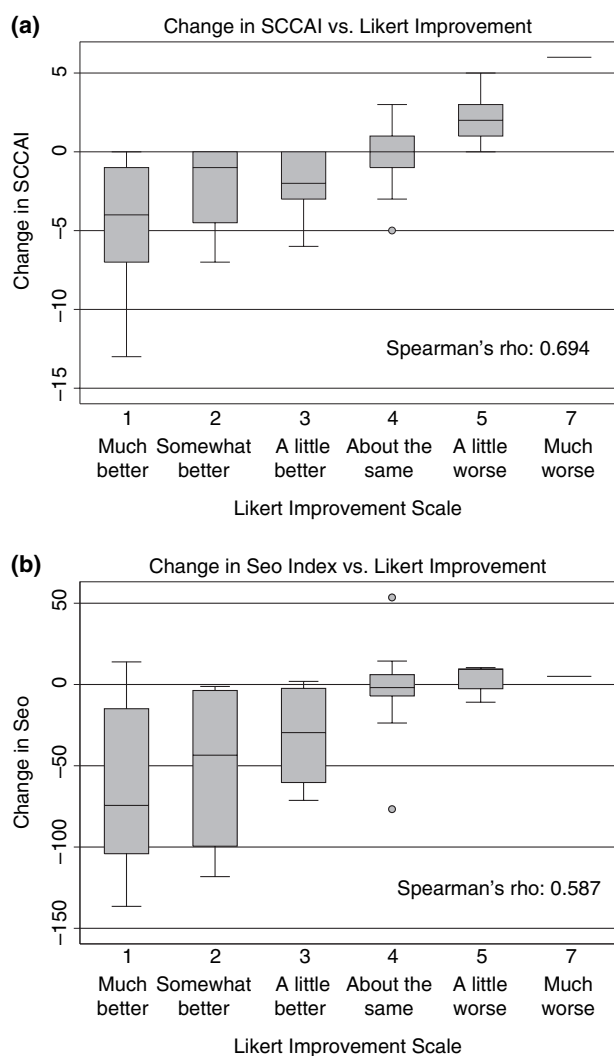


Figure 5. Responsiveness: categorical box plots of non-endoscopic indices vs. Likert Improvement Scale Ratings. (a) Box plots of the change in SCCAI categorized by Likert Improvement Scale score are shown, demonstrating a monotonic increase in the change in SCCAI with worsening change over time. The Spearman correlation is 0.694. Note that no subjects responded with 6, Somewhat Worse, on the Likert improvement scale. (b) Box plots of the change in the Seo Index categorized by Likert Improvement Scale score are shown, demonstrating a monotonic increase in the change in the Seo Index with worsening change over time. The Spearman correlation is 0.587.

'6 – somewhat worse'. We calculated Spearman correlations between the change in the indices and the Likert Improvement Scale. We found a correlation of 0.694 (95% CI: -0.587 to -0.809) (good) for the SCCAI and 0.587 (95% CI: -0.383 to -0.736) (fair) for the

Seo Index. These were used directly as the responsiveness validity scores.

Psychometric and performance validity of the SCCAI and Seo Index

We illustrate the psychometric validity and performance validity of each component of each index in Figure 6a,b. Both the SCCAI and the Seo Index have moderate to excellent psychometric validity. The performance validity of the SCCAI is good to excellent, while the performance validity of the Seo Index is moderate to good.

DISCUSSION

In this study, we developed a quantitative method to measure the psychometric and performance validity of two non-endoscopic disease activity indices for ulcerative colitis. This methodology also allows investigators to identify specific weaknesses in components of psychometric or performance validity. We found that both non-endoscopic indices have fair to excellent psychometric validity, while the SCCAI has better performance validity than the Seo Index. This is the most rigorous evaluation to date of the validity of disease activity indices in ulcerative colitis. The validity of other disease activity indices for ulcerative colitis remains unknown, and at this point, the SCCAI and the Seo Index have the best documented validity of any ulcerative colitis disease activity indices. This rigorous testing justifies the use of these non-endoscopic indices in clinical trials. Based on the results of this study, the SCCAI should be favoured over the Seo Index for measurement of disease activity in longitudinal clinical trials.

An additional benefit of this study is that it specifically identifies the weaknesses of the current indices. The SCCAI is somewhat lacking in content validity and responsiveness. The content validity could be addressed by adding items to measure the missing domains (laboratory tests and temperature) to the SCCAI, and further research is needed to improve the responsiveness of the SCCAI. It may be that the current response scales to the SCCAI questions do not have enough gradations to detect small changes in disease activity, and improving these scales might improve responsiveness.

The Seo Index is lacking in content validity, construct validity and responsiveness. The content validity could be improved by adding items to address the missing domains. Additional detailed questions about

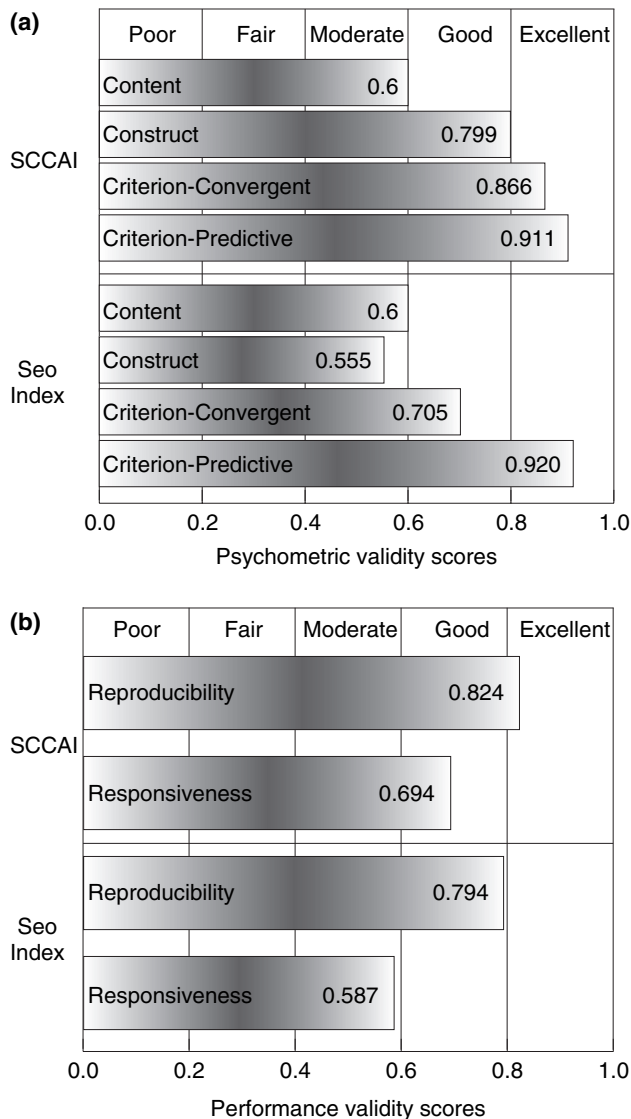


Figure 6. Summary of the Psychometric and Performance Validity Scores for the SCCAI and the Seo Index. (a) The four components of psychometric validity for both the SCCAI and Seo Index are summarized here on a 0–1 scale. The SCCAI has a weakness in content validity. The Seo Index has weaknesses in content validity, construct validity, and criterion-convergent validity. (b) The two components of performance validity for both the SCCAI and Seo Index are summarized here on a 0–1 scale. The SCCAI has only good responsiveness, while the responsiveness of the Seo Index is only fair.

bowel symptoms might improve the construct validity, as the Seo Index has only one question about bowel frequency. The responsiveness might also be improved by additional gradations in the responses to symptom questions. Despite its weaknesses, the Seo Index is

remarkable for its innovative use of laboratory tests as simple biomarkers. It is probable that the addition of biomarkers, which are not found in other disease activity indices for ulcerative colitis, causes the Seo Index to have good criterion-convergent and criterion-predictive validity.

Our data support the findings of Jowett *et al.*,¹⁶ who found that the SCCAI has good criterion-predictive ability to predict which patients are in relapse. They also support the findings of Seo *et al.*^{22,23,24}, who found that the Seo Index has good criterion-convergent validity with St Mark's Index and with endoscopic findings, and that the Seo Index has good criterion-predictive validity in its ability to predict which patients would require colectomy.

This study is limited in that we only assessed patients at a single tertiary care centre, and the results may not be generalizable to clinical study subjects in other centres, and particularly to subjects for whom English is not their first language. In order to maximize the generalizability of these findings, we deliberately included subjects with a wide range of disease activity, such as those undergoing colonoscopic surveillance, patients having colonoscopy for symptoms, and inpatients who were quite ill. A second limitation is that the comparators chosen for determining criterion-convergent validity and construct validity have not themselves been validated. St Mark's index has never been validated, and while the entire IBDQ has been validated, the individual subscores have not. This is an inherent limitation in the evaluation of psychometric validity. Future developments in the measurement of ulcerative colitis may identify better comparators for criterion-convergent validity and construct validity in ulcerative colitis.

Another limitation is the use of a definition of remission as determined by measurement of disease activity at one point in time. While this is currently the standard in clinical trials in ulcerative colitis, this approach has been criticized as inappropriate in an inherently waxing and waning disease. Some [including the Food and Drug Administration (FDA)] have advocated 'durable clinical remission'. This is a relatively new concept, and durable clinical remission does not have a consensus definition. This variability is an inherent problem in diseases with a waxing and waning course. Given the inherent variability in ulcerative colitis, perhaps durable clinical remission should be defined as remission upon repeated measurement over an extended period of time. This would add substan-

tially to the costs and difficulty of conducting clinical trials. If this were required, this would make endoscopic indices particularly unattractive, as repeated lower endoscopy is expensive and avoided by subjects. An alternative approach would be to identify levels of symptoms, biomarkers or other tests (perhaps including mucosal healing) that truly predict which subjects will have durable clinical remission over the next 6 months. Future research may identify factors that can be proven to be accurate predictors of durable clinical remission.

The definitions of clinical remission and significant clinical improvement used in this study were those of the subjects with ulcerative colitis. These are used because there is no gold standard for clinical remission in ulcerative colitis. Alternatives have been proposed, including biomarkers of inflammation, physician assessment, imaging methods, endoscopic healing or a combination of these. While one can argue that these may be more objective, none of these outcomes matters if the patient does not feel well. Biomarkers can be manipulated with biologics, but a low C-reactive protein (CRP) does not equal health. Other manipulations (i.e. topical therapy) could improve the appearance of imaging or the appearance of the colonic mucosa, but if the patient does not feel well, we have not treated the patient, but only a manifestation of the disease. From a clinical perspective, we must treat the patient, not numbers or images, so we must use the patient as a gold standard to identify which factors are valuable objective predictors of patient outcome.

For many clinicians, performance validity is more important than psychometric validity. If the psychometric validity is reasonably good, the performance validity is the critical evaluation of an instrument. In clinical trials, it is very important that the measurement tool be sensitive enough to detect clinically important changes. It is also critical that the tool be reproducible and stable enough that patients with little or no improvement have stable scores on the measurement instrument, to avoid increasing the placebo rate.

Presently, the SCCAI is the best validated index available for ulcerative colitis. The Seo Index demonstrates the value of biomarkers, as this index is able to predict clinical outcomes despite its demonstrated weaknesses in validity. The results of this study suggest that a better disease activity index for ulcerative colitis can be developed. A novel index that combined the questions of the SCCAI, the biomarker laboratory tests of the Seo Index, and a temperature item would

have significantly improved the content validity. Adding more response levels to the questions might improve the responsiveness of the novel index further. The biomarkers of the Seo Index might be improved upon with newer biomarkers, including C-reactive protein,¹⁷⁻¹⁹ faecal lactoferrin^{20, 21} or faecal calprotectin. The role of endoscopic and histologic findings, particularly as part of the current endoscopic indices, also needs to be further defined. When the value of these potential measurement items is determined, a reduced panel of symptom items, biomarkers, and possibly endoscopy or histology will probably yield an improved, valid disease activity index for ulcerative colitis. Our group is currently developing and testing new survey questions and biomarkers for ulcerative colitis for this purpose, with the support of the Crohn's and Colitis Foundation of America.

In this report, we show that the SCCAI is the most rigorously validated index in ulcerative colitis, and that it has good psychometric and performance validity. Its use should be strongly considered for longitudinal clinical trials in ulcerative colitis until improved ulcerative colitis disease activity indices are developed with better validity scores. The quantitative methodology we have introduced in this manuscript allows the direct comparison of the validity of different indices, identifies specific weaknesses in the indices for remediation, and provides a metric for evaluating the validity of future disease activity indices in ulcerative colitis and in other disease states. This methodology can bring a rigorous approach to the development, validation and improvement of future disease indices in a wide range of disease states.

ACKNOWLEDGEMENTS

Dr Higgins is supported by the NIH K12 RR017607-01 and the AGA Centocor Excellence in IBD Clinical Research Award. Dr Zimmermann is supported by the NIH R01 DK-56750-01.

We would like to thank Sheryl Korsnes for her assistance in recruiting patients, and Jack Kalbfleisch and Brenda Gillespie for their statistical assistance.

Peter D.R. Higgins has received research funding from the Otsuka America Pharmaceutical Corporation, the Crohn's and Colitis Foundation of America, the NIH K12 RR017607-01 Award, and the AGA Centocor Excellence in IBD Clinical Research Award. Ellen M. Zimmermann has received research funding from the NIH R01 DK-56750-01 Award.

REFERENCES

- 1 Truelove SC, Witts LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *Br Med J* 1955; **4947**: 1041–8.
- 2 Powell-Tuck J, Bown RL, Lennard-Jones JE. A comparison of oral prednisolone given as single or multiple daily doses for active proctocolitis. *Scand J Gastroenterol* 1978; **13**: 833–7.
- 3 Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998; **43**: 29–32.
- 4 Lichtiger S, Present DH, Kornbluth A, *et al.* Cyclosporine in severe ulcerative colitis refractory to steroid therapy. *N Engl J Med* 1994; **330**: 1841–5.
- 5 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.
- 6 Azzolini F, Pagnini C, Camellini L, *et al.* Proposal of a new clinical index predictive of endoscopic severity in ulcerative colitis. *Dig Dis Sci* 2005; **50**: 246–51.
- 7 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.
- 8 Sutherland LR, Martin F, Greer S, *et al.* 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. *Gastroenterology* 1987; **92**: 1894–8.
- 9 Rutgeerts P, Sandborn WJ, Feagan BG, *et al.* Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005; **353**: 2462–76.
- 10 Higgins PD, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. *Gut* 2005; **54**: 782–8.
- 11 Higgins PD, 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.
- 12 Irvine EJ. Quality of life – rationale and methods for developing a disease-specific instrument for inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1993; **199**: 22–7.
- 13 Guyatt G, Mitchell A, Irvine EJ, *et al.* A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 1989; **96**: 804–10.
- 14 Irvine EJ, Feagan B, Rochon J, *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.
- 15 Pallis AG, Mouzas IA, Vlachonikolis IG. The inflammatory bowel disease questionnaire: a review of its national validation studies. *Inflamm Bowel Dis* 2004; **10**: 261–9.
- 16 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.
- 17 Vermeire S, Assche GV, Rutgeerts P. The role of C-reactive protein as an inflammatory marker in gastrointestinal diseases. *Nat Clin Pract Gastroenterol Hepatol* 2005; **2**: 580–6.
- 18 Solem CA, Loftus EV Jr, Tremaine WJ, Harmsen WS, Zinsmeister AR, Sandborn WJ. Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. *Inflamm Bowel Dis* 2005; **11**: 707–12.
- 19 Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. *Inflamm Bowel Dis* 2004; **10**: 661–5.
- 20 Kane SV, Sandborn WJ, Rufo PA, *et al.* Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. *Am J Gastroenterol* 2003; **98**: 1309–14.
- 21 Sugi K, Saitoh O, Hirata I, Katsu K. Fecal lactoferrin as a marker for disease activity in inflammatory bowel disease: comparison with other neutrophil-derived proteins. *Am J Gastroenterol* 1996; **91**: 927–34.
- 22 Seo M, Okada M, Maeda K, Oh K. Correlation between endoscopic severity and the clinical activity index in ulcerative colitis. *Am J Gastroenterol* 1998; **93**: 2124–9.
- 23 Seo M, Okada M, Yao T, Okabe N, Maeda K, Oh K. Evaluation of disease activity in patients with moderately active ulcerative colitis: comparisons between a new activity index and Truelove and Witts' classification. *Am J Gastroenterol* 1995; **90**: 1759–63.
- 24 Seo M, Okada M, Yao T, Mataka H, Maeda K. Evaluation of the clinical course of acute attacks in patients with ulcerative colitis through the use of an activity index. *J Gastroenterol* 2002; **37**: 29–34.