



Published in final edited form as:

Inflamm Bowel Dis. 2008 December ; 14(12): 1660–1666. doi:10.1002/ibd.20520.

Use of the Non-invasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis

James D. Lewis, M.D., M.S.C.E.^{1,2,3,4}, Shaokun Chuai, MS^{2,4}, Lisa Nessel, MSS, MLSP², Gary R. Lichtenstein, MD^{1,3}, Faten N. Aberra, MD, MSCE^{1,3}, and Jonas H. Ellenberg, PhD^{2,4}

¹ Division of Gastroenterology, University of Pennsylvania School of Medicine, Philadelphia, PA

² Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA

³ Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA

⁴ Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA

Abstract

Background—The Mayo score and a non-invasive 9-point partial Mayo score are used as outcome measures for clinical trials assessing therapy for ulcerative colitis. There are limited data assessing what defines a clinically relevant change in these indices. We sought to assess what constitutes a clinically meaningful change in these indices using data from a recently completed placebo-controlled clinical trial.

Methods—105 patients were enrolled in a 12 week randomized, placebo-controlled trial assessing rosiglitazone for treatment of mild to moderate ulcerative colitis. We compared the change in the Mayo score, the partial Mayo score, and a 6 point score composed just of the stool frequency and bleeding components of the Mayo score to the patient's perception of disease activity at week 0 and week 12. Optimal cut points were calculated as the maximal product of sensitivity and specificity.

Results—Each index was strongly correlated with the patient's rating of disease activity at week 12 (Spearman correlations from 0.61 to 0.71, $p < 0.0001$ for all correlations). The maximal product of sensitivity and specificity to identify patient reported improvement of disease activity was achieved using cut points for change of 2.5 for the Mayo score (sensitivity 88%, specificity 80%), 2.5 for the partial Mayo score (sensitivity 88%, specificity 87%), and 1.5 for the 6 point score (sensitivity 88%, specificity 80%).

Conclusion—The partial Mayo score and the 6 point score composed solely of the stool frequency and bleeding components performed as well as the full Mayo score to identify patient perceived clinical response.

Introduction

Ulcerative colitis is a chronic relapsing disease characterized by inflammation of the colon. Randomized controlled trials designed to study the efficacy of therapeutic interventions

Address correspondence and reprint requests to: James D. Lewis, MD, MSCE, University of Pennsylvania, Center for Clinical Epidemiology and Biostatistics, 7th Floor Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, Tel: 215-573-5137, Fax: 215-573-0813, email: jlewis@cceb.med.upenn.edu.

The authors report no potential conflicts of interest related to this study.

Clinical trial registration: clinicaltrials.gov #NCT00065065

require outcome measures that characterize the activity of the disease. Such measures frequently combine findings from invasive tests (sigmoidoscopy or colonoscopy) and clinical characteristics of the patient.¹

Recruitment and retention of patients in clinical trials for ulcerative colitis is challenging. One factor that may deter patients is the need for multiple endoscopies. As such, clinical trials often only include a lower endoscopy at the time of randomization and again at the time of the primary endpoint. When such studies rely on a disease activity measure that includes data from lower endoscopy, the index can only be calculated at the time points where the endoscopy occurred. However, it is often beneficial to understand the pattern of disease activity across the entire study period.

Recent data from one institution suggest that use of invasive tests (i.e., lower endoscopy) may add little to the predictive value of the same instruments using only the non-invasive data or non-invasive measurements.^{2, 3} As such, it is appealing to use the non-invasive components of disease activity measures for interim evaluations in clinical trials.^{4, 5}

The Mayo score and the nearly identical Disease Activity Index (DAI) described by Sutherland et al. are two of the most commonly used activity indices in placebo-controlled clinical trials for ulcerative colitis.⁶ Each is composed of four categories (bleeding, stool frequency, physician assessment, and endoscopic appearance) rated from 0–3 that are summed to give a total score that ranges from 0–12. Several recent trials have utilized the non-endoscopic components of these indices as a partial index at interim visits to assess the time to onset of response.^{4, 5, 7} Higgins reported that a score of less than 2.5 on the full index has the optimal sensitivity and specificity for a patient-defined remission.² However, it is unknown what amount of change in the partial index reflects a clinically meaningful change from the perception of the patient, which is the focus of this study. We have used data from our recently completed placebo-controlled randomized trial of rosiglitazone for mild to moderately active ulcerative colitis to define the operating characteristics of the partial Mayo score using patients' rating of disease activity and change in disease activity as the gold standard.

Methods

Data for this study were acquired from a recently completed placebo-controlled randomized trial of rosiglitazone for mild to moderately active ulcerative colitis (clinicaltrials.gov #NCT00065065) which has been described in greater detail previously.⁷ The trial used a slight modification of the Mayo score to assess disease activity (Table 1). Specifically, the bleeding component as described in the Mayo index was modified such that a score of 3 required both visible blood in 50% or more of bowel movements and at least some bowel movements with blood alone.

The study included 105 patients with mild to moderately active disease defined as a total DAI score of 4 to 10, inclusively. Patients were randomized in a 1:1 ratio to receive either rosiglitazone 4 mg or placebo twice daily for 12 weeks. Disease activity was measured at randomization and every four weeks thereafter until week 12, however lower endoscopy was only completed at week 0 and week 12, such that only a partial Mayo score (9 point scale that excludes the endoscopic appearance of the mucosa) could be calculated at the interim visits. In the very early accrual period of the study, a follow-up visit was included at week 2. Without knowledge of the response rates in either arm, the Data and Safety Monitoring Board (DSMB) requested that the week 2 follow-up evaluation be eliminated with the hopes of minimizing the placebo response rate and maximizing recruitment and retention.^{6, 8, 9} Eighteen patients completed the week 2 follow-up visit.

During the course of the study, patients could be treated with other conventional medications used to treat active ulcerative colitis including mesalamine, oral corticosteroids, immunomodulators, or topical therapies (mesalamine or corticosteroids) at stable doses. Use of corticosteroids at doses greater than 20mg per day of prednisone or the equivalent was an exclusion criterion. Steroid tapering was not permitted during the study.

In anticipation of this sub-study, at each visit we also included questions about change in disease activity compared to the previous visit and compared to the randomization visit on a global seven-point scale (Table 2). The choices included much better, moderately better, a little better, unchanged, a little worse, moderately worse, and much worse. Patients also graded their current disease activity at each visit on a 6 point Likert scale – perfect, very good (minimal symptoms), good (only mild symptoms), moderately active, moderately severe, or severe. Data on quality of life were measured with the Inflammatory Bowel Disease Questionnaire (IBDQ) authored by Dr. Jan Irvine under license from McMaster University, Hamilton, Canada.¹⁰

Statistical analyses

All analyses were conducted pooling patients in both arms of the clinical trial. Descriptive statistics were used to describe patients included in the study. Continuous variables are reported as medians and interquartile ranges and categorical variables as proportions. Correlations were measured using Spearman correlation coefficients (ρ). Except where noted, only data for patients who completed the week 12 study visit were analyzed; this allowed for greater variation in disease activity than at baseline, since at randomization all patients were required to have DAI scores between 4 and 10, inclusive.

To assess sensitivity and specificity for clinical remission and clinical response, we relied on the patient's assessment of disease activity in the preceding 24 hours at randomization and at the week 12 visit. Clinical remission was defined as a self-assessment of perfect or very good (minimal disease activity). Clinical response required improvement by at least two points on the 6 point Likert scale. Receiver operating characteristics (ROC) curves were generated and the C-statistic was calculated as a summary measure of the discriminative properties of the indices. Optimal cut points were identified by the highest sensitivity \times specificity product. Area under the ROC curves for different disease indices were compared using the `roccomp` command in Stata v10 (Stata Corp, College Station, TX).

Because neither the patient rating nor the Mayo score are a pure gold standard for disease activity, we also calculated kappa statistics across a range of cut points to assess the impact of changing cut points on the agreement between the patient ratings and partial Mayo score. The kappa statistic measures the degree of agreement between the two measures beyond that which would be expected by chance. The kappa statistic can have values ranging from -1 to 1 , with values of $.41-.60$ representing moderate agreement, $.61-.80$ representing substantial agreement, and values greater than $.80$ representing almost perfect agreement.¹¹

Finally, because the physician's global assessment is likely greatly driven by the patient's report of bowel movement frequency and bleeding, we examined whether a score composed exclusively of these two factors would perform as well as the modified DAI. This six-point scale was compared to the patient's self report in the same manner described for the full and partial Mayo scores. We refer to this as the "6 point scale" in this report.

Results

105 patients were included in the study, of whom 52 and 53 were randomized to receive rosiglitazone and placebo, respectively. The median disease activity at baseline was 7.0 (of a

maximum potential of 12 points) on the full Mayo score scale. Concomitant therapy with 5-ASA, oral corticosteroids, and immunomodulators was used by 78.1%, 28.6%, and 26.7% of the patients, respectively.

Correlation of change in the Mayo Score with patient assessment of change in disease activity

Seventy-five patients completed the week 12 visit. Change in disease activity from randomization to week 12 of therapy as measured by the full Mayo score correlated well with the patient's perception of change in disease activity whether measured with a single question at week 12 ($\rho=0.60$, $p<0.0001$) or examining the change in response to a single question about disease activity in the previous 24 hours (perfect, very good, good, moderately active, moderately severe, or severe) completed at randomization and again at the week 12 visit ($\rho=0.65$, $p<0.0001$). All subsequent analyses of change in disease activity were based on response to a single question about disease activity in the previous 24 hours completed at randomization and again at the week 12 visit.

Correlation of patient assessment of disease activity with the full and partial Mayo scores and the 6 point scale

Patient assessment of disease activity was strongly correlated with the full Mayo score at week 12 ($\rho=0.71$, $p<0.0001$) (Table 3). The correlation of the partial Mayo score with the patient assessment of disease activity at week 12 was essentially identical to that of the full Mayo score ($\rho=0.70$, $p<0.0001$). The six point scale that included only the bowel movement frequency and bleeding score was highly correlated with 9 point partial Mayo score ($\rho=0.96$, $p<0.0001$), the 12 point full Mayo score ($\rho=0.88$, $p<0.0001$), and the patient's assessment of disease activity ($\rho=0.61$, $p<0.0001$) at the week 12 visit. The physician's rating of disease activity component of the Mayo score correlated well with the patient's rating of disease activity at week 12 ($\rho=0.64$, $p<0.0001$). The physician rated the disease activity as milder than the patient rated the disease in 4 of the 5 highly discordant pairs (data not shown). The IBDQ correlated a little less strongly with the full Mayo score ($\rho=-0.48$, $p<0.0001$) and the partial Mayo score ($\rho=-0.47$, $p<0.0001$). The correlation of the IBDQ with the patient assessment of disease activity ($\rho=-0.63$, $p<0.0001$) at the week 12 visit was similar to that between the physician assessment of disease activity and the patient assessment.

Sensitivity and specificity to identify remission

Of the 75 patients completing the 12 week visit, 17 rated their disease activity as perfect or very good. Using a cut point of 2.5 as suggested by Higgins,² the full Mayo score had sensitivity of only 41% but a specificity of 95% to identify patient defined remission. Increasing the cut point to 3.5 increased the sensitivity to 59%, but at the expense of specificity (84%). However, the maximal product of sensitivity \times specificity was observed at a cut point of 4.5 (88% and 78%, respectively). The product of sensitivity (71%) and specificity (84%) was maximized at a cut point of 2.5 using the 9 point partial Mayo score to identify patient-defined remission. For the 6 point scale, a cut point of 1.5 produced the maximal product of sensitivity (65%) and specificity of (81%). The area under the ROC curve was similar for the full Mayo score (0.85) and the partial Mayo score (0.85), but lower for the 6 point scale (0.78, $p=0.03$ for comparison across the three indices) (Table 4, Figure 1).

Sensitivity and specificity to identify clinical improvement

Sixteen of the 75 patients who completed the week 12 visit reported at least a 2 category improvement on the patient self rating scale from week 0 to week 12. In ROC curve

analysis, C statistics were similar ($p=0.48$) for the full Mayo score (0.89), the 9 point partial Mayo score (0.90), and the 6 point scale (0.89) (Table 5, Figure 2). Optimal cut points of change in the full Mayo score and change in the partial Mayo score were 2.5 points (sensitivity 88%, specificity 80%) and 2.5 points (sensitivity 88%, specificity 86%), respectively (Figure 2). Of note, a cut point of 1.5 for change in the partial Mayo score had identical sensitivity to the 2.5 cut point, but with slightly lower specificity (73%). Using the 6 point scale, the optimal cut point to identify a clinical response was 1.5 point change, with sensitivity of 88% and specificity of 80%.

Kappa statistic

As a final test of the reliability of our methods to identify an optimal cut point for change in the 9 point partial Mayo score to identify a clinical response, we computed Kappa statistics for various combinations of change in patient reported disease activity on the Likert scale from baseline to week 12 compared with various cut points for change in the partial Mayo score. The maximal Kappa statistic (0.65) was obtained using a 2 point change on the Likert scale (our a priori definition of response) and a 2.5 point change in the partial Mayo score, the optimal cut point identified in our ROC analysis.

Discussion

Measurement of disease activity in patients with ulcerative colitis is important for the conduct of any clinical trial. Because sigmoidoscopy and colonoscopy are expensive, relatively unpleasant, and can potentially worsen ulcerative colitis,¹² development of valid non-invasive measurements of disease activity are needed. At the same time, it is helpful to avoid collection of excessive and redundant data in clinical trials. We and others have previously used a modification of the Mayo score to follow disease activity over time by utilizing the non-invasive components. In this study, we have provided new information on the test operating characteristics of this modified and non-invasive measurement. Our results confirm that the addition of the endoscopy component adds little to the correlation of the Mayo score with patient-perceived clinical activity. In addition, this study establishes that the partial Mayo score and the 6 point score composed solely of the stool frequency and bleeding components performed as well as the full Mayo score to identify patient-perceived clinical response. The maximum product of sensitivity and specificity was achieved with reduction in the Mayo score by ≥ 3 points, the partial Mayo score by either ≥ 2 or ≥ 3 points, and the 6 point scale by ≥ 2 points. However, our results and those of Higgins also confirm that these indices do not correlate perfectly with patients' perception of disease activity and raise questions about the optimal definition of remission using the Mayo score.

Our results are complimentary to that obtained by Higgins et al.^{2, 3} The Higgins study was conducted on patients in routine clinical practice from a single center. Our data were collected as part of a multicenter clinical trial. We were surprised that we came to different cut points to maximize the product of sensitivity \times specificity for patient defined remission. Higgins identified an optimal cut point to identify remission of 2.5, while in our study, a cut point of 4.5 achieved the maximal product of sensitivity \times specificity. This could be due to the many differences in the two studies. The patient populations were likely quite different, with our patients participating in a clinical trial of active disease, having defined management of their colitis therapies, and being required to have mild to moderately active disease at the outset. In contrast, the Higgins study included patients in routine clinical practice. Another difference was the definition of self-reported remission. We defined this as having no or minimal symptoms where as Higgins asked the patient directly whether they were in "remission."

Despite these differences, it is noteworthy that the global performance of the Mayo score to discriminate patient-perceived active from inactive disease was relatively similar in the two studies. For example, at the point of maximal sensitivity \times specificity, the sensitivity and specificity in the Higgins study were 82% and 89%, respectively; in this study, the sensitivity and specificity were 88% and 78%, respectively. Likewise, the area under the ROC was 0.94 and 0.85, both suggesting fairly good discriminative ability of the measure. Thus, while we are uncertain of the optimal definition of clinical remission using the Mayo score, it appears reasonably accurate at discriminating patients who perceive themselves to have ongoing disease activity from those whose disease is clinically quiescent.

Several different definitions of clinical remission based on the Mayo score have been used in clinical trials, ranging from a Mayo score of 2 or less to a score of 0.6 The major reason in favor of using a more stringent definition of remission is to increase the specificity (i.e. minimize false positives). However, our results and those of Higgins suggest that specificity is very high using a Mayo score of 2 or less to define remission.² This suggests that the more stringent definition be limited to defining endoscopic remission.

As with remission, different definitions of response have been used, most commonly a reduction of the baseline Mayo score of either 2 or 3 points.⁶ The data from the analysis in this report suggest that to identify a clinically important improvement in disease activity, a 3 point or greater decrease in the Mayo score is the optimal cut point, with relatively high sensitivity (88%) and specificity (80%). Of note, Su et al. demonstrated that the rate of response in placebo arms of ulcerative colitis clinical trials is substantially higher when a response is defined as a 2 point decrease rather than a 3 point decrease.⁶ In our clinical trial, we observed the same. The response rate in the placebo arm was nearly double when response was defined as a 2 point decrease, while the response rate in the active treatment arm changed much less.⁷

A major limitation of clinical trials that rely on the invasive indices is the ability to monitor disease activity at multiple time points. One approach to this has been to incorporate the non-invasive IBDQ. However, in both our study and that of Higgins², the IBDQ appeared to be less well correlated with patient assessment of disease activity than the Mayo score or partial Mayo score.

Our results provide important guidance on how to interpret changes in partial Mayo score. To identify a clinical response, a 3 point or greater change resulted in the maximal product of sensitivity \times specificity, a finding that was confirmed in our assessment of the Kappa statistic at various combinations of cut points. However, it should be noted that these conclusions are based on small numbers of subjects and the sensitivity of a 2 point and 3 point change were identical. Similar studies of this nature are needed to confirm these results.

The data elements included in the partial Mayo score are routinely recorded during office visits and as such can be used in retrospective studies of ulcerative colitis. In addition, the bleeding and stool frequency components of the instrument are ideally suited for questionnaire-based observational studies. Interestingly, the scores of the 6 point scale using only the bleeding and stool frequency components correlated extremely well with the full Mayo score, the partial Mayo score, and the patient rating of disease activity, suggesting that these two questions are informative in assessing patient-perceived disease activity in a reproducible fashion.

Comparison of ROC curves demonstrated nearly identical results for the 12, 9, and 6 point scales to identify clinical response. However, the 6 point scale performed slightly less well for identifying clinical remission. Taken together, these data suggest that the 6 point scale

can fairly reliably measure disease activity and change in clinical status without the need for interaction with a physician, and as such could be a valuable tool for questionnaire-based observational research. Of note, the Simple Clinical Colitis Activity Index, which relies heavily on questions about stool frequency, bleeding, and urgency, does not require blood tests or endoscopy and can be self administered.¹³ However, this index is slightly more complicated, including questions about extraintestinal manifestations of disease. Future research is warranted to clarify whether these additional questions are important in characterizing remission and clinical response.

One might ask why the patient's assessment of disease activity that we used as the "quasi gold standard" in this study should not be used as the outcome measure in most trials. The patient's global perception of disease improvement is entirely subjective and likely at greater risk for bias. Previous research has shown that the "placebo effect" is greater in studies using subjective outcome measures which may explain the higher placebo remission rate.¹⁴ The patient-defined measure also has a limited range, particularly for patients with mild to moderate disease as were included in this study. In contrast, the Mayo score allows for finer separation of response and remission. Finally, and perhaps most importantly, the patient-defined response measure is more subject to variability due to factors unrelated to the ulcerative colitis activity, such as presence of increased stress, fatigue, duration of disease, prior history of mild versus severe disease course, etc.

Although the Mayo score, the partial Mayo score, and the 6 point scale all seem to discriminate responders from non-responders relatively well, each is imperfect. It is important to consider the implications of an imperfect outcome measure which results in misclassification bias. In a clinical trial, such misclassification should usually be nondifferential between the two treatment arms and will bias the results toward the null. Consider a hypothetical medication that produces a 40% response rate while placebo therapy results in a 20% response rate (i.e. risk difference of 20%). If the outcome measure has either sensitivity or specificity of only 80%, the measured risk difference will be only 16%. Accounting for imperfect sensitivity and specificity reduces the measured risk difference even further (e.g., sensitivity 80%, specificity 90% yielding a risk difference of approximately 14%). Such misclassification bias results in substantial loss of statistical power. Given the human and financial cost of conducting clinical trials, the need for precise outcome measures is obvious.

There are several important limitations of this analysis. First, it was based on a small sample size and from a single multicenter, clinical trial. Likewise, we focused on patients with mild to moderately active disease. It is possible that the indices examined in this study perform differently among patients with more severe ulcerative colitis. Likewise, the small sample size prevented us from examining how other factors, such as use of corticosteroids, antidepressant medications, or narcotics, influence the operating characteristics of these instruments. The small sample size and limited number of patients who had mucosal healing also prevented us from examining if the non-invasive indices can identify patients with mucosal healing, an outcome hypothesized to predict long term outcomes and of interest to regulatory authorities.¹⁵ As such, this study should not be interpreted to suggest that inclusion of endoscopic healing among outcomes in ulcerative colitis trials is unnecessary. Finally, patient-defined outcome measures have not been validated in terms of reproducibility and responsiveness to change. For all of these reasons, we believe that these findings should be tested for reproducibility in future clinical trials and observational studies.

In conclusion, utilizing data from a recently completed clinical trial, we have demonstrated that the partial Mayo score and the 6 point score composed solely of the stool frequency and

bleeding components performed as well as the full Mayo score to identify patient-perceived clinical response. In addition, these data provide further evidence that the Mayo score has reasonable test operating characteristics to distinguish patient-perceived active disease from quiescent disease and to identify a patient defined response (i.e. clinical improvement). Based on our data, clinical response may be best defined using a 3 point or greater decrease in the Mayo score or the partial Mayo score. Finally, for observational studies or studies involving data collection in the absence of direct patient contact, use of only the bleeding and stool frequency components of the Mayo score should provide reasonable estimates of disease activity and change in disease activity. Further refinement and validation of these indices is warranted in order to achieve optimal efficiency in ulcerative colitis research.

Acknowledgments

This study was supported by NIH grant DK059961, Drug and placebo were provided by GlaxoSmithKline. Use of the Inflammatory Bowel Disease Questionnaire, authored by Dr. Jan Irvine was made under license from McMaster University, Hamilton, Canada. The study was designed and conducted, the data were analyzed, and the manuscript was written independently by the investigators.

References

1. D'Haens G, Sandborn WJ, Feagan BG, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology*. 2007; 132:763–86. [PubMed: 17258735]
2. 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. [PubMed: 15888785]
3. Higgins PD, Schwartz M, Mapili J, Zimmermann EM. Is endoscopy necessary for the measurement of disease activity in ulcerative colitis? *American Journal of Gastroenterology*. 2005; 100:355–61. [PubMed: 15667493]
4. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine*. 2005; 353:2462–76. [PubMed: 16339095]
5. Lewis JD, Lichtenstein GR, Stein RB, et al. An open-label trial of the PPAR-gamma ligand rosiglitazone for active ulcerative colitis. *American Journal of Gastroenterology*. 2001; 96:3323–8. [PubMed: 11774944]
6. Su C, Lewis JD, Goldberg B, Brensinger C, Lichtenstein GR. A meta-analysis of the placebo rates of remission and response in clinical trials of active ulcerative colitis. *Gastroenterology*. 2007; 132:516–26. [PubMed: 17258720]
7. Lewis J, Lichtenstein G, Deren J, et al. Rosiglitazone for ulcerative colitis: a placebo-controlled trial. *Gastroenterology*. 2008; 134:688–95. [PubMed: 18325386]
8. Ilnyckyj A, Shanahan F, Anton PA, Cheang M, Bernstein CN. Quantification of the placebo response in ulcerative colitis. *Gastroenterology*. 1997; 112:1854–8. [PubMed: 9178676]
9. Su C, Lichtenstein GR, Krok K, Brensinger CM, Lewis JD. A meta-analysis of the placebo rates of remission and response in clinical trials of active Crohn's disease. *Gastroenterology*. 2004; 126:1257–69. [PubMed: 15131785]
10. 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. [PubMed: 8299896]
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 8435711]
12. Menees S, Higgins P, Korsnes S, Elta G. Does colonoscopy cause increased ulcerative colitis symptoms? *Inflammatory Bowel Diseases*. 2007; 13:12–8. [PubMed: 17206634]
13. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut*. 1998; 43:29–32. [PubMed: 9771402]

14. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine*. 2001; 344:1594–602. [PubMed: 11372012]
15. Froslic KF, Jahnsen J, Moum BA, Vatn MH, Group I. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. *Gastroenterology*. 2007; 133:412–22. [PubMed: 17681162]

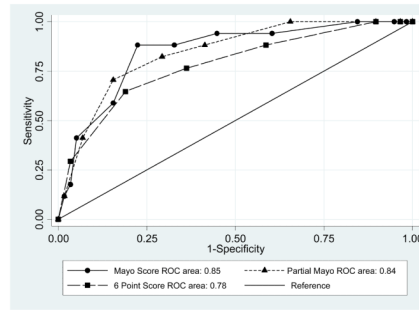


Figure 1. Comparison of ROC curves for the Mayo score, the partial Mayo score, and the 6 Point Score to identify patient-defined remission. The area under the curve was less for the 6 Point Score than for the Mayo and partial Mayo scores ($p=0.03$).

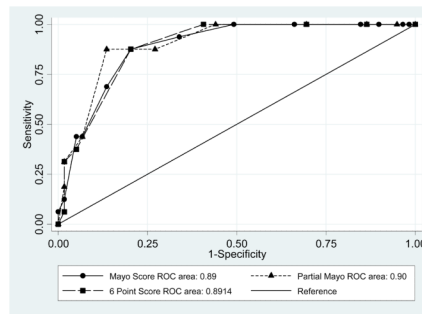


Figure 2. Comparison of ROC curves for the Mayo score, the partial Mayo score, and the 6 Point Score to identify patient-defined clinical improvement. There was no significant difference in the area under the three curves ($p=0.48$).

Table 1**Components of the Mayo Score****Stool Frequency**

- 0 = Normal
- 1 = 1–2 stools/day more than normal
- 2 = 3–4 stools/day more than normal
- 3 = >4 stools/day more than normal

Rectal bleeding*

- 0 = None
- 1 = Visible blood with stool less than half the time
- 2 = Visible blood with stool half of the time or more
- 3 = Passing blood alone

Mucosal appearance at endoscopy[†]

- 0 = Normal or inactive disease
- 1 = Mild disease (erythema, decreased vascular pattern, mild friability)
- 2 = Moderate disease (marked erythema, absent vascular pattern, friability, erosions)
- 3 = Severe disease (spontaneous bleeding, ulceration)

Physician rating of disease activity

- 0 = Normal
- 1 = Mild
- 2 = Moderate
- 3 = Severe

* A score of 3 for bleeding required patients to have at least 50% of bowel motions accompanied by visible blood and at least one bowel motion with blood alone.

[†]The mucosal appearance at endoscopy is not included in the Partial Mayo Score

Table 2

Timing and structure of outcome measures

Disease activity measure	Structure	Week 0	Week 4	Week 8	Week 12
Patient assessment of clinical activity	Perfect, minimal, mild, moderate, moderately severe, severe	X	X	X	X
Patient assessment of change from week 0	Much better, moderately better, a little better, no change, a little worse, moderately worse, much worse		X	X	X
Mayo score	Stool frequency, bleeding, physician rating, mucosal appearance	X			X
Partial Mayo score	Stool frequency, bleeding, physician rating	X	X	X	X
6 point Mayo score	Stool frequency, bleeding	X	X	X	X
IBDQ	Full version	X	X	X	X

Table 3

Correlation of disease activity measures with patient assessment of disease activity* (n=75)

Disease activity measure	Spearman correlation coefficient	P value
Mayo score	0.71	<0.0001
Partial Mayo score	0.70	<0.0001
6 point Mayo score	0.61	<0.0001
IBDQ	-0.63	<0.0001

* Patient assessment of disease activity based on a single 6 point scale ranging from perfect to severe disease.

Table 4

Sensitivity and specificity of disease activity measures to identify patient assessment of disease remission*
(n=75)

Disease activity measure	Cut point [#]	Sensitivity	Specificity	Area under the curve
Mayo score	4.5	88%	78%	0.85
Partial Mayo score	2.5	71%	84%	0.84
6 point Mayo score	1.5	65%	81%	0.78

* Remission defined as patient assessment of disease activity as perfect or very good (minimal symptoms).

[#] Cut point – maximal product of sensitivity and specificity

Table 5

Sensitivity and specificity of disease activity measures to identify patient assessment of clinical response*
(n=75)

Disease activity measure	Cut point [#]	Sensitivity	Specificity	Area under the curve
Mayo score	3.5	88%	80%	0.89
Partial Mayo score	3.5	88%	87%	0.90
6 point Mayo score	2.5	88%	80%	0.89

* Response defined as patient assessment of disease activity improving at least 2 points on a 6 point scale ranging from perfect to severe disease.

[#] Cut point – maximal product of sensitivity and specificity