Journal of Crohn's and Colitis (2014) **8**, 1444–1453

Available online at www.sciencedirect.com

ScienceDirect

A randomized clinical trial of mesalazine suppository: The usefulness and problems of central review of evaluations of colonic mucosal findings☆

Kiyonori Kobayashi [a,*], Fumihito Hirai [b], Makoto Naganuma [c], Kenji Watanabe [d], Takafumi Ando [e], Hiroshi Nakase [f], Katsuyoshi Matsuoka [g], Mamoru Watanabe [h]

[a] *Department of Research and Development Center for New Medical Frontiers, School of Medicine, Kitasato University, Kanagawa, Japan*
[b] *Department of Gastroenterology, Fukuoka University Chikushi Hospital, Fukuoka, Japan*
[c] *Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan*
[d] *Department of Gastroenterology, Osaka City University Graduate School of Medicine, Osaka, Japan*
[e] *Department of Gastroenterology, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[f] *Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
[g] *Division of Gastroenterology and Hepatology, Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan*
[h] *Department of Gastroenterology and Hepatology, Graduate School of Tokyo Medical and Dental University, Tokyo, Japan*

Received 3 April 2014; received in revised form 22 May 2014; accepted 22 May 2014

KEYWORDS
Ulcerative colitis;
Colonoscopy;
Mucosal findings;
Central review;
Inter-rater reliability

Abstract

Background: The methods of evaluating endoscopic mucosal findings and the definition of mucosal healing in inflammatory bowel disease have not been standardized.
Aim: To examine a third-party central review of colonic mucosal evaluations.
Methods: A double-blind, placebo-controlled, parallel-group trial was performed for 4 weeks, which involved continuous administration of a 1-g mesalazine suppository to 129 patients with mild to moderate ulcerative colitis and active rectal inflammatory findings. Mucosal findings were evaluated by using a 4-grade score (0, 1, 2, 3). Reviews by attending physicians were considered the primary evaluations. Concurrently, a central review committee of 7 gastroenterologists served as the third party.

Abbreviations: ICC, intra-class correlation coefficients.

☆ The clinical trial registration: JapicCTI-111421.

[*] Corresponding author at: Department of Research and Development Center for New Medical Frontiers, School of Medicine, Kitasato University, 2-1-1 Asamizodai, Minami, Sagamihara, Kanagawa 252-0380, Japan. Tel.: +81 42 748 9111; fax: +81 42 749 8690.
E-mail address: koba-eus@kitasato-u.ac.jp (K. Kobayashi).

http://dx.doi.org/10.1016/j.crohns.2014.05.005

Results: The endoscopic remission induction rate from the attending physicians' evaluations was 82.8% in the mesalazine suppository group and 31.1% in the placebo suppository group, whereas the respective rates from the central review committee were 90.6% and 59.0%. However, there was a difference of 27.9 percentage points between the remission induction rates of the placebo group found by the two groups of raters. Differences in the evaluations of mucosal finding scores were also found among the third-party reviewers.

Conclusions: The evaluations of the attending physicians were consistent with those of the central review committee in showing the effectiveness of mesalazine suppository through the index of mucosal healing. However, differences were observed among the raters in their evaluations of mucosal finding scores. Therefore, standardizing evaluation criteria and improving review methods for mucosal findings would enable the more effective use of third-party central reviews in clinical drug trials.

1. Introduction

Ulcerative colitis is an inflammatory bowel disease with primary symptoms that include frequent diarrhea, hemafecia, and abdominal pain. The disease involves repeated stages of active subjective symptoms and stages of remission of these symptoms. Patients are rarely cured completely, and the disease tends to be chronic.[1–3] The cause of ulcerative colitis remains unclear; however, it involves erosion and ulceration of the colonic mucosa.[4] Therefore, a definitive diagnosis requires not only the presence of clinical symptoms such as persistent or recurrent diarrhea or stool with mucous and blood, but also an evaluation of mucosal findings through colonoscopy or confirmation with histopathological findings. Recent advancements in colonoscopy equipment have enabled a more precise evaluation of mucosal findings in ulcerative colitis.

Traditionally, the aim of ulcerative colitis treatment is to ameliorate clinical symptoms such as frequent bowel movements and hemafecia. However, mucosal healing is becoming a therapeutic target with the use of long-term, high-dose mesalazine, anti-tumor necrosis-α antibody drugs, and immunomodulators such as azathioprine.[5–7] Moreover, there have been reports on methods for evaluating ulcerative colitis activity, such as qualitatively categorizing clinical symptoms and physical and mucosal findings, as well as quantitatively scoring activity indices.[8–12] However, as methods of evaluating mucosal findings or defining mucosal healing have yet to be standardized, evaluations are left to the discretion of individual physicians. Therefore, naturally, large physician-dependent differences in the evaluations of mucosal findings have been reported.[13–15]

In everyday medical care, treatment based on the attending physicians' evaluation of mucosal findings is not considered problematic. However, there are concerns that in clinical trials, differences between the assessments of individual physicians could affect the evaluation of drug effectiveness. Therefore, our objective was to confirm the reliability of the attending physicians' evaluations for the performance of uniform evaluations of mucosal findings in clinical trials. To achieve this goal, we recruited third parties not involved in the clinical trial (a central review committee) to also perform evaluations.[11,16–20] However, considering the evaluations by a central review committee as the results of a clinical trial would require many stipulations over the mucosal images presented to the committee, such as concerning the capabilities of the imaging device and the photographic methods used. The central review committee would also have to perform its evaluations quickly. The more members the committee has, the more difficult it would be to perform speedy evaluations. Therefore, when a central review committee is formed to perform evaluations, it is important to find an evaluation method that can be executed both quickly and precisely under a limited number of conditions.

In this double-blind, parallel-group mesalazine suppository trial[21] of patients with mild to moderate ulcerative colitis and active inflammatory findings in the rectal area, the endoscopic remission induction rate from the mucosal finding scores given by attending physicians were considered as the primary evaluations. To confirm the reliability of those results, a central review committee was formed consisting of 7 gastroenterologists who did not participate in the trial. For each case, the committee evaluated mucosal findings only from the end of the trial (or at drop-out). These evaluations were used to examine the reliability of the results of the attending physicians' evaluations, as well as to check for differences between the evaluations of the attending physicians and those of the central review committee, and among the 7 members of the committee. This was expected to clarify the issues related to uniformity in evaluating mucosal findings and help with proposing countermeasures.

2. Materials and methods

2.1. Outline of the clinical trial

This phase III, randomized, placebo-controlled, double-blind, multi-institutional, parallel-group trial[21] was performed across 45 institutions in Japan after enrolling 129 patients with mild to moderate active ulcerative colitis and active inflammatory rectal findings. The subjects were men and women aged ≥15 years and ≤74 years who had ulcerative colitis and met the following criteria: (i) a score of 4–8 on the ulcerative colitis disease activity index[22,23] and a score of ≥2 considering the mucosal findings in the rectum; and (ii) initial episode-type patient or flare-up and remission-type patient. Patients who met any of the following criteria were excluded: (i) having a score of ≥2 considering the colonic mucosal findings in areas other than the rectum at the start of the trial; (ii) receiving any of the following treatments within 4 weeks after initiating the investigational drug (oral mesalazine exceeding 2400 mg/day; oral salazosulfapyridine exceeding 4500 mg/day; mesalazine enema; mesalazine or salazosulfapyridine suppository; corticosteroid drugs administered orally, as an enema, suppository, injection, or anally such as in the form of an ointment); cytapheresis; (iii) receiving any of the following treatments within 12 weeks after initiating the investigational drug (oral or injectable immunosuppressants or immunomodulators; other investigational drugs); (iv) having a history of hypersensitivity to mesalazine or salicylic acid drugs; and (v) were pregnant or nursing.

The patients were randomly assigned to receive a 1-g mesalazine suppository (Pentasa; Ferring Pharmaceuticals, Saint-Prex, Switzerland) or placebo at the start of this study, according to a computer-generated randomization scheme. Subjects in both groups had 1 suppository placed inside the rectum every day continuously for 4 weeks. The investigational drugs (mesalazine suppository, placebo suppository) were not distinguishable from each other and were allotted by using a randomized schema created with a computer when starting their administration.

The main evaluation of effectiveness was the endoscopic remission induction rate derived from mucosal finding scores in the rectum evaluated after 4 weeks of administration (or at drop-out). To achieve uniformity in the reviewers' evaluations of mucosal findings observed by colonoscopy, the evaluations of mucosal findings were made by using a 4-grade score (0–3) based on a mucosal finding atlas (not shown) created from the Mayo endoscopic subscore[10] under the supervision of a gastroenterologist. Endoscopic remission was defined as a mucosal finding score of 0 or 1. In this system, a score of 1 is given when "redness, a reduced vascular pattern, and/or mild fragility" are observed in endoscopic findings. In particular, the presence of redness or mild fragility is sometimes considered to indicate mild inflammatory states; therefore, some researchers oppose correlating a score of 1 with mucosal healing. However, in patients experiencing repeated flare-ups and remissions of ulcerative colitis, a reduced mucosal vascular pattern is sometimes evident even when other signs of inflammation have disappeared. In addition, it can be difficult to determine whether the fragility is due to artifacts owing to the insertion of an endoscope. Therefore, in this clinical trial, endoscopic remission, or mucosal healing, was defined as a Mayo endoscopic subscore of 0 or 1, which were based on the results of trials performed in other countries.[24,25]

This clinical trial observed the ethical principles outlined in the Declaration of Helsinki, criteria for clinical trials of medical products (Good Clinical Practice), and other related rules and regulations. Prior approval was acquired from the institutional review board of each participating institution. In addition, written informed consent was obtained from all the patients after they received a full explanation of the trial. For patients who were minors, written consent was obtained from both the patient and his/her legal representative. This trial was registered on a clinical trial website. Analyses for this study were performed as part of the trial.

2.2. Evaluation of mucosal findings

Before initiating administration of the investigational drug, the attending physicians observed the entire colon by means of colonoscopy, determined the area of the rectum where inflammation was most severe, and finally evaluated the mucosal findings. After 4 weeks of administration of the investigational drug (or at drop-out), the mucosal findings at the predetermined area of rectal inflammation were re-evaluated, and colonoscopic images were taken after removing as much stool, mucous, and other residue as possible. The mucosal findings were evaluated with a 4-grade score (0, 1, 2, 3) by using a mucosal finding atlas created before the trial (0 = normal or non-active findings; 1 = redness, reduced vascular pattern, mild fragility; 2 = marked redness, no visible vascular pattern, fragility, erosion; 3 = spontaneous bleeding, ulceration). The mucosal finding scores were not intended to reflect bleeding or redness caused by bowel preparation for colonoscopy. In addition, a score of 2 could not be given on the basis of a lack of a visible vascular pattern alone; marked redness, fragility, or erosion also had to be present.

The central review committee used the endoscopic images submitted by the attending physicians (digital or printed images; at least 4 images including an overall image and images adjacent to the specified area) taken after 4 weeks of drug administration (or at drop-out) to evaluate the mucosal findings with a 5-grade score (0, 1, 2, 3, review impossible) by using the mucosal finding atlas. These endoscopic images were devoid of any information that could identify the patients. The central review committee was not involved in removing such information from the images.

The central review committee consisted of 7 gastroenterologists in the field of inflammatory bowel disease who were not participating in the trial ("central review members"). Each member reviewed the mucosal findings on all the colonoscopic images submitted by the attending physicians. The results of their evaluations were then summed up, and when at least 5 members agreed on a score for a case ("agreed group"), this score was adopted as the committee's review. Cases in which (i) the evaluations of at least 5 members did not agree, (ii) 2 or more evaluations were "review impossible," and (iii) the evaluation was postponed for a separate meeting ("review meeting group"), were discussed in a review meeting attended by at least 5 central review members. The central review committee's final review was decided after the colonoscopic images of these cases were reexamined by all the participants. The attending physicians and central review members had roughly the same experience as that of gastroenterologists; the former had a mean of 23.94 years (13–40 years) of experience and the latter had 19.57 years (13–26 years) of experience.

2.3. Statistical evaluation

The subjects of the evaluation underwent a colonoscopy after 4 weeks of receiving the investigational drug (or at drop-out). The subjects of the clinical trial were 129 patients with mild to moderate active ulcerative colitis and active inflammatory findings in the rectal area (mesalazine suppository group, 65 cases; placebo suppository group, 64 cases). Four patients dropped out because of exacerbation of ulcerative colitis, pregnancy, an adverse event, and poor adherence in consuming the investigational drug. These patients did not undergo a colonoscopy upon their exit from the trial, and therefore their mucosal images were not submitted to the central review committee. Therefore, a total of 125 cases were used for inter-rater comparison of mucosal findings.

Table 1 Endoscopic remission induction rates from the attending physicians and the central review committee for each drug group.

Item	Mesalazine group (n = 64)		Placebo group (n = 61)	
	Attending physicians	Central review committee	Attending physicians	Central review committee
Endoscopic remission [a]	53	58	19	36
Endoscopic remission induction rate (%) [b]	82.8	90.6	31.1	59.0
Difference in endoscopic remission induction rates (%) [c]	−7.8		−27.9	
P value [d]	0.1250		0.0005	

[a] Patients with a 0 or 1 mucosal finding score after 4 weeks of administration (or at drop-out).

[b] Endoscopic remission induction rate (%) = (no. of remission-induced patients/total patients) × 100.

[c] (Attending physicians' evaluation) minus (central review committee's evaluation).

[d] McNemar test (5% significance level).

Endoscopic remission induction rates were calculated by using the evaluations provided by the attending physicians' and the central review committee of the mesalazine suppository group and the placebo suppository group. Inter-rater differences for each drug group were examined by using the McNemar test (5% significance level). Weighted kappa coefficients were calculated to measure inter-rater reliability for the mucosal finding scores of the 2 groups of raters. Furthermore, intra-class correlation coefficients (ICC) were calculated to consider variations between the 7 central review members. Next, endoscopic remission induction rates were calculated for each mucosal finding score among all the cases, the agreed group, and the review meeting group. Inter-rater differences were compared for each drug group by using the Wilcoxon signed rank test and McNemar test (5% significance level). Weighted kappa coefficients were calculated as a measure of inter-rater reliability for the mucosal finding scores given by the attending physicians and the central review committee, and ICCs were calculated to consider variations between the 7 central review members. Differences between raters for each mucosal finding score in each drug group were compared by using the McNemar test (5% significance level). Similarly, differences between raters for each mucosal finding score in each drug group were compared between the agreed group and the review meeting group by using the McNemar test (5% significance level).

Weighted kappa coefficients and ICC are often used as criteria for examining inter-rater reliability while using evaluation scales, such as the mucosal finding scores used here. Weighted kappa coefficients and ICCs ≤0.40 are considered "poor," 0.41–0.74 are "fair to good," and 0.75–1.00 are "excellent."[26–28] A score of 1 represents complete agreement among reviewers, with an increase in the gap between reviewers as the value becomes smaller.

3. Results

The endoscopic remission induction rate from the attending physicians' evaluations was 82.8% in the mesalazine suppository group (53 of 64 cases) and 31.1% in the placebo suppository group (19 of 61 cases); the rate from the central review committee's evaluations was 90.6% in the mesalazine suppository group (58 of 64 cases) and 59.0% in the placebo suppository group (36 of 61 cases). Both groups of raters found that the therapeutic effect of mesalazine suppository was significantly superior to that of a placebo. In the mesalazine suppository group, there was a difference of 7.8 percentage points between the endoscopic remission induction rate of the attending physicians and that of the central review committee, which was not statistically significant. However, the 27.9 percentage point difference observed for the placebo suppository group was significant (McNemar test, $P < 0.01$) (Table 1). To determine the reliability of the attending physicians' evaluations, their results were compared with those of the central review committee and examined for consistency. The weighted kappa coefficients for all cases, the mesalazine suppository group, and the placebo suppository group (each

Table 2 Consistency of the central review committee's evaluations with regard to the attending physicians' evaluations.

Mucosal finding scores		Central review committee overall					Central review committee's evaluations of each drug group									
							Mesalazine group					Placebo group				
		0	1	2	3	Total	0	1	2	3	Total	0	1	2	3	Total
Attending physicians	0	14	7	0	0	21	12	7	0	0	19	2	0	0	0	2
	1	10	37	4	0	51	9	24	1	0	34	1	13	3	0	17
	2	0	25	24	0	49	0	5	5	0	10	0	20	19	0	39
	3	0	1	3	0	4	0	1	0	0	1	0	0	3	0	3
Total		24	70	31	0	125	21	37	6	0	64	3	33	25	0	61
Weighted kappa coefficient		0.613					0.542					0.460				

Table 3 Consistency of the 7 central review members' evaluations with regard to the attending physicians' evaluations.

Applicable no. of patients	Overall (n = 125)	Mesalazine group (n = 64)	Placebo group (n = 61)
True estimate of variance	0.331	0.279	0.200
Overall estimate of variance	0.548	0.515	0.397
ICC [a]	0.604	0.541	0.504

[a] ICC: intra-class correlation coefficient = (true estimate of variance) / (overall estimate of variance).

drug group) were 0.613, 0.542, and 0.460, respectively (Table 2). Comparing the consistency between the 7 central review members' evaluations and the attending physicians' evaluations revealed ICCs of 0.604, 0.541, and 0.504 for all cases, and each drug group, respectively (Table 3).

Endoscopic remission induction rates were also calculated from the attending physicians' and the central review committee's evaluations for each drug group in the agreed group (75 cases) and the review meeting group (50 cases). The differences between the evaluations of the 2 groups of raters in the mesalazine suppository group were 2.6 percentage points and 15.4 percentage points, respectively, which were not significant. However, the differences between the raters in the placebo suppository group were 24.3 percentage points and 33.3 percentage points, respectively, which were significant (McNemar test, $P < 0.05$) (Table 4).

Next, to determine the reliability of the attending physicians' evaluations, the results of the attending physicians' evaluations were compared with those of the central review committee and examined for consistency. In the agreed group, the weighted kappa coefficients for all cases and each drug group were 0.6887, 0.6945, and 0.4658, respectively, whereas those for the review meeting group were 0.4946, 0.3090, and 0.4375, respectively (Table 5).

Comparing the consistency of the 7 central review members' evaluations with the attending physicians' evaluations in the agreed group revealed ICCs of 0.728, 0.674, and 0.638 for all cases, and each drug group, respectively, whereas those in the review meeting group were 0.406, 0.365, and 0.246, respectively (Table 6).

The evaluations of the attending physicians and the central review committee were also compared for each mucosal finding score in each drug group. No differences were observed between the groups in the mesalazine suppository group; significant differences between the raters (McNemar test, $P < 0.01$) were observed for mucosal finding scores of 1 and 2 in the placebo suppository group (Table 7). The attending physicians' and central review committee's evaluations were also compared for each mucosal finding score in each drug group in the agreed group (75 cases) and the review meeting group (50 cases). No differences were observed for any mucosal finding score between the raters in the mesalazine suppository group. However, a significant difference (McNemar test, $P < 0.05$) was observed for a mucosal finding score of 1 in the placebo suppository group between the raters in the review meeting group (Table 8).

4. Discussion

There are problems with uniformity while evaluating mucosal findings from patients with inflammatory bowel disease when

Table 4 Endoscopic remission induction rates from the attending physicians and the central review committee for each drug group in the agreed group and the review meeting group.

Item	Agreed group (n = 75)				Review meeting group (n = 50)			
	Mesalazine group (n = 38)		Placebo group (n = 37)		Mesalazine group (n = 26)		Placebo group (n = 24)	
	Attending physicians	Central review committee	Attending physicians	Central review committee	Attending physicians	Central review committee	Attending physicians	Central review committee
Endoscopic remission [a]	33	34	14	23	20	24	5	13
Endoscopic remission induction rate (%) [b]	86.8	89.5	37.8	62.2	76.9	92.3	20.8	54.2
Difference in endoscopic remission induction rates (%) [c]	−2.6		−24.3		−15.4		−33.3	
P value [d]	1.0000		0.0225		0.2188		0.0215	

[a] Patients with a 0 or 1 mucosal finding score after 4 weeks of administration (or at drop-out).
[b] Endoscopic remission induction rate (%) = (no. of remission-induced patients/total patients) × 100.
[c] (Attending physicians' evaluation) minus (central review committee's evaluation).
[d] McNemar test (5% significance level).

Table 5 Consistency of the central review committee's evaluations with regard to the attending physicians' evaluations in the agreed group and the review meeting group.

Mucosal finding score [a]		Agreed group (n = 75)														
		Overall					Mesalazine group					Placebo group				
		0	1	2	3	Total	0	1	2	3	Total	0	1	2	3	Total
Attending physicians	0	11	5	0	0	16	10	5	0	0	15	1	0	0	0	1
	1	5	24	2	0	31	4	14	0	0	18	1	10	2	0	13
	2	0	12	15	0	27	0	1	4	0	5	0	11	11	0	22
	3	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1
Total		16	41	18	0	75	14	20	4	0	38	2	21	14	0	37
Coefficient [b]		0.6887					0.6945					0.4658				

Mucosal finding score [a]		Review meeting group (n = 50)														
		Overall					Mesalazine group					Placebo group				
		0	1	2	3	Total	0	1	2	3	Total	0	1	2	3	Total
Attending physicians	0	3	2	0	0	5	2	2	0	0	4	1	0	0	0	1
	1	5	13	2	0	20	5	10	1	0	16	0	3	1	0	4
	2	0	13	9	0	22	0	4	1	0	5	0	9	8	0	17
	3	0	1	2	0	3	0	1	0	0	1	0	0	2	0	2
Total		8	29	13	0	50	7	17	2	0	26	1	12	11	0	24
Coefficient [b]		0.4946					0.3090					0.4375				

[a] Mucosal finding score after 4 weeks of administration (or at drop-out).
[b] Weighted kappa coefficient.

mucosal healing is used as a therapeutic target. This is an important issue in regular medical care but is particularly significant when changes in mucosal findings are used as an index of the effectiveness of test drugs in clinical trials, as the results need to be both accurate and reproducible.

Travis et al.[29] used endoscopic images to examine evaluations of ulcerative colitis activity, as well as intra- and inter-observer variations in the evaluations. They found that disease activity classifications made by using endoscopy could help in the accurate evaluation of endoscopic severity, but that variation and reactivity should be thoroughly investigated if it was to be used as an evaluation index in clinical trials.

In clinical trials on inflammatory bowel disease, forming a central review committee consisting of individuals not participating in the trial is one method to reduce variations in the endoscopic evaluations of mucosal findings and ensure uniformity of the evaluations. Evaluations by a central review committee are considered advantageous because they are performed by specialists in the disease who aim to achieve uniform evaluations, which leads to highly reliable results. However, when performing multi-institutional or international joint clinical trials, a number of issues need to be resolved, such as concerning the speed of the evaluations, and stipulations over the quality and quantity of mucosal images that need to be submitted to the review committee from each participating institution.

Therefore, in this clinical trial of mesalazine suppository, the attending physicians' evaluations were adopted as the primary evaluation of effectiveness, while a central review committee was also formed with 7 specialists in inflammatory bowel disease who were not participating in the trial. As the simplest method for the central review that took both uniformity and speed into consideration, the attending physicians' evaluations of mucosal findings were compared with those of the central review committee only for colonoscopic images taken at the end of the trial (or at drop-out) to examine inter-rater differences. Both the attending physicians' and the central review committee's evaluations confirmed that the therapeutic effect of mesalazine suppository was significantly higher than that of a placebo. However, although the central review committee confirmed the reliability of the attending physicians' evaluations when the endoscopic remission induction rates of each drug group were compared, it was found that the raters' results were similar in the mesalazine suppository group, but the central review committee's endoscopic remission induction rate for the placebo suppository group was significantly higher than the rate given by the attending physicians (Table 1).

Interestingly, there were significant differences between the attending physicians' and the central review committee's reviews of the placebo suppository group among the 75 cases on which at least 5 central review members agreed (agreed group) and the 50 cases on which the committee's evaluations were divided, and a final evaluation was decided in a review meeting (review meeting group) (Table 4). This shows that there was a clear difference between the attending physicians' and the central review committee's evaluations of the placebo suppository group. The variations between the 7 central review members were particularly high in the review meeting group. As this influenced the final evaluations by the committee, ensuring reliability with regard to the attending physicians' evaluations became difficult (Table 6). This shows that there are mucosal images that can produce divided evaluations even

Table 6 Consistency of the 7 central review members' evaluations with regard to the attending physicians' evaluations in the agreed group and the review meeting group.

Applicable no. of patients	Agreed group (n = 75)			Review meeting group (n = 50)		
	Overall (n = 75)	Mesalazine (n = 38)	Placebo (n = 37)	Overall (n = 50)	Mesalazine (n = 26)	Placebo (n = 24)
True estimate of variance	0.370	0.280	0.249	0.234	0.223	0.095
Overall estimate of variance	0.508	0.416	0.391	0.576	0.611	0.386
ICC [a]	0.728	0.674	0.638	0.406	0.365	0.246

[a] ICC: intra-class correlation coefficient = (true estimate of variance) / (overall estimate of variance).

between members of a central review committee formed by specialists in inflammatory bowel disease. The differences in the attending physicians' and central review committee's evaluations of the placebo suppository group were found to lie in the scores of 1 and 2 (Table 7). It is worth noting that the mucosal finding scores of 1 in the review meeting group were decided as a committee in a meeting attended by the members of the central review committee (6 members in attendance) after they discussed the digital and/or printed images submitted by the attending physicians and while referencing the mucosal finding atlas. Focusing on the mucosal finding score of 1, while the attending physicians' and central review committee's evaluations of the mesalazine suppository group were mostly identical, the central review committee more often evaluated the placebo suppository group with a score of 1 relative to the attending physicians (Table 8). As the endoscopic remission induction rates were determined from mucosal finding scores of 0 or 1, endoscopic remission induction rates of the committee were higher than those of the attending physicians.

In the evaluations of mucosal finding scores, the question arises why the raters' endoscopic remission induction rates were nearly identical in the mesalazine suppository group, while there was a clear difference between the rates for the placebo suppository group. Mucosal findings greatly improved in the patients who received mesalazine suppositories, and, although this was a blind trial, it appears that both the attending physicians and the central review committee were able to definitively score many of these cases as 1 or lower. This is supported by the fact that almost the same number of cases was evaluated as 0 by the raters. Moreover, even if the central review committee tended to give low scores, it was assumed that as the attending physicians gave mucosal finding scores of 0 or 1 in many cases, the endoscopic remission induction rates from both groups of raters were nearly identical. In the placebo suppository group, however, the central review committee did give lower scores than the attending physicians (Table 7).

Lange et al.[30] studied differences in the evaluations of endoscopic images of ulcerative colitis between experienced and inexperienced endoscopic physicians. They found that experienced physicians gave significantly higher scores; among the factors of vascular pattern, erosion, ulceration, and fragility, differences were particularly significant while considering ulceration. The 7 central review members in this trial were well versed in the diagnosis and treatment of inflammatory bowel disease and were selected because they were specialists in the lower gastrointestinal tract who regularly performed endoscopic examinations and evaluated mucosal findings in their own practice. Therefore, it is difficult to believe that they would assign relatively low scores.

One reason why evaluations of mucosal findings differed between the attending physicians and the central review committee is thought to be in the method used to evaluate findings. The attending physicians were able to evaluate findings from the end of the trial (or at drop-out) by referencing findings with active inflammation from when the subject registered; however, the central review committee only evaluated findings from the end of the trial (or at drop-out). In this study, a central review was conducted only at the end of the trial (or at drop-out). However, if evaluations had also been performed at registration, it is highly possible that mucosal findings given a score of 2 by the attending physicians would have been scored as 1 by the central review members.

Table 7 Comparison of each mucosal finding score from the attending physicians and the central review committee for each drug group.

Mucosal finding score [a]	Mesalazine group (n = 64)			Placebo group (n = 61)		
	Attending physicians	Central review committee	Test [b] (*P* value)	Attending physicians	Central review committee	Test [b] (*P* value)
0	19	21	0.8036	2	3	1.0000
1	34	37	0.6776	17	33	0.0015
2	10	6	0.2188	39	25	0.0094
3	1	0	–	3	0	–
Total	64	64		61	61	

[a] Mucosal finding score after 4 weeks of administration (or at drop-out).

[b] McNemar tests (5% significance level.)

Table 8 Comparison of each mucosal finding score from the attending physicians and the central review committee in the agreed group and the review meeting group for each drug group.

Mucosal finding score[a]	Agreed group (n = 75)						Review meeting group (n = 50)					
	Mesalazine group			Placebo group			Mesalazine group			Placebo group		
	Attending physicians	Central review committee	Test[b] (*P* value)	Attending physicians	Central review committee	Test[b] (*P* value)	Attending physicians	Central review committee	Test[b] (*P* value)	Attending physicians	Central review committee	Test[b] (*P* value)
0	15	14	1.0000	1	2	1.0000	4	7	0.4531	1	1	–
1	18	20	0.7539	13	21	0.0574	16	17	1.0000	4	12	0.0215
2	5	4	1.0000	22	14	0.0574	5	2	0.3750	17	11	0.1460
3	0	0	–	1	0	–	1	0	–	2	0	–
Total	38	38		37	37		26	26		24	24	

[a] Mucosal finding score after 4 weeks of administration (or at drop-out).
[b] McNemar tests (5% significance level).

Therefore, at the end of the trial, the mucosal finding scores of these patients in the placebo suppository group would not change, keeping them at a score of 1.

Another possible reason for the discrepancy between the attending physicians' and the central review committee's evaluations is a difference in the degree of their grasp of the mucosal findings. One of the conditions of this trial was a score of at least 2 in rectal endoscopic findings at registration. In addition to no visible vascular pattern, there should have been marked redness, fragility, or erosion to give a score of 2. If attending physicians encountered intestinal residue, mucous, or secretions adhering to the area while performing colonoscopies, they were allowed to clean the area with water or use other methods to remove the material as much as possible according to the protocol. To determine the area of the rectum with the most severe inflammation, they could use endoscopic retroflexion in the rectum, thereby gaining a dynamic and detailed view of the mucosal surface. This allowed them to precisely understand the mucosal findings (erosion, ulceration, color changes, etc.) in the area to be evaluated. In contrast, the central review committee evaluated the endoscopic mucosal findings by using digital or printed images submitted by the attending physicians. Moreover, when printed images were submitted, the images were scanned to create digital images and then printed out. Among these digital images and printed photographs, there were some that included residue, mucous, or secretions adhering to the inner walls of the intestines. However, final evaluations for these images were made in the central review, and no cases were determined to be unsuitable for evaluation due to poor pretreatment of the intestines, but there might have been cases that could not recognize erosions dottled exactly. While the digital images were close to what was observed by the attending physicians, the printed images were sometimes different from what was actually observed during colonoscopy, such as the color of the mucosa, and it could be difficult to identify not only areas of erosion but also small ulcerations. Moreover, the central review committee evaluated a limited number of still images (approximately 4), and in some images there was mucous or stool adhering to the mucosa, which could have prevented the reviewers from accurately identifying scatterings of small erosions and other features. The presence of erosion would acquire a mucosal finding score of 2; therefore, an inability to identify erosion, or differences among reviewers, could have influenced the scoring of mucosal findings. The basis for this is that although the central review members examined the digital and printed images together at a central review meeting, differences of opinion arose among the committee members regarding their evaluations of erosion.

This study reconfirmed that there are inter-rater differences in endoscopic evaluations of mucosal findings in ulcerative colitis. Accurate evaluation of small, active lesions such as erosion is particularly important to improve the uniformity of endoscopic evaluations. Erosion is a form of minor mucosal damage that accompanies inflammation. As erosion progresses to ulceration, it is an extremely important finding in evaluating drug effectiveness in clinical trials. To allow for accurate endoscopic evaluations of these small lesions in central reviews, it is necessary to create uniform inter-rater evaluation criteria, as well as improve submission methods, such as increasing the quality of the digital images presented. In the future, the knowledge gained in this study could be used to improve the review methods and further increase the reliability of evaluations by the central review, which could enable the evaluation of mucosal findings in inflammatory bowel disease to be completely performed by third parties.

Conflict of interests

This trial was funded entirely by Kyorin Pharmaceutical Co., Ltd. Employees of Kyorin Pharmaceutical Co., Ltd., were involved in the analysis and interpretation of the data, and in the writing of the manuscript.

Kiyonori Kobayashi, Fumihiro Hirai, Makoto Naganuma, Kenji Watanabe, Takafumi Ando, Hiroshi Nakase, and Katsuyoshi Matsuoka have served on advisory boards and received consultancy funding from Kyorin Pharmaceutical Co. Ltd. Mamoru Watanabe have counseled the central review committee as the medical expert of this study and received a consultancy funding from Kyorin Pharmaceutical Co. Ltd.

Acknowledgments

The authors would like to thank all the patients and supervising physicians who participated in this trial. All authors have contributed to the creation of the atlas of mucosal findings and the assessment of the mucosal finding score in the implementation of the trial described in this article, and all authors participated in the preparation, review and approval of the manuscript content.

References

1. Moum B, Ekbom A, Vatn MH, Aadland E, Sauar J, Lygren I, et al. Clinical course during the 1st year after diagnosis in ulcerative colitis and Crohn's disease: results of a large, prospective population-based study in southeastern Norway, 1990–93. *Scand J Gastroenterol* 1997;**32**:1005–12.
2. Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002;**347**:417–29.
3. Lakatos PL, Lakatos L. Ulcerative proctitis: a review of pharmacotherapy and management. *Expert Opin Pharmacother* 2008;**9**:741–9.
4. Kishi H. Endoscopic characteristics of the healing process of ulcerative colitis. *Diagn Ther Endosc* 1998;**5**:37–48.
5. Lichtenstein GR, Rutgeerts P. Importance of mucosal healing in ulcerative colitis. *Inflamm Bowel Dis* 2010;**16**:338–46.
6. Rutgeerts P, Vermeire S, Van AG. Mucosal healing in inflammatory bowel disease: impossible ideal or therapeutic target? *Gut* 2007;**56**:453–5.
7. Frøslie KF, Jahnsen J, Moum BA, Vatn MH, IBSEN Group. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. *Gastroenterology* 2007;**133**:412–22.
8. Rachmilewitz D. Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. *BMJ* 1989;**298**:82–6.
9. 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.
10. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. *N Engl J Med* 1987;**317**:1625–9.
11. Baron JH, Connell AM, Lennard-Jones J. Variation between observers in describing mucosal appearances in proctocolitis. *BMJ* 1964;**1**:89–92.
12. Matts SG. The value of rectal biopsy in the diagnosis of ulcerative colitis. *Q J Med* 1961;**30**:393–407.
13. Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SP. Outcome measurement in clinical trials for ulcerative colitis: toward standardisation. *Trials* 2007;**8**:17.
14. Travis S, Cooney R, Lukas M, Butruk E, Kotzev I, Warren BF, et al. Conduct of clinical trials in UC: impact of independent scoring of endoscopic severity on results of a randomised controlled trial with a peptide and 5-ASA. *Am J Gastroenterol* 2006;**101**(Suppl 9):S429.
15. Orlandi F, Brunelli E, Feliciangeli G, Svegliati-Baroni G, Di Sario A, Benedetti A, et al. Observer agreement in endoscopic assessment of ulcerative colitis. *Ital J Gastroenterol Hepatol* 1998;**30**:539–41.
16. Ito H, Iida M, Matsumoto T, Suzuki Y, Sasaki H, Yoshida T, et al. Direct comparison of two different mesalamine formulations for the induction of remission in patients with ulcerative colitis: a double-blind, randomized study. *Inflamm Bowel Dis* 2010;**16**:1567–74.
17. Feagan BG, Sandborn WJ, D'Haens G, Pola S, McDonald JW, Rutgeerts P, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;**145**:149–57.
18. Delvaux M, Papanikolaou IS, Fassler I, Pohl H, Voderholzer W, Rösch T, et al. Esophageal capsule endoscopy in patients with suspected esophageal disease: double blinded comparison with esophagogastroduodenoscopy and assessment of interobserver variability. *Endoscopy* 2008;**40**:16–22.
19. Lee YC, Lin JT, Chiu HM, Liao WC, Chen CC, Tu CH, et al. Intraobserver and interobserver consistency for grading esophagitis with narrow-band imaging. *Gastrointest Endosc* 2007;**66**:230–6.
20. Edebo A, Tam W, Bruno M, Van Berkel AM, Jönson C, Schoeman M, et al. Magnification endoscopy for diagnosis of nonerosive reflux disease: a proposal of diagnostic criteria and critical analysis of observer variability. *Endoscopy* 2007;**39**:195–201.
21. Watanabe M, Nishino H, Sameshima Y, Ota A, Nakamura S, Hibi T. Randomised clinical trial: evaluation of the efficacy of mesalazine (mesalamine) suppositories in patients with ulcerative colitis and active rectal inflammation—a placebo-controlled study. *Aliment Pharmacol Ther* 2013;**38**:264–73.
22. Farup PG, Hinterleitner TA, Lukás M, Hébuterne X, Rachmilewitz D, Campieri M, et al. Mesalazine 4 g daily given as prolonged-release granules twice daily and four times daily is at least as effective as prolonged-release tablets four times daily in patients with ulcerative colitis. *Inflamm Bowel Dis* 2001;**7**:237–42.
23. Hanauer S, Schwartz J, Robinson M, Roufail W, Arora S, Cello J, et al. Mesalamine capsules for treatment of active ulcerative colitis: results of a controlled trial. Pentasa Study Group. *Am J Gastroenterol* 1993;**88**:1188–97.
24. Ngo Y, Gelinet JM, Ivanovic A, Kac J, Schénowitz G, Vilotte J, et al. Efficacy of a daily application of mesalazine (Pentasa) suppository with progressive release, in the treatment of ulcerative proctitis: a double-blind versus placebo randomized trial. *Gastroenterol Clin Biol* 1992;**16**:782–6.
25. Lucidarme D, Marteau P, Foucault M, Vautrin B, Filoche B. Efficacy and tolerance of mesalazine suppositories vs. hydrocortisone foam in proctitis. *Aliment Pharmacol Ther* 1997;**11**:335–40.
26. Feinstein AR. Clinimetrics. New Haven: Yale University Press; 1987.
27. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;**33**:613–9.
28. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.

29. Travis SP, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, et al. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 2013;**145**:987–95.

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