

Research article

Open Access

## Inter-observer agreement in the assessment of endoscopic findings in ulcerative colitis

Thomas de Lange<sup>\*1</sup>, Stig Larsen<sup>2</sup> and Lars Aabakken<sup>3</sup>

Address: <sup>1</sup>Department of Gastroenterology, Ullevaal University Hospital, Oslo, Norway, <sup>2</sup>Department of Epidemiology, Norwegian School of Veterinary Medicine, Oslo, Norway and <sup>3</sup>Department of Gastroenterology Rikshospitalet University Hospital, Oslo, Norway

Email: Thomas de Lange\* - t.d.lange@medisin.uio.no; Stig Larsen - stig.larsen@veths.no; Lars Aabakken - lars.aabakken@klinmed.uio.no

\* Corresponding author

Published: 18 May 2004

Received: 18 November 2003

*BMC Gastroenterology* 2004, **4**:9

Accepted: 18 May 2004

This article is available from: <http://www.biomedcentral.com/1471-230X/4/9>

© 2004 de Lange et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Endoscopic findings are essential in evaluating the disease activity in ulcerative colitis. The aim of this study was to evaluate how endoscopists assess individual endoscopic features of mucosal inflammation in ulcerative colitis, the inter-observer agreement, and the importance of the observers' experience.

**Methods:** Five video clips of ulcerative colitis were shown to a group of experienced and a group of inexperienced endoscopists. Both groups were asked to assess eight endoscopic features and the overall mucosal inflammation on a visual analogue scale. The following statistical analyses were used; Contingency tables analysis, kappa analysis, analysis of variance, Pearson linear correlation analysis, general linear models, and agreement analysis. All tests were carried out two-tailed, with a significance level of 5%.

**Results:** The inter-observer agreement ranged from very good to moderate in the experienced group and from very good to fair in the inexperienced group. There was a significantly better inter-observer agreement in the experienced group in the rating of 6 out of 9 features ( $p < 0.05$ ). The experienced and inexperienced endoscopists scored the "ulcerations" significantly different. ( $p = 0.05$ ). The inter-observer variation of the mean score of "erosions", "ulcerations" and endoscopic activity index in mild disease, and the scoring of "erythema" and "oedema" in moderate-severe disease was significantly higher in the inexperienced group.

A correlation was seen between all the observed endoscopic features in both groups of endoscopists. Among experienced endoscopists, a set of four endoscopic variables ("Vascular pattern", "Erosions", "Ulcerations" and Friability") explained 92% of the variation in EAI. By including "Granularity" in these set 91% of the variation in EAI was explained in the group of inexperienced endoscopists.

**Conclusion:** The inter-observer agreement in the rating of endoscopic features characterising ulcerative colitis is satisfactory in both groups of endoscopists but significantly higher in the experienced group. The difference in the mean score between the two groups is only significant for "ulcerations". The endoscopic variables "Vascular pattern", "Erosions", "Ulcerations" and Friability" explained the overall endoscopic activity index. Even though the present result is quite satisfactory, there is a potential of improvement. Improved grading systems might contribute to improve the consistency of endoscopic descriptions.

## Background

Ulcerative colitis (UC) is one of the major challenges in a gastroenterology practice and the endoscopic findings are essential in evaluating the disease activity [1,2]. Nevertheless, our previous work demonstrated that the detailed description of mucosal inflammation is quite often unsatisfactory in the endoscopy reports [3]. The severity of mucosal inflammation in UC is usually evaluated with Barons' endoscopic activity index (EAI), introduced and validated forty years ago [4]. Later minor modifications have never been validated [5-7]. This index assesses the four items "vascular pattern", "friability", "erosions" and "ulcerations" [7], but even "erythema", "oedema", "granularity", and "blood in lumen" are signs of mucosal inflammation. However, it is not clear to what extent these different endoscopic features contribute to the overall assessment of the EAI [4].

Endoscopic findings are most frequently assessed on fixed point scales, or simply described by dichotomous variables (present/absent) [8]. However, endoscopic features of mucosal inflammation are continuous variables posing potential drawbacks with discrete scales for scoring. Previous studies have shown benefits of visual analogue scale (VAS) in the assessment of mucosal lesions while at the same time improving study power [9-12].

The aim of this study was to evaluate how endoscopists assess signs of mucosal inflammation in ulcerative colitis, the inter-observer agreement, the variance of the mean score, and the influence of the observers' experience, as well as the correlation between the eight endoscopic features and their individual contribution to the EAI.

## Methods

Five patients, presenting with varying degrees of ulcerative colitis were admitted for endoscopic examination. The examination was videotaped with a Sony DV CAM DSR-20 MDP digital video recorder permitting lossless editing and copying. The videos were edited and a clip of about 30 s from each patient was shown by means of a high resolution video projector to an audience of endoscopists. Fifteen experienced endoscopists, (more than 750 colonoscopies), and 21 inexperienced (less than 200 colonoscopies) were asked to assess eight endoscopic signs of colitis and perform an overall score of the mucosal inflammation on a VAS according to the EAI. (Figure 1, Table 1)

For each clip the mean EAI of the experienced group was used as the "gold standard". To determine the inter-observer agreement the observers' scores for the five video clips were inter-rated by transforming the VAS score to an ordinal value ranging from 1-5.

## Statistical analysis

All continuously distributed variables are expressed by mean values with standard deviations and 95% confidence intervals [13]. All tests were carried out two-tailed, with a significance level of 5%. In order to analyse the inter observer agreement the following agreement procedure was used. The inter-rated scores of the 5 video clips for all observers were compared to the "Gold Standard", expressed in contingency tables and the inter-observer agreement presented by a  $\kappa$ -value [14]. The strength of agreement is respectively very good for a  $\kappa$ -value between 0.81-1.00, good for 0.61-0.80, moderate for 0.41-0.60 and fair for 0.21-0.40 [13]. Kappa analysis was used for comparison of experienced and inexperienced observers with regard to agreement [14].

Observers' scores were compared using analysis of variance (ANOVA) [15]. The coefficient of variance normally expressed as the relation  $SD/\bar{X}$  is inappropriate to compare scores of wide range. This coefficient was modified and expressed as the relation  $SD/MP$  (MP = midpoint of the scale).

Pearson linear correlation analysis was performed in order to study the correlation pattern between variables within groups of observer. In order to express EAI as a function of the set of independent variables, general linear models (GLM) was used [15].

To study the agreement between EAI and the Gold Standard within groups of observers the agreement analysis was used [13,16]. The two sets of measurements were compared using Student T-test for paired samples [13]. The EAI and the Gold Standard were plotted against each other and Pearson linear regression analysis performed [13]. The intercept was tested against 0 and the regression coefficient against 1 by using the studied test methodology [15]. The agreement limits and the agreement index were calculated [13,15].

## Results

### Inter-observer agreement

The observers rating of the endoscopic findings are represented in a contingency table (Table 2). The kappa analysis showed that the inter-observer agreement ranged from very good to moderate in the experienced group and from very good to fair in the inexperienced group. In the experienced group it was respectively *very good* for the items "erosions", "ulcerations", "friability" and "EAI". It was *good* for the findings "erythema", "oedema", "granularity", and "blood in lumen" and *moderate* for "vascular pattern". In the inexperienced group this agreement was *very good* for the item "friability", *good* for the items "ulceration" and "EAI", *moderate* for the items "erythema", "oedema", "granularity", "blood in lumen" and "erosions" and only *fair* for the item "vascular pattern" (Table 3).

Vascular pattern	Normal	Absent
Erythema	Normal	Absent
Oedema	Absent	Present
Granularity	Absent	Present
Blood in lumen	Absent	Present
Erosions	None	Multiple
Ulcerations	None	Multiple
Friability	No bleeding	Spontaneous bleeding
EAI	0	3

**EAI**

- 0=Normal      Normal, visible vascular pattern
- 1=Mild        Reduced vascular pattern, some vulnerability
- 2=Moderate   Absent vascular pattern, vulnerable mucosa, erosions
- 3=Severe      Spontaneous bleeding, ulcerations

**Figure 1**  
**Visual analogue scales for the assessment of endoscopic features**

**Table 1: Definition of endoscopic features of inflammation**

Vascular pattern	Changes in normal vascular pattern, decreased or absent
Erythema	Reddening of the mucosa
Oedema	Swollen mucosa
Granularity	Micronodular mucosal surface
Blood in lumen	Free intraluminal blood
Erosions	Small superficial defect in the mucosa, of a white or yellow colour, with a flat edge
Ulcerations	Compared to erosions, larger and deeper excavated fibrin covered lesions
Friability	Bleeding mucosa either at contact or spontaneously

**Table 2: Observers' rating of endoscopic findings compared to the gold standard within the group of experienced and inexperienced endoscopists.**

Endoscopic variables	Endoscopist group	Gold standard	Normal	Endoscopist Score			
				Mild	Mild-Moderate	Moderate Severe	Severe
Vascular pattern	Inexperienced (N = 21)	Normal	19	2	0	0	0
		Mild	1	7	9	3	1
		Mild-Moderate	1	7	5	6	2
		Moderate-Severe	0	5	6	7	3
		Severe	0	0	1	5	15
	Experienced (N = 15)	Normal	15	0	0	0	0
		Mild	0	6	8	0	1
		Mild-Moderate	0	5	5	4	1
		Moderate-Severe	0	3	1	9	2
		Severe	0	1	1	2	11
Granularity	Inexperienced (N = 21)	Normal	16	5	0	0	0
		Mild	5	10	4	2	0
		Mild-Moderate	0	3	14	3	1
		Moderate-Severe	0	4	3	13	1
		Severe	0	0	0	2	19
	Experienced (N = 15)	Normal	15	0	0	0	0
		Mild	0	11	4	0	0
		Mild-Moderate	0	1	9	5	0
		Moderate-Severe	0	3	2	8	2
		Severe	0	0	0	2	13
Erosions	Inexperienced (N = 21)	Normal	14	5	2	0	0
		Mild	7	7	6	1	0
		Mild-Moderate	0	8	11	2	0
		Moderate-Severe	0	0	2	15	4
		Severe	0	0	0	4	17
	Experienced (N = 15)	Normal	14	1	0	0	0
		Mild	1	13	1	0	0
		Mild-Moderate	0	1	13	1	0
		Moderate-Severe	0	0	0	12	3
		Severe	0	0	1	2	12
Ulcerations	Inexperienced (N = 21)	Normal	14	6	1	0	0
		Mild	6	10	4	1	0
		Mild-Moderate	1	5	13	2	0
		Moderate-Severe	0	0	3	18	0
		Severe	0	0	0	0	21
	Experienced (N = 15)	Normal	14	0	1	0	0
		Mild	0	12	3	0	0
		Mild-Moderate	1	3	11	0	0
		Moderate-Severe	0	0	0	13	2
		Severe	0	0	0	2	13

**Table 2: Observers' rating of endoscopic findings compared to the gold standard within the group of experienced and inexperienced endoscopists. (Continued)**

Friability	Inexperienced (N = 21)	Normal	17	3	1	0	0
		Mild	3	11	7	0	0
		Mild-Moderate	1	6	9	5	0
		Moderate-Severe	0	1	5	13	2
		Severe	0	0	0	2	19
	Experienced (N = 15)	Normal	15	0	0	0	0
		Mild	0	13	1	1	0
		Mild-Moderate	0	2	11	2	0
		Moderate-Severe	0	0	3	12	0
		Severe	0	0	0	0	15

**Table 3: Inter-observer agreement in the rating of mucosal lesion characterising ulcerative colitis**

Endoscopic findings	Exp		Inexp		p-value
	κ	CI	κ	CI	
Vascular pattern	0.51	0.38–0.66	0.38	0.26–0.50	0.05
Erythema	0.61	0.48–0.75	0.51	0.39–0.63	N.S.
Oedema	0.65	0.52–0.78	0.53	0.42–0.63	N.S.
Granularity	0.68	0.56–0.81	0.60	0.49–0.72	N.S.
Blood in lumen	0.76	0.65–0.88	0.57	0.46–0.69	<0.05
Erosions	0.81	0.71–0.92	0.51	0.39–0.63	<0.05
Ulcerations	0.80	0.69–0.91	0.65	0.55–0.76	<0.05
Friability	0.85	0.76–0.94	0.57	0.46–0.69	<0.05
Grade of inflammation	0.97	0.92–1.00	0.79	0.71–0.49	<0.05

**Mean score**

*Normal Mucosa*

Comparing experienced to inexperienced endoscopists there was no significant difference between the mean score of any endoscopic findings, but the coefficient of variance of the mean score was significantly higher (p < 0,01) in the inexperienced group when scoring "erythema" (Table 4)

*Mucosal inflammation*

A significant difference in the observed mean score was only seen in the assessment of ulcerations in moderate – severe and severe disease (p = 0.05). The coefficient of variance was significantly higher in the inexperienced group for the scoring of "erosions", "ulcerations" "blood in lumen" and "EAI" in mild disease (p < 0.01), for the scoring of vascular pattern" and "blood in lumen" in mild-moderate disease (p < 0,01), for the scoring of "vascular pattern", erythema and oedema in moderate-severe disease (p < 0.01, p < 0.01, p = 0.04 respectively) and for the scoring of "vascular pattern"erythema", "erosions" and "friability" in severe disease (p = 0.04, p < 0.01, p < 0.01, p = 0.01 respectively) (Table 4).

**Inter-correlation between endoscopic findings and the EAI**

Significant correlation was found between all the observed endoscopic variables in both groups of endoscopists (Table 5). In the group of experienced endoscopists, the set of endoscopic variables "Vascular pattern", "Erosions", "Ulcerations" and Friability" linearly explains 92% of the variation in EAI. By including interaction between the observed endoscopic variables, the explanation increases to 97%. "Erosions" with its interactions was found to be the most important endoscopic variable for the experienced endoscopists explaining the variation in EAI.

By including "Granularity" in the set, 91% of the variation in EAI was explained in the inexperienced group. This explanation increased to 95% by including interactions between the observed variables.

**Correlation between the observed EAI and the "Gold standard"**

Comparing the EAI to the "Gold standard" no significant difference was detected between the two groups of endoscopists (Table 6). However, the linear relationship

**Table 4: Mean score for the endoscopic assessment of mucosal lesions comparing experienced and inexperienced endoscopists**

		Normal		Mild		Mild-Moderate		Moderate-Severe		Severe	
		Exp	Unexp	Exp	Unexp	Exp	Unexp	Exp	Unexp	Exp	Unexp
Vascular pattern	$\bar{X}$ (SD)	0,4 (0,6)	0,5 (0,8)	7,7 (1,7)	6,8 (2,5)	8,1 (1,0)	7,4 (2,4)	8,2 (0,9)	7,2 (2,6)	8,9 (0,5)	8,9 (0,8)
	CV(%)	13%	17%	34%	53%	21%	51%	19%	55%	11%	17%
Erythema	$\bar{X}$ (SD)	1,0 (1,4)	2,3 (3,4)	6,7 (2,2)	6,9 (2,9)	6,9 (2,1)	6,5 (3,0)	8,1 (0,9)	7,6 (2,4)	9,0 (0,4)	8,8 (0,8)
	CV(%)	30%	72%	47%	62%	45%	64%	19%	51%	9%	17%
Oedema	$\bar{X}$ (SD)	0,4 (0,6)	0,7 (1,1)	5,3 (2,9)	5,5 (3,1)	5,9 (2,1)	6,8 (2,8)	7,1 (1,9)	6,4 (3,2)	8,7 (0,7)	8,9 (0,7)
	CV(%)	13%	23%	62%	66%	45%	60%	40%	68%	15%	15%
Granularity	$\bar{X}$ (SD)	0,5 (0,6)	0,5 (0,5)	4,6 (3,2)	2,7 (3,2)	5,9 (2,7)	6,3 (2,8)	6,5 (2,3)	6,2 (3,2)	8,7 (0,8)	8,8 (1,1)
	CV(%)	13%	11%	68%	68%	46%	60%	49%	68%	17%	23%
Blood in lumen	$\bar{X}$ (SD)	0,4 (0,6)	0,4 (0,4)	0,5 (0,5)	1,7 (3,0)	1,2 (1,4)	2,3 (3,2)	5,4 (2,7)	4,6 (3,7)	8,8 (0,7)	8,8 (0,9)
	CV(%)	13%	9%	11%	64%	30%	68%	57%	79%	15%	17%
Erosions	$\bar{X}$ (SD)	0,3 (0,5)	0,3 (0,3)	0,9 (1,2)	2,6 (3,3)	3,8 (2,9)	3,2 (3,4)	7,4 (1,7)	7,0 (2,7)	8,6 (0,8)	9,0 (0,4)
	CV(%)	11%	6%	26%	70%	62%	72%	36%	57%	17%	8%
Ulcerations	$\bar{X}$ (SD)	0,3 (0,3)	0,3 (0,3)	0,4 (0,3)	1,3 (2,3)	1,1 (2,3)	2,1 (2,2)	6,2 (2,9)	4,1 (3,2)	7,0 (2,9)	8,6 (1,8)
	CV(%)	6%	6%	6%	49%	49%	47%	62%	68%	55%	38%
Friability	$\bar{X}$ (SD)	0,4 (0,5)	0,4 (0,5)	1,7 (2,0)	1,4 (2,1)	2,4 (2,1)	2,8 (2,8)	5,0 (2,9)	6,1 (2,5)	8,8 (0,5)	8,9 (0,9)
	CV(%)	11%	11%	43%	45%	60%	60%	62%	53%	11%	19%
EAI	$\bar{X}$ (SD)	0,3 (0,3)	0,3 (0,5)	2,7 (0,9)	2,9 (2,3)	3,9 (1,3)	3,9 (2,1)	7,3 (1,1)	6,3 (1,8)	9,0 (0,6)	9,1 (0,4)
	CV(%)	6%	11%	19%	49%	28%	45%	23%	38%	13%	9%

X = Mean score, SD = Standard deviation, CV(%) = Coefficient of variance, Exp = Experienced, Inexp = Inexperienced, Vasc patt = Vascular pattern, Blood in lu = blood in lumen, EAI = Endoscopic activity index

**Table 5: Correlation between the endoscopic variables within the group of experienced and inexperienced endoscopists**

		Vascular pattern	Erythema	Oedema	Granularity	Blood in lumen	Erosions	Ulcerations	Friability	EAI
Inexperienced Endoscopists	Vascular pattern	I	0,79	0,85	0,72	0,58	0,65	0,52	0,61	0,74
	Erythema	0,79	I	0,73	0,64	0,55	0,58	0,47	0,59	0,65
	Oedema	0,85	0,73	I	0,79	0,58	0,61	0,54	0,59	0,71
	Granularity	0,72	0,64	0,79	I	0,62	0,56	0,62	0,66	0,79
	Blood in lumen	0,58	0,55	0,58	0,62	I	0,81	0,73	0,90	0,81
	Erosions	0,65	0,58	0,61	0,56	0,81	I	0,76	0,84	0,86
	Ulcerations	0,52	0,47	0,54	0,62	0,73	0,76	I	0,77	0,83
	Friability	0,61	0,59	0,59	0,66	0,90	0,84	0,77	I	0,87
	EAI	0,74	0,65	0,71	0,79	0,81	0,86	0,83	0,87	I
Experienced Endoscopists	Vascular pattern	I	0,90	0,84	0,75	0,50	0,61	0,46	0,56	0,73
	Erythema	0,90	I	0,83	0,77	0,61	0,68	0,56	0,65	0,77
	Oedema	0,84	0,83	I	0,91	0,71	0,73	0,62	0,75	0,81
	Granularity	0,75	0,77	0,91	I	0,65	0,70	0,64	0,70	0,79
	Blood in lumen	0,50	0,61	0,71	0,65	I	0,84	0,80	0,92	0,86
	Erosions	0,61	0,68	0,73	0,705	0,84	I	0,86	0,82	0,92
	Ulcerations	0,46	0,56	0,62	0,645	0,80	0,86	I	0,77	0,83
	Friability	0,56	0,65	0,75	0,705	0,92	0,82	0,77	I	0,84
	EAI	0,73	0,77	0,81	0,79	0,86	0,92	0,83	0,84	I

**Table 6: Agreement between "EAI" and "Gold standard" in the group of inexperienced and experienced endoscopists.**

	Inexperienced endoscopists	Experienced endoscopists
Mean of [17]	4,57	4,47
SD of {(EAI+Gold standard)/2}	3,18	3,19
Mean of {Diff.(EAI+Gold standard)}	-0,14	-0,01
SD of {Diff.(EAI+Gold standard)}	1,65	0,88
Agreement Index (AI)	0,28	0,61
Correlation between the mean and the absolute difference of EAI and Gold Standard	0,04	0,11

between the two variables was found to differ significantly ( $p < 0.01$ ) from the line of equality ( $Y = X$ ) in the group of inexperienced endoscopists (Fig. 2a). No such differences were detected for the experienced endoscopists (Fig. 2b). The agreement index in the inexperienced group was found fairly good compared to excellent for the experienced endoscopists and the percent of outliers was 6.7% in the inexperienced group compared to 5.3% in the experienced (Fig. 2c & 2d). (Table 6). No significant correlation was found between the mean and the absolute difference of EAI and "Gold Standard".

## Discussion

This study demonstrated an overall satisfactory inter-observer agreement in rating the endoscopic lesions, but significantly better in the experienced group. Comparing experienced to inexperienced endoscopists the only significant difference observed was in the mean score of "ulcerations" but the inter-observer variation of the mean score was more pronounced in the inexperienced group particularly in the assessment of mild and moderate disease.

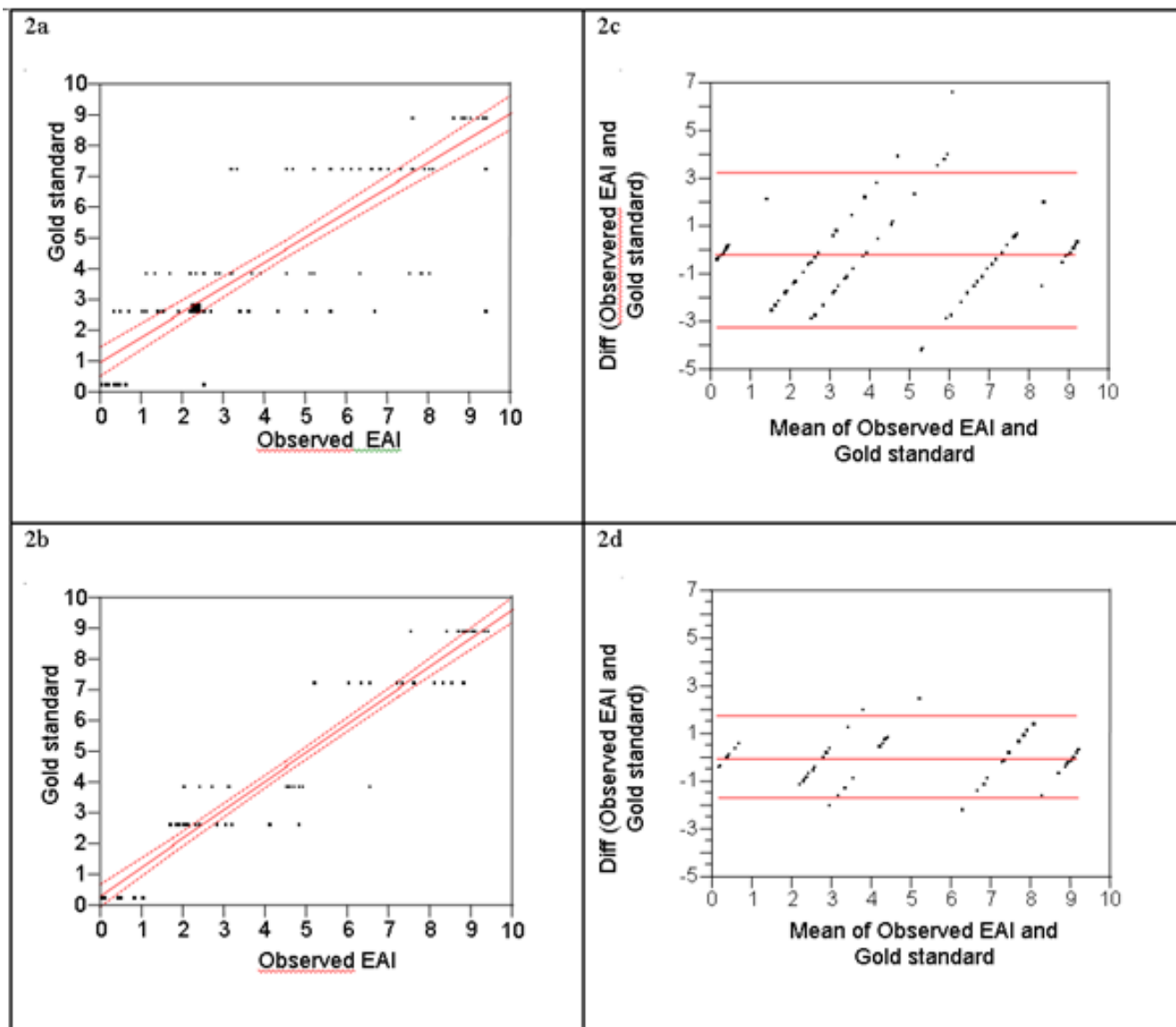
These results are to some extent in contrast to earlier studies. The study of Baron et al. in the era of fiberendoscopic instruments presented the advantage that the endoscopic findings were assessed live in 60 patients by three different observers [4], nevertheless this study was of limited value because not all types of mucosal lesions were present and only three observers working in team for 20 months prior to the study were involved. However, various endoscopic lesions showed a variable degree of observer agreement and the only feature presenting a good agreement was friability. No attempt was made to correlate the endoscopic findings neither to the histological nor the clinical severity. According to the result they proposed an endoscopic grading of "activity" in ulcerative colitis still in use, further evaluation of this grading has not been performed. Orlandi et al. demonstrated variable inter-observer agreement, when 46 still images of ulcerative colitis were evaluated by four experienced and 11 inexperienced endoscopists [17]. Minoli et al. emphasised the importance of common recognition and definition of endo-

scopic findings. Prior to their study the observers were briefed to agree on the definition of different endoscopic signs, and the study was only started when an agreement on the definition of lesions was reached. They found a good concordance in the assessments of 12 out of 16 findings in IBD [18]. Some of the terms and their attributes in these studies are either not currently used or are not according to the Minimal Standard Terminology version 2.0 (MST) approved by most of the international endoscopy organisations (OMED, ESGE, ASGE, JSGE) [19].

The differences between the present study and the previous ones might be explained by differences in study design, and analysis. In fact neither of the studies correlated the agreement and the variance to the degree of inflammation, and the methods of presenting the images was different, as was the method of assessing endoscopic findings.

The difference in mean score and in coefficient of variance was lowest in the assessment of the EAI, this might partially be due to the four step score normal-mild-moderate-severe. An identical grading of "erythema", "oedema", and "granularity" might reduce the differences between experienced and inexperienced endoscopists as well as the inter-observer variation of the mean score. This grading was not used because it is not a part of the MST but should probably be introduced as attribute values to these items. The VAS might be less appropriate for the scoring of countable lesions like ulcers and erosions and for sensitive features of mucosal inflammation present in nearly any degree of inflammation like absence of vascular pattern explaining the tight score range from 7.7 in mild disease to 8.9 in severe disease. This might bias the inter-rating of the "vascular pattern" and reduce the inter-observer agreement and explain why the present study showed a lower kappa value for "vascular pattern" compared to the observations of Orlandi et al. The inter-observer agreement of all other endoscopic signs was superior in the present study.

There was a good correlation between the grading of all endoscopic features. In the experienced group four

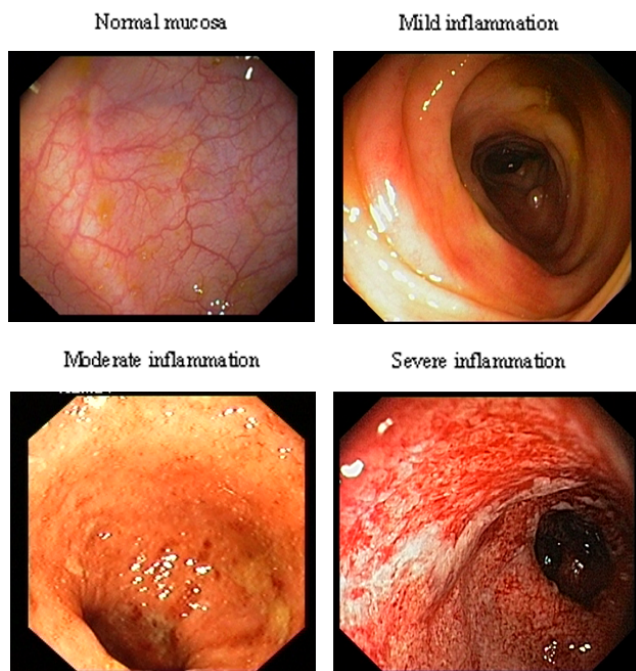


**Figure 2**  
**Regression analysis and agreement plots for experienced and inexperienced endoscopists** The linear relation between "Observed EAI" and "Gold standard" in the group of inexperienced (2a) and experienced (2b) endoscopists. The results expressed by the linear regression line with 95% confidence interval for the lines. The agreement plot in the group of inexperienced (2c) and experienced (2d) endoscopists. The results expressed by the line of mean with 95% confidence interval for the lines.

features "vascular pattern", "friability", "erosions", and "ulcerations" explained the EAI, but changing the attributes values of other features might affect this result. Even if in the inexperienced group granularity was added to explain the EAI we still assume that the first four items are sufficient to describe inflammatory changes of the mucosa.

Inflammatory mucosal changes present in other GI diseases are also graded, eg oesophagitis with the LA-classification [20]. However the LA-classification, evaluating the extent of mucosal breaks in a short segment, is quite different from the more complex score of diffuse mucosal inflammation in UC characterised by at least 4 different endoscopic features. This might explain why Lundell et al





**Figure 3**  
Endoscopic images of ulcerative colitis

did not demonstrate any difference between experienced and inexperienced endoscopists [20].

The "Gold Standard" was chosen to be the mean EAI of the experienced endoscopists because one of the main purposes was to assess the influence of experience.

Technical factors like natural light spectre reproduction of the CCD chip might contribute to the variance in the grading of erythema [21].

The live examination is of course the gold standard for the evaluation of the colonoscopy with the possibility to re-examine any suspicious lesion, but it is not practically feasible to show the same endoscopy to a sufficient number of observers at the same time. This live situation is probably best imitated by reviewing high quality videotapes with the possibility for the observer to play and rewind the movie. It permits more detailed examination of the mucosa. The friability might be easier to assess if the intubation and the retraction of the scope is recorded.

In routine examinations the problems of variance in assessing endoscopic findings might be reduced by systematic image documentation of gastrointestinal endoscopies as recommended by ESGE [22] (fig 3).

## Conclusions

The inter-observer agreement in the rating of endoscopic features characterising ulcerative colitis is satisfactory in both groups of endoscopists but significantly higher in the experienced group. The difference in the mean score between the two groups is only significant for "ulcerations". But the variance of the mean score might influence the follow up of the patients and bias the results in studies especially of mild and moderate ulcerative colitis. The endoscopic variables "Vascular pattern", "Erosions", "Ulcerations" and Friability" explained the overall endoscopic activity index. Even though the present result is quite satisfactory, there is a clear potential of improvement. Improved grading systems might contribute to improve the consistency of endoscopic descriptions.

## Competing interests

None declared.

## Abbreviations

DV: Digital Video

EAI: Endoscopic Activity Index

GML: General Linear Models

MST: Minimal Standard Terminology

UC: Ulcerative Colitis

VAS: Visual Analogue Scale

## Authors' contributions

Tdl the main designer, the coordinator of the study and the corresponding author also carried out the video presentation, participated in the statistical analysis, and drafted the manuscript. SL performed the statistical analysis. LA participated in its design and in drafting the manuscript. All authors read and approved the final manuscript.

## References

1. **The role of colonoscopy in the management of patients with inflammatory bowel disease.** American Society for Gastrointestinal Endoscopy. *Gastrointest Endosc* 1998, **48**:689-690.
2. Marteau P: **Inflammatory bowel disease.** *Endoscopy* 2002, **34**:63-68.
3. De Lange T, Moum BA, Tholfsen JK, Larsen S, Aabakken L: **Standardization and quality of endoscopy text reports in ulcerative colitis.** *Endoscopy* 2003, **35**:835-840.
4. Baron JH, Connell AM, Lennard-Jones JE: **Variation Between Observers in Describing Mucosal Appearances in Proctocolitis.** *Br Med J* 1964, **1**:89-92.
5. Gomes P, du BC, Smith CL, Holdstock G: **Relationship between disease activity indices and colonoscopic findings in patients with colonic inflammatory bowel disease.** *Gut* 1986, **27**:92-95.
6. Holmquist L, Ahren C, Fallstrom SP: **Clinical disease activity and inflammatory activity in the rectum in relation to mucosal inflammation assessed by colonoscopy. A study of children**

- and adolescents with chronic inflammatory bowel disease. *Acta Paediatr Scand* 1990, **79**:527-534.
7. Roseth AG, Aadland E, Jahnsen J, Raknerud N: **Assessment of disease activity in ulcerative colitis by faecal calprotectin, a novel granulocyte marker protein.** *Digestion* 1997, **58**:176-180.
  8. D'Argenio G, Cosenza V, Riegler G, Della VN, Deritis F, Mazzacca G: **Serum transglutaminase correlates with endoscopic and histopathologic grading in patients with ulcerative colitis.** *Dig Dis Sci* 2001, **46**:649-657.
  9. Aabakken L, Larsen S, Osnes M: **Visual analogue scales for endoscopic evaluation of nonsteroidal anti-inflammatory drug-induced mucosal damage in the stomach and duodenum.** *Scand J Gastroenterol* 1990, **25**:443-448.
  10. Larsen S, Aabakken L, Lillevold PE, Osnes M: **Assessing soft data in clinical trials.** *Pharmaceutical Medicine* 1991, **5(1)**:29-36.
  11. Osnes M, Larsen S, Eidsaunet W, Thom E: **Effect of diclofenac and naproxen on gastroduodenal mucosa.** *Clin Pharmacol Ther* 1979, **26**:399-405.
  12. Aabakken L, Olaussen B, Mowinckel P, Osnes M: **Gastroduodenal lesions associated with two different piroxicam formulations. An endoscopic comparison.** *Scand J Gastroenterol* 1992, **27**:1049-1054.
  13. Altman DG: *Practical statistics for medical research* London: Chapman and Hall; 1991.
  14. Agresti A: *Categorical data analysis* New York: Wiley; 1990.
  15. Kleinbaum DG: *Applied regression analysis and other multivariable methods* Boston: PWS-Kent Pub Co; 1988.
  16. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1**:307-310.
  17. 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**:539-541.
  18. Minoli G, Pera A, Barletti C, Vanni R, Rocca R: **Endoscopy concordance in diagnosis of IBD.** *Acta Endoscopica* 1990, **20(3 SUPPL A)**:273-277.
  19. Delvaux M, Korman LY, Armengol-Miro JR, Crespi M, Cass O, Hagenmuller F, et al.: **The minimal standard terminology for digestive endoscopy: introduction to structured reporting.** *Int J Med Inf* 1998, **48**:217-225.
  20. Lundell LR, Dent J, Bennett JR, Blum AL, Armstrong D, Galmiche JP, et al.: **Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification.** *Gut* 1999, **45**:172-180.
  21. Vakil N, Knyrim K, Everbach EC: **The appreciation of colour in endoscopy.** *Baillieres Clin Gastroenterol* 1991, **5**:183-194.
  22. Rey JF, Lambert R: **ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy.** *Endoscopy* 2001, **33**:901-903.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-230X/4/9/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

