ABSTRACT

Background: Ulcerative colitis (UC), a chronic inflammatory disorder that affects the large bowel, may lead to significant morbidity and early mortality. Medical treatment for UC targets several endpoints in clinical practice and clinical trials including clinical, endoscopic, radiologic, biochemical and histologic disease activity. Because histologic disease activity has been linked with long-term sequelae, even in the absence of endoscopically visible disease, healing of the inner colonic lining (mucosal healing) is a potentially important treatment goal. Although many histologic indices have been described, their operating properties are unknown.

Objectives: To conduct a systematic review to identify partially or fully validated histologic indices of disease activity in UC and to test the identified indices for reliability by comparing intra- and inter-rater intra-class correlation coefficients (ICC).

Methods: A 2-phase study was undertaken. In Phase 1, a systematic review was conducted to identify all UC histologic indices and evaluate their item selection process and use of psychometric methods. In Phase 2, five expert gastroenterology pathologists rated disease activity on 150 digitized histology slides on three occasions using the previously identified indices and a 100 mm visual analog scale (VAS) global measure of histologic disease activity. Histology slides used in this study were from a randomized controlled trial conducted in 2005 that evaluated the effectiveness of, a now proven to be effective, leukocyte trafficking inhibitor called vedolizumab. Intra- and inter-rater
agreement was measured using ICC with 95% confidence intervals. Subsequently, a consensus process was conducted among a panel of expert pathologists to understand major sources of disagreement.

**Results:** Eighteen histologic indices were identified of which 2 were developed using a clearly defined item selection process and had undergone some appropriate psychometric testing. These two indices, the Geboes Score (GS) and modified Riley Score (MRS) showed “substantial” to “almost perfect” intra-rater agreement but only “moderate” inter-rater agreement. Based on the qualitative consensus results, rules and standard definitions were developed to aid in future studies.

**Conclusions:** Most histologic indices of disease activity in UC did not employ psychometric testing during their development phase. The GS and MRS histologic indices showed “substantial” to “almost perfect” intra-rater agreement, but less than satisfactory inter-rater agreement. We recommend utilizing standardized definitions and rules to minimize disagreement between pathologists in the development and validation of a histologic index for disease activity in UC. Further studies that implement these recommendations are underway.

**Key Words:** Ulcerative colitis, disease activity, index, histologic, agreement.
RÉSUMÉ

Contexte: La colite ulcéreuse (CU), une maladie inflammatoire chronique qui affecte le gros intestin, peut conduire à une morbidité importante et une mortalité précoce. Le traitement médical de l'UC cible plusieurs points de terminaison dans la pratique clinique et les essais cliniques, y compris endoscopique, radiologique, l'activité clinique de la maladie, biochimiques et histologiques. Parce que l'activité de la maladie histologique a été liée à des séquelles à long terme, même en l'absence de la maladie par voie endoscopique visible, la guérison de la muqueuse colique intérieure (cicatrisation des muqueuses) est un objectif de traitement potentiellement important. Bien que de nombreux indices histologiques ont été décrits, leurs propriétés de fonctionnement ne sont pas connus.

Objectifs: effectuer un examen systématique pour identifier les indices histologiques partiellement ou entièrement validées de l'activité de la maladie dans la RCH et pour tester les indices identifiés pour la fiabilité en comparant inter-et intra-juges intra-classe coefficients de corrélation (CPI).

Méthodes: Une étude de phase 2 a été réalisée. Dans la phase 1, un examen systématique a été menée afin d'identifier tous les indices histologiques de communications unifiées et d'évaluer leur processus de sélection de l'article et l'utilisation de méthodes psychométriques appropriées. Dans la phase 2, cinq gastroentérologie pathologistes experts notés activité de la maladie sur 150 lames histologiques numérisées à trois reprises à l'aide des indices précédemment identifiées et une échelle de 100 mm visuelle analogique (EVA)
de mesure globale de l'activité de la maladie histologique. Diapositives histologiques utilisées dans cette étude étaient d'un essai contrôlé randomisé mené en 2005 qui a évalué l'efficacité de, un maintenant prouvé pour être efficace, un inhibiteur de la traite des leucocytes appelé vedolizumab. Inter et intra-évaluateur a été mesurée à l'aide de la CPI avec des intervalles de confiance à 95%. Par la suite, un processus de consensus a été mené auprès d'un panel de médecins experts pour comprendre les principales sources de désaccord.

**Résultats:** Dix-huit indices histologiques ont été identifiés, dont 2 ont été développés en utilisant un processus de sélection d'un élément clairement défini et avaient subi des tests psychométriques approprié. Ces deux indices, les Geboes Note (GS) et Score modifié Riley (MRS) ont montré "substantielle" de l'accord "presque parfait" intra-évaluateur mais seul accord "modérée" inter-évaluateur. Sur la base des résultats du processus de consensus qualitatives, des règles et des définitions normalisées ont été élaborées pour aider dans les études futures.

**Conclusions:** La plupart des indices histologiques de l'activité de la maladie en UC n'ont pas utilisé des tests psychométriques au cours de leur phase de développement. Les indices histologiques GS et MRS ont montré à un accord intra-évaluateur "substantielle" "presque parfait", mais moins que satisfaisante accord inter-juges. Nous recommandons l'utilisation de définitions et règles standardisées afin de minimiser le désaccord entre les pathologistes dans le développement et la validation d'un indice histologique de l'activité de la maladie.
dans la RCH et d'autres études qui mettent en œuvre ces recommandations sont en cours.

**Mots clés:** colite ulcéreuse, l'activité de la maladie, index, histologiques, accord.
Acknowledgments:

I offer my sincere gratitude to both of my thesis supervisors Dr. Brian Feagan and Dr. Maida Sewitch for all of the help and guidance they provided me while preparing this thesis. I especially thank Dr. Feagan for enlarging my vision of science, expanding my passion towards clinical research and providing coherent answers to my endless questions and Dr. Sewitch for her continuous feedback and constant push towards perfection. I would like to also acknowledge Dr. Barrett Levesque, Peggy Vandervoort, Dr. Larry Stitt, Dr. GuangYong Zou, Dr. William Sandborn, Dr. Geert D’Haens and all of my fellow colleagues and co-investigators at the wonderful Robarts Research Institute for their unconditional support. I thank the faculty, staff and my colleagues at McGill University for inspiring me to continue my work in this field. Grateful and warm thanks are reserved for my loving parents who have supported me throughout my years of education, both morally and financially, and to my siblings for always being there for me. Last but never least; special thanks are owed to my dear wife Solafa for her years of love, patience and everlasting support.
Preface and Contributions of Authors:

This dissertation in total represents research completed under the supervision of Dr. Brian Feagan and Dr. Maida Sewitch and is presented in three main chapters. Chapter One includes a comprehensive literature review presented as a systematic review that was published in the journal Inflammatory Bowel Diseases. Chapter Two presents original work from a clinical study that evaluated agreement among pathologists on measures of histologic disease activity in ulcerative colitis. This chapter has resulted in a manuscript that is under peer review at a major gastroenterology journal, and findings have been published in abstract form following presentations at the 9th European Crohn’s and Colitis Organization (ECCO)* and the 2014 Digestive Disease Week (DDW)† meetings. Chapter Three presents the strengths, limitations and conclusions of my dissertation including recommendations and future research plans.


Author contributions for Chapter 1: Mahmoud H Mosli (MHM), Brian G Feagan (BGF), and Barrett G Levesque (BGL) contributed to the conception and design of the study, analysis and interpretation of data, and drafting the article; William J Sandborn (WJS), Geert D'Haens (GD'H), Cynthia Behling (CB), Keith Kaplan (KK), David K Driman (DKD), John K MacDonald (JKM), Margaret K Vandervoort (MKV), Karel Geboes (KG), Lisa M Shackelton (LMS), and Kenneth A Baker (KAB) contributed to the analysis and interpretation of the data and revising the manuscript for important intellectual content.

Author contribution for Chapter 2: MHM, Maida Sewitch (MS), BGF, WJS, GD'H, Reena Khanna (RK), CB, KK, DKD, LMS, KAB, JKM, MKV, KG, Michael Valasek (MAV), Rish Pai (RP), Cord Langner (CL), Robert Riddell (RR), Noam Harpaz (NH), Michael Peterson (MP), Larry Stitt (LWS), GuangYong Zou (GYZ), and BGL were involved in the development of study concept and design. MHM, BGF, WJS, GD'H, RK, MKV, KG, BGL supervised the study. Acquisition, analysis, and interpretation of the data was done by MHM, BGF, WJS, GD'H, RK, CB, KK, DKD, LMS, KAB, JKM, MKV, KG, MAV, RP, CL, RR, NH, MS, MP, LWS, GYZ, BGL. Statistical analysis was primarily performed by LWS and GYZ. Drafting of the manuscript was done by MHM with assistance from BGF, LMS, LWS, GYZ, MS and BGL. MHM, BGF, WJS, GD'H, RK, CB, KK, DKD, LMS, KAB, JKM, MKV, KG, MAV, RP, CL, RR, NH, MS, MP, LWS, GYZ and BGL critically reviewed the manuscript for important intellectual content.
Author contributions for Chapter 3

MHM, with assistance from MS and BGF, authored chapter 3.

Ethics Approval: The histology slides analyzed in this study were obtained during the execution of a clinical trial protocol (Feagan et al, 2005). The informed consent obtained during the clinical trial complied with ICH-GCP and all applicable regulatory requirement(s). The consent of study subjects included the use of the collected data for other research purposes, and additional consent was not required. This study was reviewed by the Robarts Research Ethics Board to safeguard the rights of the study subjects from whom the endoscopy examinations were originally obtained. All of the pathologists involved in the agreement study were employed by Robarts Research Institute and their contractual agreement served as informed consent.
Introduction:

Ulcerative colitis (UC) is a chronic inflammatory disorder that can result in important morbidity and an increased risk of mortality. Evaluation of disease activity in UC is currently based on clinical symptom indices such as the Mayo Clinic Score (MCS), endoscopic scores such as the Modified Baron Score (MBS) and patient-reported outcomes such as the Inflammatory Bowel Disease Questionnaire (IBDQ) {Bitton et al Gastro 2001}. Histology is infrequently used to assess disease activity in clinical trials of therapy for UC. Although the potential reasons for this circumstance are complex they include: 1) the lack of validated instruments; 2) disagreement among pathologists in rating severity of inflammation; 3) the problem of sampling error due to “patchiness” of colonic inflammation; 4) presence of a lag between relief of symptoms, endoscopic remission and histologically defined resolution of inflammation and 5) lack of consensus regarding the clinical relevance of histologic changes. Despite these limitations, identification of a validated histologic index of disease severity in UC is necessary for the following two reasons: 1) histologic assessment of mucosal healing is likely a relevant treatment goal as existing data suggest that mucosal healing, defined either by histology or endoscopy, predicts important outcomes such as colectomy and corticosteroid-free remission {Cooney et al Trials 2007, D'Haens et al Gastro 2007, Floren et al Scand J Gastro 1987, Rutter et al Gastro 2004}; and 2) the literature supports a relationship between persistent inflammation and the development of colorectal cancer {Gupta et al Gastro 2007}. Given these observations; microscopic or histologic remission, also
known as “deep remission”, is being considered as a potential endpoint for both clinical trials and clinical practice {Froslie et al Gastro 2007, Truelove BMJ 1956}. As new medical therapies become available for UC, clinicians need to know which measures (e.g. symptoms, endoscopy, histopathology) best predict long-term health outcomes. Methodologically rigorous evaluative instruments are needed to adequately answer this question.

Although considerable progress has been made in the development of validated endoscopic instruments {Travis et al Gut 2012}, none of the histologic indices currently available for use as endpoints in clinical trials in UC have been fully validated. In order to objectively use histologic disease activity as an endpoint in clinical practice and clinical trials, a relevant index must be identified or developed and then validated. Accordingly, it is necessary to assess the currently available histologic indices to identify instruments that were developed using a formal methodological framework that included a methodologically rigorous item selection process and appropriate assessments of validity, reliability and responsiveness.

In this thesis, a systematic review of the literature was performed to identify all UC histologic disease activity indices and select instruments that were developed through a clearly defined item selection process and underwent partial or full psychometric testing. We then performed an agreement study to evaluate the reliability of the selected indices using previously obtained colonic biopsies as test material.
Preface to Chapter 1:

Chapter 1 provides a review of the literature on histologic disease activity indices in UC with the aim of identifying indices that were developed using a clearly defined item selection process and psychometric property testing. Results from this literature review served as the basis for the clinical study presented in Chapter 2. The scores of the indices identified were subsequently used to assess agreement among pathologists in the evaluation of histologic disease activity in UC.

This chapter lays the groundwork for the following chapters.
Chapter 1: Literature Search: A Systematic Review Of Histologic Disease Activity Indices For Ulcerative Colitis.

Title: Histologic evaluation of ulcerative colitis: a systematic review of disease activity indices

Authors: Mahmoud H Mosli, MD1,2,3 Brian G Feagan, MD1,3,4 William J Sandborn, MD3,5 Geert D’Haens, MD, PhD3,6 Cynthia Behling, MD7 Keith Kaplan, MD8 David K Driman, MD9 Lisa M Shackelton, PhD3 Kenneth A Baker, PhD3 John K MacDonald, MS1 Margaret K Vandervoorst, MS3 Karel Geboes, MD, PhD10 Barrett G Levesque, MD3,5

1Department of Medicine, Western University, London, Ontario, Canada
2Department of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia
3Robarts Clinical Trials, Inc, Robarts Research Institute, Western University, London, Ontario, Canada
4Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada
5Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA
6Department of Gastroenterology, Academic Medical Center, Amsterdam, The Netherlands
7Pacific Rim Pathology Medical Corporation, San Diego, CA, USA
Grant support: None

Abbreviations: Active Ulcerative Colitis Trial (ACT); confidence interval (CI); Digestive Diseases Week (DDW); hazard ratio (HR); inflammatory bowel disease (IBD); intraclass coefficient (ICC); modified Riley score (MRS); mucosal healing (MH); randomized controlled trial (RCT); relative risk (RR); ulcerative colitis (UC)

Disclosures: The authors have nothing to disclose

Writing assistance: None
ABSTRACT:

Background: Ulcerative colitis (UC) is an idiopathic inflammatory disorder. Currently, the main goals of treatment are to induce and maintain clinical and/or endoscopic remission. However, emerging evidence indicates that persistent disease activity on colonic biopsies in the setting of clinical or endoscopic remission is an independent predictor of poor health outcomes. A number of prior studies have proposed histologic indices for use in specific trials of UC. The aim of this study is to systematically review all existing histologic indices used to assess UC disease activity and identify those that were developed using an item selection process and have at least one clearly defined psychometric property (reliability, responsiveness and validity).

Methods: We performed a systematic review of histologic indices evaluating disease activity in UC. MEDLINE, EMBASE, PubMed, the Cochrane Library (CENTRAL), and Digestive Diseases Week (DDW) abstracts of randomized and/or controlled clinical trials were searched from inception to February 2013 for applicable studies. Data from these studies were reviewed and analyzed.

Results: After systematically applying inclusion criteria, we identified 108 scientific papers including 88 clinical studies and 21 clinical reviews. Eighteen indices of histologic activity in UC were identified and reviewed, two of which were constructed through a clearly outlined item selection process and have at least one clearly defined psychometric property, namely the Geboes and Riley scores.
**Conclusions:** Although multiple histologic scoring indices for assessment of UC disease activity currently exist, most of these indices were not developed using a formal validation process and their psychometric properties are poorly understood. Future studies are needed to address these deficiencies.

**Keywords:** Ulcerative colitis; histology; indices; outcome measures; disease activity
INTRODUCTION:

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) of unknown etiology with a wide spectrum of disease severity (1). UC can be complicated by toxic mega-colon and colorectal cancer (2). Pharmacologic management includes aminosalicylates, corticosteroids, purine antimetabolites, and tumor necrosis factor antagonists, used sequentially or in combination (3-5). Induction and maintenance of remission are important treatment goals; however, there is no universally accepted definition of remission and no consensus on the best way to assess disease activity.

In clinical practice, disease activity is assessed through evaluation of symptoms and severity of colonic inflammation by sigmoidoscopy or colonoscopy (6). The imprecision of this approach likely contributes to large variances in disease management and sub-optimal patient outcomes. Therefore, clinical investigators advocate for the use of quantitative endoscopic indices as outcome measures in randomized controlled trials (RCTs) (7, 8). Mucosal healing (MH), evaluated by defined endoscopic criteria, confers greater long-term benefit than symptom control (9, 10). Ardizzone et al prospectively evaluated a cohort of 157 newly diagnosed patients with moderate-to-severely active UC who received corticosteroid therapy and were followed for up to 5 years. Patients without complete MH were more likely to receive immunosuppressive therapy (hazard ratio [HR], 10.6; 95% confidence interval [CI], 2.2–51.0), have greater rates of hospitalization (HR, 3.6; 95% CI, 1.6–8.5) and undergo colectomy (HR, 8.4; 95% CI, 1.3–55.2) than those with complete MH (10). A population-based Norwegian
cohort study had similar findings (11). An endoscopic score of 0 (normal mucosa), 1 (light erythema or granularity), or 2 (granularity, friability, and bleeding, with or without ulcerations) 1 year after treatment initiation was associated with a significantly lower rate of colectomy at 5 years compared to higher endoscopic scores (relative risk [RR], 0.22; 95% CI: 0.06–0.79). The Active Ulcerative Colitis Trials (ACT-1 and ACT-2) of infliximab showed that MH, defined as an absolute endoscopy sub-score of 0 (inactive) or 1 (mild disease [erythema, decreased vascular pattern, mild friability]) at week 8, was associated with a lower rate of colectomy after 54 weeks than that observed in patients without MH ($P$=0.0004)(12). These findings suggest that treatments resulting in MH might yield better long-term outcomes than those based on symptom resolution.

While MH is associated with a favourable long-term outlook, endoscopy alone is a poor predictor of histologically defined healing (13, 14). Truelove and Richards first reported that histologic evidence of active inflammation was common in patients with endoscopically normal mucosa following successful induction therapy (15). This observation is clinically relevant as histology may be useful for the prediction of relapse (16). Patients with chronic UC in symptomatic and endoscopic remission with histologic evidence of acute inflammation had a 2- to 3-fold greater risk of relapse during a 12-month follow-up period, which was positively correlated with the severity of the inflammatory infiltrate (17). Patients whose inflammatory infiltrate was graded as severe had a 2-fold risk of relapse compared to those with lower scores. Bitton et al also found that the presence of
residual histologic inflammatory activity was an independent predictor of early clinical relapse in 74 patients with clinically and endoscopically quiescent UC (18). Similarly, a retrospective analysis of 75 adult patients with endoscopically inactive UC showed that basal plasmacytosis and a Geboes histologic score >3.1 (19) were associated with a marked increase in relapse rate (14). A stringent definition of remission that incorporates both endoscopic healing and complete resolution of the inflammatory infiltrate might be a valuable treatment goal, and if so, early assessment of microscopic healing may predict response to treatment. For these reasons, identification of histologic features of disease activity that can be accurately and reproducibly measured (20) is a clinical and research priority.

Multiple scoring systems have been developed to measure histologic features of UC including: degree of acute (Figure 1) or chronic (Figure 2) inflammatory cell infiltrates, the presence or absence of architectural distortion of colonic crypts, and the integrity of the colonic epithelium (21). Although these indices have potential value for both informing clinical practice and as outcome measures in clinical trials, little is known about the methodology used to create these scales and their psychometric properties. We therefore reviewed all existing histologic indices used to assess disease activity in UC and assessed their psychometric properties in an effort to characterize their potential role in clinical management and as outcome measures in clinical trials.
MATERIALS and METHODS:

The Cochrane Library (CENTRAL) (issue 2, 2013), MEDLINE, EMBASE, PubMed and Digestive Disease Week (DDW) abstracts were electronically searched from their inception dates (MEDLINE from 1946, EMBASE from 1974, DDW abstracts from 1980 and PubMed from 1966) to February 2013. Each database was searched for ‘ulcerative colitis’ AND (‘histology’ OR ‘pathology’ OR ‘immunohistochemistry’ OR ‘biopsy’) AND (‘index’ OR ‘indice’ OR ‘scale’ OR ‘score’ OR ‘Riley’ OR ‘Geboes’)‡.

All studies that utilized histologic indices of disease activity in patients with UC, including randomized and/or controlled trials, case-controlled studies, and cohort studies were included. Case reports, editorials, clinical guidelines, commentary, letters to the editor, and meeting reports were excluded. Clinical reviews were included for reference review and hand searching. Studies cited in the review articles, that were not identified through the literature search but were relevant and applicable, were added manually. Efforts were made to contact authors for missing information.

Two reviewers (MHM and KAB) independently screened citations and abstracts and retrieved full-text publications of all potentially eligible articles according to PRISMA guidelines. No language restrictions were applied; publications were translated into English if required. Indices were categorized according to their method of classification and degree of validation. The two

‡ The full literature search strategy for each database is presented in the appendix section.
reviewers assessed study eligibility and disagreements were resolved by consensus.

**RESULTS:**

Our literature search retrieved 4514 citations. After exclusion of duplicates (2179), 2335 papers were screened; 516 animal studies were identified and removed (Figure 3). Eligibility criteria were applied to the remaining citations. Sixty-four papers were identified by search and forty-five additional papers were added. Eighty-eight clinical studies were identified that included a novel histologic index or used a histologic index as a clinical endpoint. Eighteen indices with 2 types of scoring systems (stepwise [categorically progressive] and numerical [quantitative]) were identified. The more commonly used stepwise systems divided disease activity into subjectively assessed grades, whereas the quantitative systems numerically scored features. The 18 indices identified in our search are described in Table 1. The following description focuses on the most commonly cited indices.

**Truelove and Richards Index**

The first histologic index developed for UC was reported in a study of 111 serial biopsy specimens from 42 UC patients with varying stages of clinical and endoscopic activity and 24 controls without UC (15). Specimens were categorized as having: no, mild-to-moderate, or severe inflammation. Although no formal correlation analyses were performed, over half of the patients with clinical remission showed evidence of endoscopic or histologic activity. Histologic activity was also observed in 37% of biopsies from endoscopically
normal mucosa. This simple and subjective scoring system has been frequently used in clinical trials (22-24). One weakness of this index is that severe inflammation is imprecisely defined (i.e., ‘heavy’ infiltration by neutrophils and eosinophils, crypt abscesses, and erosions). The method of item selection was not clear but appears to be subjective in nature. The operating properties of this index were partially evaluated in a prospective study that assessed agreement between clinical (Simple Clinical Colitis Activity Index (SCCAI))(43), endoscopic (Baron score)(44), and histologic grading of disease activity. In this study, 4 gastroenterologists and 2 pathologists independently graded biopsies from 91 patients with varying stages of disease activity. Moderate inter-observer agreement, as assessed by the kappa statistic, was shown between histologic and endoscopic assessments (κ=0.58), fair agreement between clinical and endoscopic assessments (κ=0.27), and moderate agreement between clinical and histologic assessments (κ=0.47), or among all 3 methods (κ=0.44)(42). The intra- and inter-rater reliability of this index for the evaluation of disease activity in UC was not evaluated.

**Saverymuttu Index**

This index was first described in a prospective trial that compared Indium-111 (111In) granulocyte scanning with endoscopy, histology, and fecal 111In-granulocyte excretion for the assessment of disease extent and severity in 52 patients with Crohn’s disease or UC (45). This widely used histologic scoring system (46-53) generates a total score based on 4 sub-scores (see Table 1). Excellent correlations between endoscopy, histology and 111In scans were shown
(r=0.90 [endoscopy] and r=0.90 [histology] for extent; r=0.86 [endoscopy] and r=0.91 [histology] for disease activity). This index is simple and comprehensive but was not developed using a clearly defined item selection process, did not undergo formal psychometric property testing and was never validated in an independent study.

**Initial Riley Scoring System**

The initial Riley score (Table 2) was described in a randomized, double blind, parallel group trial that compared delayed release mesalamine and enteric coated sulfasalazine maintenance therapy for quiescent UC (defined as endoscopically normal colonic mucosa or the presence of erythema only)(54). Biopsy sections were evaluated using a 5-point scale to measure the degree of chronic inflammatory cell infiltrate and tissue destruction. Relapse rates were not significantly different between treatment groups at the end of the 48-week trial, and Riley scores were low in patients who maintained remission in both treatment groups. This scoring system has not been validated or used in clinical trials.

**Riley and Modified Riley Scoring Systems**

In 1991, Riley et al prospectively examined the value of histologic inflammation to predict clinical recurrence over a 12-month period in 82 asymptomatic UC outpatients in endoscopic remission (17). Unlike the initial Riley score that graded chronic inflammatory cell infiltration and tissue destruction exclusively, this study employed a 4–point scale (none, mild, moderate, or severe) to independently score 6 items: a) presence of an acute
inflammatory cell infiltrate (neutrophils in the lamina propria), b) crypt abscesses, c) mucin depletion, d) surface epithelial integrity, e) chronic inflammatory cell infiltrate (round cells in the lamina propria), and f) crypt architectural irregularities, by two pathologists whose scores were averaged. These additional histologic features were included to better define and isolate features characteristic of mucosal inflammation. The following frequencies of findings were noted: chronic inflammatory infiltrate (100%), crypt architectural irregularities (58%), acute inflammatory activity (32%), acute inflammatory cell infiltrate (28%), crypt abscesses (11%) and mucin depletion (22%). Although 90%–98% inter-observer agreement was described for these items, these estimates were not adjusted for chance and intra-observer agreement was not assessed. Twenty-seven patients (33%) relapsed after a median of 18 weeks (range: 3-44 weeks). Importantly, the presence of residual histologic disease activity was associated with relapse whereas no such relationship was observed for endoscopically defined inflammation. Relapse occurred at similar rates to patients with endoscopically inflamed mucosa in patients who had macroscopically normal mucosa or only erythema at study entry (35% vs. 32%). Relapse rates were higher in the presence of an acute inflammatory infiltrate (52% vs. 25%, $P=0.02$), crypt abscesses (78% vs. 27%, $P<0.0005$), mucin depletion (56% vs. 26%, $P<0.002$) and any breach in the surface epithelium (75% vs. 31%, $P=0.10$). The 4–point Riley score has been adopted as an outcome measure in multiple RCTs (55-61).

The 4–point Riley score was empirically modified by Feagan et al (Table 3) to exclude items such as structural alterations (i.e. crypt branching) that are
probably not responsive to clinically relevant changes in inflammation (62). This modified Riley score (MRS) ranks the degree of inflammation hierarchically, allowing for an un-weighted aggregation of the scores that facilitates the comparison of mean values. This instrument was used as an outcome measure in a RCT of the α4β7 antagonist, vedolizumab (MLN02), for the treatment of active UC (62). The MRS was calculated at baseline and at weeks 4 and 6. Mean histology scores, endoscopically defined disease activity, and symptoms significantly improved at weeks 4 and 6 in patients assigned to vedolizumab. Although these results suggest that the MRS may be a useful outcome measure, the clinical relevance of the changes detected by the MRS remains unknown.

**Geboes Score**

Geboes et al developed a scoring system for microscopic disease activity that incorporated several previously reported histologic items (19). The score was generated with the premise that the major grades and subgrades are progressive and correlate with increasing disease severity or activity (Table 4). To develop this index, 3 pathologists examined a convenience sample of 99 biopsy slides obtained on 2 occasions from the colonic mucosa of patients with actively inflamed (n=68) and quiescent (n=31) distal UC. In this study, kappa values measuring inter-observer agreement between pairs of three readers were 0.20, 0.26 and 0.42. After refining the initially developed scale, kappa values increased to 0.59, 0.62 and 0.70, reflecting better agreement. Inter-observer agreement was higher for samples taken from endoscopically inflamed mucosa (66%) than for those from endoscopically inactive mucosa (64%). Lemmens et al
examined the correlation between endoscopic activity, based on the Mayo score, and histologic activity, based on both the Riley and the Geboes scores in 263 biopsy specimens from 131 patients with UC. A significant correlation was found between endoscopic and histologic activity (Kendall's $\tau = 0.482$, $P < 0.0001$)(63). The GS is the only histologic disease activity score in UC that is comprised of objectively selected items and that has been evaluated for reliability using kappa statistics rather than percent agreement unadjusted for chance.

DISCUSSION:

The Need for Validated Scores

For histologic indices of UC disease severity and activity to be clinically useful, their operating properties must be accurately defined. Validity (extent to which an instrument measures the intended outcome); responsiveness (ability to detect a meaningful change in health status); reliability (consistency or reproducibility of an instrument); and feasibility (ease with which an instrument can be utilized), are essential properties of evaluative instruments. However, based upon the findings of the present review, none of the indices used to evaluate histopathology in clinical trials of UC has been fully validated.

Inter- and intra-observer variability in histologic scoring is a formidable problem. A statistical method that measures agreement between observers beyond chance should be used to assess reliability. The kappa statistic (or kappa coefficient) is most commonly used for this purpose (64). A kappa of 1 indicates perfect agreement, whereas a kappa of 0 indicates agreement by chance (65). A limitation of the kappa statistic is that it is affected by the
prevalence of the finding under observation. Inter-observer agreement can also be calculated by the intra-class coefficient (ICC), which is equivalent to weighted data in the case of ordinal scores (66). Other factors such as the number and quality of samples, and the feasibility of the scoring must also be considered when assessing agreement between readers.

The currently available scoring systems have been applied to biopsy material that was collected under diverse protocols. The original Riley study employed one biopsy sample from the anterior rectal wall, whereas two samples were used in the Geboes study. Histologic disease severity changes during the natural disease course and following administration of effective treatment (16,67). Therefore, while initially one biopsy sample may be reliable, histologic activity may differ among samples collected at follow-up. This difference may be mitigated by collecting two or more samples (68).

Sample quality may be a potential confounder of histologic disease activity as low quality slides may hinder the pathologists’ ability to accurately detect inflammation and, conversely, optimal quality slides may increase their ability to detect subtle changes. In the Geboes study, less than one-third of the samples were considered of good quality. Of 99 total samples, 31 were good, 36 substandard, and 22 were of poor quality (of which, 13 were not examined). This problem has considerable implications for precision of scoring. Some index items, such as basal plasmacytosis, require perpendicular sections of well-oriented samples for accurate interpretation. Section thickness may also affect the interpretation of results. In clinical trials, the operating characteristics of
scores will vary depending on the quality of samples obtained and every effort should be made to standardize and optimize collection and processing of biopsy samples.

Additional research is needed to define which, if any, of the existing histologic indices are most reliable and valid in large clinical trials with heterogeneous sample quality. The results obtained in these circumstances may differ from those from agreement studies performed on samples of optimal quality. Determining the relative merits of numerical versus stepwise scoring is also relevant. In some samples, for instance those obtained from ulcers, granulation tissue may be the major element of the sample, making it difficult if not impossible to evaluate cryptitis or crypt abscesses. Similarly, basal plasmacytosis may be impossible to assess if a section is not correctly orientated.

**Future Directions:**

Validated, reliable, and responsive histologic scoring systems are needed in UC. Validated scoring systems used for both patient management and in clinical trials exist for other diseases, including the non-alcoholic fatty liver disease activity score (69), the METAVIR or Ishak score for chronic hepatitis (70), the Gleason score for prostate cancer (71), and the follicular lymphoma score (72). Validated scoring systems typically are comprised of items that are selected by regression analysis. The operating properties of the scoring system should then be tested by conducting an appropriately powered agreement study with calculation of ICC between multiple central readers followed by assessing
responsiveness of the index to treatments of known efficacy. Several methodological frameworks for the development and validation of evaluative instruments exist (73). Highly responsive indices facilitate early drug development since they are statistically efficient in detecting meaningful treatment differences.

In summary, histopathology is an important component of UC assessment both in clinical practice and for clinical trials, with potential long-term implications for predicting remission rates, future surgery, and malignancy risk (77, 78). Although multiple histologic disease activity scoring systems have been proposed for UC, none is optimal according to established methodological criteria. Further studies are needed to either improve the existing instruments or to develop new measures for use in both clinical practice and in clinical trials.

ACKNOWLEDGEMENT:

We wish to thank Dr. David T. Rubin and Dr. Noam Harpaz for providing us with valuable data regarding the Chicago index and the Harpaz index, respectively.
FIGURE 1. Acute inflammatory changes seen on colonic biopsies of ulcerative colitis patients using hematoxylin and eosin staining showing (A) crypt abscess, (B) cryptitis, and (C) neutrophils in lamina propria. Magnification for A-C: 400x
FIGURE 2. Chronic inflammatory changes seen on colonic biopsies of ulcerative colitis using hematoxylin and eosin staining showing (A) crypt branching and increased eosinophils, (B) inflammatory gap with basal lymphoid aggregates, and (C) inflammatory gap with basal plasma cells. Magnification for A, B: 200x, C: 100x
FIGURE 3. The PRISMA algorithm followed in the systematic review

- Total = 4514 papers
  - Duplicates excluded (2179)
    - Animal Studies Removed (516)
      - Exclusion by title/abstract (1325)
        - Exclusion by applying exclusion criteria (430)
          - Studies added manually (+45)
            - 88 studies involving 18 scoring systems
              - 21 related reviews and studies
TABLE 1. Histologic scoring indices for evaluating disease activity in UC

<table>
<thead>
<tr>
<th>Index</th>
<th>Setting</th>
<th>Description of Scale</th>
<th>Extent of Use</th>
<th>Level of Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Truelove and Richards index (1956)</em>&lt;sup&gt;(15)&lt;/sup&gt;</td>
<td>Prospective study</td>
<td><strong>Stepwise 3-grade scale</strong>&lt;br&gt;1) no inflammation&lt;br&gt;2) mild to moderate inflammation&lt;br&gt;3) severe inflammation</td>
<td>Multiple clinical studies and RCT&lt;sup&gt;(22-41)&lt;/sup&gt;</td>
<td>Partially validated:&lt;br&gt;Agreement between histologic and endoscopic activity was $k=0.58$, between clinical and endoscopic activity was $k=0.27$, between clinical and histologic activity was $k=0.47$ and between all three measures of disease activity was $k=0.44$&lt;sup&gt;(42)&lt;/sup&gt; Unclear item selection process</td>
</tr>
<tr>
<td><em>Matts Score (1961)</em>&lt;sup&gt;(79)&lt;/sup&gt;</td>
<td>Prospective study</td>
<td><strong>Stepwise 5-point grading system</strong>&lt;br&gt;1) normal&lt;br&gt;2) some infiltration of the mucosa or lamina propria, with either round cells or polymorphonuclear cells&lt;br&gt;3) much cellular infiltration of the mucosa, lamina propria and sub-mucosa&lt;br&gt;4) presence of crypt abscesses, with much infiltration of all layers of the mucosa&lt;br&gt;5) ulceration, erosion or necrosis of the mucosa, with cellular infiltration of some or all of its layers</td>
<td>Multiple clinical studies and RCT&lt;sup&gt;(90-83)&lt;/sup&gt;</td>
<td>Not validated</td>
</tr>
<tr>
<td>Score Name</td>
<td>Type</td>
<td>Grading Scale</td>
<td>Validation Status</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------------------------------------</td>
<td>-------------------</td>
<td></td>
</tr>
</tbody>
</table>
| **Watts Score**   | Prospective study | **Stepwise 4-point grading scale**<br>
(0) normal<br>
(1) no significant epithelial changes but increased number of chronic inflammatory cells in the lamina propria<br>
(2) mild epithelial changes, usually regenerative, leukocytes may be seen in the crypts or ducts along with paneth cells. Severe inflammatory cells can be seen in the lamina propria<br>
(3) severe inflammatory changes with evidence of crypt abscesses, inflammatory erosions and frank ulceration, occasional multinucleated giant cells can be seen | One prospective study<sup>84</sup> | Not validated |
| **Keren Score**   | Prospective study | **Descriptive**<br>active versus inactive | None | Not validated |
| **The Friedman Index** | RCT | **Stepwise 4-grade scale**<br>
(0) normal<br>
(1) lamina propria inflammation<br>
(2) crypt injury<br>
(3) ulceration | Multiple clinical studies and RCT<sup>87,94</sup> | Not validated |
| **Gomes Score**   | Prospective study | **Numerical 5-grade scale**<br>
(0) normal<br>
(1) mild edema and inflammation in the lamina propria<br>
(2) crypt abscess formation and lamina propria involvement<br>
(3) destructive crypt abscesses with or without granulomas<br>
(4) active ulceration<br>
**Scores are summed to generate an overall total score (maximum score of 24)** | Clinical studies<sup>96, 97</sup> | Not validated |
| **Saverymuttu Index (1986)**<sup>(45)</sup> | Prospective study | **Numerical grading system generating a total score composed of 4 different variable sub-scores:**

**a) Enterocyte damage:**
(0) Normal, (1) loss of single cells, (2) loss of groups of cells, (3) frank ulcerations

**b) Crypt abnormalities:**
(0) Normal, (1) single inflammatory cell, (2) cryptitis, (3) crypt abscesses

**c) Lamina propria involvement:**
(0) Normal, (1) slight increase in mononuclear cells, (2) moderate increase in mononuclear cells, (3) marked increase in mononuclear cells

**d) Acute inflammatory infiltration in the lamina propria:**
(0) normal, (1) mild increase, (2) moderate increase, (3) marked increase |

Multiple clinical studies and RCT<sup>(46-53)</sup> | Not validated |

| **Floren Index (1987)**<sup>(98)</sup> | Prospective study | **Stepwise 5-point grading system**

(1) normal
(2) enhanced glands with intraepithelial granulocytes and stromal enhancement beyond normal of lympho-plasmacytic cells or eosinophilic granulocytes (slight inflammation)
(3) goblet cell depletion, loss of tubular parallelism, and reduced mucin production in some glands with intraepithelial granulocytes, marked increase of inflammatory cells in the stroma (intermediate inflammation)
(4) marked gland and mucosal atrophy, evident crypt abscesses and pus on the surface, massive increase of acute inflammatory cells and follicle formation |

Multiple clinical studies and RCT<sup>(98-106)</sup> | Not validated |
<table>
<thead>
<tr>
<th>Index</th>
<th>Study Type</th>
<th>Grade System</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanauer Index (1993)(^{107})</td>
<td>RCT</td>
<td><strong>Stepwise 4-point grading system</strong></td>
<td>Few clinical studies and RCT(^{108-110}) Not validated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0) normal colonic mucosa, to (3) high-grade active inflammatory bowel disease</td>
<td></td>
</tr>
<tr>
<td>Odze Index (1993)(^{68})</td>
<td>Prospective study</td>
<td><strong>Stepwise 4-point grading system</strong></td>
<td>None Not validated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(a) normal (b) active (ulcerations)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) chronically active (cryptitis, crypt abscesses, or surface erosions)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(d) chronically inactive (not defined)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Categorized based on 6 parameters:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1) crypt architecture, (2) villiform surface contour, (3) mixed inflammation in the lamina propria, (4) basal plasmacytosis, (5) basally located lymphoid aggregates, (6) paneth cell metaplasia</td>
<td></td>
</tr>
<tr>
<td>Sandborn Index (1993)(^{111})</td>
<td>Prospective study</td>
<td><strong>Stepwise 4-point grading system</strong></td>
<td>Few clinical studies and RCT Not validated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0) inactive chronic colitis (normo-cellular or hyper-cellular lamina propria; PMNs absent)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1) mild active chronic colitis (scant PMN’s in lamina propria, occasional cryptitis but few crypt abscesses, minimal glandular destruction or ulceration)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) moderately active chronic colitis (moderate numbers of PMN’s in lamina</td>
<td></td>
</tr>
</tbody>
</table>
propria with cryptitis and crypt abscesses prominent, some glandular destruction)
(3) severely active chronic colitis (numerous PMN's with abundant cryptitis, crypt abscesses, extensive glandular destruction; ulceration may be prominent)

**Scheppach/ D’Argenio Score (2001)**

**Prospective study**

**Numerical scoring system**
Measures six variables (infiltration of the mucosa by lymphocytes and plasma cells, infiltration of crypts by neutrophils, crypt abscesses, ulceration, mucous cell depletion and crypt architectural distortion) each scored on a 0 to 2 scale according to severity.

**Maximum overall score of 12**

**Harpaz Index (2003)**

**Prospective study**

**Stepwise grading system**
Categorizes microscopic disease activity into:

- a) inactive (no cryptitis)
- b) mildly active colitis (cryptitis in <50% of crypts)
- c) moderately active colitis (cryptitis in >50% of crypts)
- d) severely active colitis (ulcerations or erosions)

**Two clinical studies**

**Partially validated:** Harpaz index was highly correlated with the Mayo clinic endoscopic sub-score ($r_s=0.76$)

**Unclear item selection process**

**Initial Riley Score (1988)**

**RCT**

**Stepwise 5-point scale**
Measures the degree of chronic inflammatory cell infiltrate and tissue destruction

**None**

**Not validated**

**Riley Score (1991)**

**Prospective study**

**Stepwise 4-point scale**
None (0), mild (1), moderate (2), or

**Multiple clinical studies and**

**Partially validated:** Reliability measured

**Not validated**
severe (3) Independently scores 6 different features:

- a) presence of an acute inflammatory cell infiltrate (neutrophils in the lamina propria)
- b) crypt abscesses
- c) mucin depletion
- d) surface epithelial integrity
- e) chronic inflammatory cell infiltrate (round cells in the lamina propria)
- f) crypt architectural irregularities

Modified Riley score (2005)<sup>(62)</sup>  RCT  Stepwise grading system

| Consists of select items from the original score that were believed to be responsive to changes in acute inflammation |

Geboes Score (2000)<sup>(19)</sup>  Prospective study  Stepwise grading system

<table>
<thead>
<tr>
<th>Categorizes patients into 7 grades based on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) architectural changes</td>
</tr>
<tr>
<td>2) chronic inflammatory infiltrate</td>
</tr>
<tr>
<td>3) lamina propria eosinophils</td>
</tr>
<tr>
<td>4) lamina propria neutrophils</td>
</tr>
<tr>
<td>5) epithelial neutrophils</td>
</tr>
<tr>
<td>6) crypt destruction</td>
</tr>
<tr>
<td>7) erosions or ulcerations</td>
</tr>
</tbody>
</table>

RCT<sup>(55-61)</sup> by percent agreement (94%) but unadjusted for chance

None  Not validated

**Partially validated:**
Inter-observer agreement was assessed among three readers. Kappas=0.26, 0.42 and 0.20<sup>19</sup>

Few clinical studies<sup>(58, 119, 120)</sup>
<table>
<thead>
<tr>
<th>&quot;Chicago&quot; score (2007)</th>
<th>Case-control study</th>
<th><strong>Stepwise 6-point scale</strong></th>
<th>One clinical study</th>
<th>Not validated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0) Normal (completely uninvolved, no architectural distortion, no infiltrates)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1) Quiescent (architectural distortion, increased lamina propria lymphocytes, but no activity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) Increased lamina propria granulocytes without definite intraepithelial granulocytes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) Intraepithelial granulocytes (e.g. cryptitis) without crypt abscesses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4) Crypt abscesses in less than 50% of crypts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5) Crypt abscesses in greater than 50% of crypts, or erosion/ulcerations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**TABLE 2. Initial Riley Scoring System**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Microscopic Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>1</td>
<td>Mild increase in chronic inflammatory cell infiltrate, no tissue destruction</td>
</tr>
<tr>
<td>2</td>
<td>Moderate increase in chronic inflammatory cell infiltrate, no tissue destruction</td>
</tr>
<tr>
<td>3</td>
<td>Marked increase in chronic inflammatory cell infiltrate, mild tissue destruction</td>
</tr>
<tr>
<td>4</td>
<td>Marked increase in chronic inflammatory cell infiltrate, obvious tissue destruction</td>
</tr>
</tbody>
</table>
### TABLE 3. Modified Riley Scoring System

<table>
<thead>
<tr>
<th>Activity</th>
<th>Histological Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0: Neutrophils in epithelium = None (3.0)</td>
</tr>
<tr>
<td></td>
<td>1: Neutrophils in epithelium = &lt;25% crypts involved (3.1 or 3.2A)</td>
</tr>
<tr>
<td>Mild</td>
<td>2: Neutrophils in epithelium = ≥25% to ≤75% crypts involved (3.2B or 3.3A)</td>
</tr>
<tr>
<td></td>
<td>3: Neutrophils in epithelium = &gt;75% crypts involved (3.3B)</td>
</tr>
<tr>
<td></td>
<td>4: Lamina propria neutrophils = Mild but unequivocal increase (2B.1)</td>
</tr>
<tr>
<td>Moderate</td>
<td>5: Lamina propria neutrophils = Moderate increase (2B.2)</td>
</tr>
<tr>
<td></td>
<td>6: Lamina propria neutrophils = Marked increase (2B.3)</td>
</tr>
<tr>
<td>Severe</td>
<td>7: Erosion or ulceration = Present (5.1 or 5.2 or 5.3 or 5.4)</td>
</tr>
</tbody>
</table>
### TABLE 4. The Geboes Scoring System

<table>
<thead>
<tr>
<th>Grade 0 - Structural (architectural change)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subscore</strong></td>
<td><strong>Characteristics</strong></td>
</tr>
<tr>
<td>0.0</td>
<td>No abnormality</td>
</tr>
<tr>
<td>0.1</td>
<td>Mild abnormality</td>
</tr>
<tr>
<td>0.2</td>
<td>Mild or moderate diffuse or multifocal abnormalities</td>
</tr>
<tr>
<td>0.3</td>
<td>Severe diffuse or multifocal abnormalities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 1 - Chronic inflammatory infiltrate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>No increase</td>
</tr>
<tr>
<td>1.1</td>
<td>Mild but unequivocal increase</td>
</tr>
<tr>
<td>1.2</td>
<td>Moderate increase</td>
</tr>
<tr>
<td>1.3</td>
<td>Marked increase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 2A - Lamina propria eosinophils</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2A.0</td>
<td>No increase</td>
</tr>
<tr>
<td>2A.1</td>
<td>Mild but unequivocal increase</td>
</tr>
<tr>
<td>2A.2</td>
<td>Moderate increase</td>
</tr>
<tr>
<td>2A.3</td>
<td>Marked increase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 2B - Lamina propria neutrophils</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2B.0</td>
<td>No increase</td>
</tr>
<tr>
<td>2B.1</td>
<td>Mild but unequivocal increase</td>
</tr>
<tr>
<td>2B.2</td>
<td>Moderate increase</td>
</tr>
<tr>
<td>2B.3</td>
<td>Marked increase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 3 - Neutrophils in epithelium</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>None</td>
</tr>
<tr>
<td>3.1</td>
<td>&lt;5% crypts involved</td>
</tr>
<tr>
<td>3.2</td>
<td>&lt;50% crypts involved</td>
</tr>
<tr>
<td>Grade 4 - Crypt destruction</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>4.0</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>4.1</strong></td>
<td>Probable - local excess of neutrophils in part of crypt</td>
</tr>
<tr>
<td><strong>4.2</strong></td>
<td>Probable - marked attenuation</td>
</tr>
<tr>
<td><strong>4.3</strong></td>
<td>Unequivocal crypt destruction</td>
</tr>
</tbody>
</table>

**Grade 5 - Erosion or ulceration**

| **5.0** | No erosion, ulceration, or granulation tissue |
| **5.1** | Recovering epithelium + adjacent inflammation |
| **5.2** | Probable erosion - focally stripped |
| **5.3** | Unequivocal erosion |
| **5.4** | Ulcer or granulation tissue |
REFERENCES:


36. Kruis W, Schreiber S, Theuer D, et al. Low dose balsalazide (1.5 g twice daily) and mesalazine (0.5 g three times daily) maintained remission of ulcerative colitis but high dose balsalazide (3.0 g twice daily) was superior in preventing relapses. Gut. 2001;49:783-789.


63. Lemmens B, Arijs I, Van Assche G, et al. Correlation between the endoscopic and histologic score in assessing the activity of ulcerative


72. Matutes E, Owusu-Ankomah K, Morilla R, et al. The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of


87. D'Arienzo A, Manguso F, Castiglione GN, et al. Beclomethasone dipropionate (3 mg) enemas combined with oral 5-ASA (2.4 g) in the


94. Roseth AG, Aadland E, Grzyb K. Normalization of faecal calprotectin: a
predictor of mucosal healing in patients with inflammatory bowel disease.


108. Auwerda JJA, Zijlstra FJ, Tak CJAM, et al. Ridogrel enemas in distal


pathologists with different levels of experience. Mod Pathol. 2003;16:118A.


Preface to Chapter 2:

This chapter presents a study that evaluated agreement among pathologists in the evaluation of histologic disease activity in UC. The GS and the MRS, the two evaluative indices developed through a clearly defined item selection process that included psychometric property testing, were assessed using a central information management system (CIMS) that allowed multiple expert pathologists to score digitized histologic images that were obtained during a clinical trial of induction therapy for UC. Intra- and inter-observer agreement was estimated and sources of disagreement between readers were identified. Recommendations for standardizing scoring of readers were also developed.

This study is a component of a long-term plan to develop and validate histopathology as an outcome measure in both clinical trials and clinical practice.
Chapter 2: Agreement Among Pathologists in the Assessment of Histologic Disease Activity in Ulcerative Colitis.

ORIGINAL ARTICLE

Agreement Among Pathologists in the Assessment of Histologic Disease Activity in Ulcerative Colitis

Correspondence to Dr. Barrett G Levesque; Division of Gastroenterology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA USA 92093-0956, 92103; bglevesque@ucsd.edu; phone (858) 657-7060; fax (858) 657-5022

Authors: Mahmoud H Mosli,1,2,3,4 Brian G Feagan,1,2,5 William J Sandborn,1,6 Geert D’Haens,1,7 Reena Khanna1,2, Cynthia Behling,8 Keith Kaplan,9 David K Driman,10 Lisa M Shackelton,1 Kenneth A Baker,1 John K MacDonald,1 Margaret K Vandervoort,1 Karel Geboes,11 Mark A Valasek12, Rish Pai13, Cord Langner14, Robert Riddell15, Noam Harpaz16, Maida Sewitch17, Michael Peterson18, Larry W Stitt5, GuangYong Zou5, Barrett G Levesque 1,6

1 Robarts Clinical Trials, Robarts Research Institute, Western University, London, Ontario, Canada
2 Department of Medicine, Western University, London, Ontario, Canada
3 Department of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia
Department of Epidemiology, Biostatistics and Occupational Medicine, McGill University, Montreal, Canada.

Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada

Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA

Department of Gastroenterology, Academic Medical Center, Amsterdam, The Netherlands

Pacific Rim Pathology Medical Corporation, San Diego, CA, USA

Carolinhas Medical Center, Charlotte, NC, USA

Department of Pathology, Western University, London, Ontario, Canada

Department of Pathology, University Hospital of KU Leuven and UZ Gent, Leuven, Belgium

Department of Pathology, University of California, San Diego, USA

Department of Anatomic Pathology, Cleveland Clinic, Cleveland, OH, USA

Department of Pathology, Medical University of Graz, Graz, Austria

Department of Pathology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

Department of Pathology, Mount Sinai Hospital, New York, NY, USA

Department of Medicine, McGill University, Montreal, Canada

Department of Pathology, Western Washington Pathology and Multicare Health System, Tacoma, Washington, USA
ABSTRACT:

Objective: Histopathology is potentially an important outcome measure in ulcerative colitis (UC). Many histologic disease activity indices have been developed, but their operating properties are not well defined. In the present study, we assessed intra- and inter-rater agreement for the Geboes and Modified Riley scores.

Design: An agreement study with repeated assessments at baseline, 2 and 4 weeks was undertaken. Five experienced pathologists with gastrointestinal pathology fellowship training and expertise in inflammatory bowel disease (IBD) evaluated 49 UC colon biopsies using the Geboes score (GS), modified Riley score (MRS) and a 100 mm visual analogue scale (VAS) global rating of histologic severity. Intra- and inter-rater agreement for each grading system and for individual instrument items were determined using intra-class correlation coefficients (ICC). Correlation between VAS, GS and MRS was calculated. Ad hoc, a consensus process was used to understand the common sources of measurement disagreement among the biopsies responsible for the greatest disagreement. Rules for minimizing disagreement were subsequently generated by consensus.

Results: Intra-rater ICCs (95% confidence intervals) for the total GS, MRS and VAS scores were 0.82 (0.77, 0.87), 0.71 (0.64, 0.78) and 0.79 (0.73, 0.85), respectively. Corresponding inter-rater ICCs were substantially lower: 0.56 (0.45, 0.68), 0.48 (0.38, 0.61) and 0.61 (0.51, 0.72). Good correlation was observed between VAS scores and the Geboes and modified Riley scores.
Consensus participants identified lack of standardised definitions and sub-optimal slide quality as potential causes of disagreement.

**Conclusions:** Although substantial to “almost perfect” intra-rater agreement was found in the assessment of histologic disease activity in UC, inter-rater agreement was considerably lower. Standardization of item definitions and modification of the existing indices is required to create a psychometrically sound UC histological disease activity index.

**Key words:** ulcerative colitis, histology, central reading, disease activity
**What is already known about this subject?**

- Endoscopic bowel healing is associated with favourable long-term outcomes, but does not assure histologically inactive disease.
- Histologic inflammation is associated with an increased risk of relapse.
- Quantitative histologic indices are preferable to subjective assessment of symptoms as outcome measures in clinical trials.
- Although multiple histologic scoring systems have been developed and used in clinical trials, their operating properties have not been systematically validated.
- Identification, and standardization of the histologic features of UC disease activity may help define both treatment goals in clinical practice and outcome measures for clinical trials.

**What are the new findings?**

- Intra-rater agreement for histologic disease activity using Geboes score (GS), modified Riley score (MRS) and 100 mm visual analogue scale (VAS) in UC is “substantial” to “almost perfect” when pathologists centrally score biopsies.
- Inter-rater agreement is considerably lower than intra-rater agreement.
- Higher agreement was observed when GS was assessed as a continuous variable.
- There is good correlation between global VAS of histologic disease activity with the Geboes and modified Riley scores.
- A consensus process identified causes of disagreement and potential solutions for improving inter-rater agreement.
<table>
<thead>
<tr>
<th>How might it impact on clinical practice in the foreseeable future?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Validated histologic scoring systems are needed in UC.</td>
</tr>
<tr>
<td>• Partially validated indices, such as GS or MRS can provide preliminary indicators of disease activity and/or progression until fully validated instruments are available.</td>
</tr>
<tr>
<td>• Standardization of definitions and measures of slide quality may improve future use of currently available indices.</td>
</tr>
</tbody>
</table>
INTRODUCTION:

Clinical symptoms and endoscopy findings are traditionally used to evaluate disease activity in ulcerative colitis (UC). However, there is growing interest in the assessment of histologic disease activity based on the concept that resolution of bowel inflammation beyond endoscopic healing may provide additional clinical benefit. For example, in a prospective study of 82 patients with quiescent UC, the presence of residual histologic inflammation on rectal biopsy was associated with a three-fold greater risk of relapse at 12 months of follow-up (1). However, before a measure of histologic disease activity can be accepted as a useful metric in clinical research and clinical practice, the measures’ reliability and reproducibility must be demonstrated (2).

The two most commonly used histologic indices to evaluate disease activity in UC are the Geboes Score (GS) and the Riley Score. The GS is a 7-item instrument that has been utilized as an outcome measure in clinical trials. It classifies histologic changes as Grade 0 (structural change only); Grade 1 (chronic inflammation); Grade 2 (2a: lamina propria eosinophils, 2b: lamina propria neutrophils); Grade 3 (neutrophils in the epithelium); Grade 4 (crypt destruction); and Grade 5 (erosions or ulcers) and generates a score from 0 to 5.4, with higher scores indicating greater inflammation. A decrease of the GS to Grade 0 or 1 has been empirically designated as histologic healing (3). The Riley score is a 6-item instrument that grades each item on a four-point scale (none, mild, moderate, or severe). Scores range from 0 (no inflammation) to 7 (severe acute inflammation)(1, 4). The modified Riley score (MRS) excluded the
item ‘architectural distortion’ (1), based upon the premise that it is unlikely to be responsive to change following therapy (4) and is considered an improved version of the Riley score. Both the Geboes and the Riley scores were developed through an item selection process that partially included psychometric property testing. In contrast, both the Chicago (5) and Harpaz (6) scores, are histologic indices of UC disease activity that were developed through an unclear item selection process. Global severity of histologic disease activity is generally assessed using a 100 mm visual analog scale (VAS), where no disease activity was scored as 0 and the most severe activity was scored as 100.

Neither the GS nor the MRS was developed using an optimal structured framework for index development, which is why their operating properties are essentially unknown (7). Further characterization of these indices therefore requires complete assessment of validity, responsiveness and reliability. In this paper, we assessed intra- and inter-rater agreement for the GS and the MRS and for individual items that comprise these indices in addition to several other items that were selected by expert pathologists and identified the items of highest disagreement. Furthermore, correlations between the VAS, the Geboes and the modified Riley scores were also assessed as measures of validity.

METHODS:

Study design and population:

A multi-center agreement study with assessments at baseline, 2 and 4 weeks was undertaken. Five pathologists with gastrointestinal pathology fellowship training and experience in inflammatory bowel disease (IBD) (KG, CB,
KK, RP and CL) participated; they were selected for their expertise and willingness to commit to the initiative, and then trained in the use of a central image management system (CIMS) that hosted the digital histologic images. Standardized training materials were provided on the GS and MRS that included examples of ideal digital images for each individual index item. Points of disagreement between readers regarding the definitions of items were discussed and consensus was reached prior to study initiation. Each pathologist independently reviewed 50 randomly selected digital slide images three times, approximately two weeks apart. All of the individual items that comprise the GS and MRS, and other items considered potentially relevant by the participating pathologists were evaluated at each reading. The pathologists also rated global histologic disease activity using a 100 mm VAS. Digital images were reviewed in the absence of clinical information. Each pathologist provided subjective assessments of the overall quality of each slide that were based on three criteria (stain, section, and image quality).

Following completion of the initial reading process, the sources of disagreement among readers were evaluated using a two-step procedure. First, outlying images were statistically identified. Second, a consensus process was performed ad-hoc in which five additional specialized pathologists (NH, RR, DD, MV and MP) were invited to join the initial central reading pathologists to reassess digital images with the greatest amount of disagreement. Each pathologist participated in this multi-phase consensus process to identify the reasons for disagreement. In Phase 1, each pathologist individually examined a
sample of slides that were identified statistically to be responsible for the majority of disagreement, and provided feedback on their overall quality and scored them using the GS and MRS study indices. Phase 2 was an open teleconference where feedback was compiled and presented by the conference moderator (BL). Slides and the concerns raised during Phase 1 were presented and then each concern and potential explanations and solutions were discussed among the group. Consensus was reached regarding the score of each slide according to the GS and the MRS and the level of quality.

Rectal biopsies used in the present study had been obtained from patients with mild to moderately active UC, who participated in a phase 2 randomized controlled trial (RCT) of vedolizumab, a monoclonal antibody directed to the alpha-4-beta-7 integrin (4). In the trial, disease activity was measured according to clinical and endoscopic disease activity scales, the ulcerative colitis clinical score (UCCS) and the modified Baron score (MBS), respectively (8, 9). At the time of the RCT, biopsies were paraffin embedded, sectioned and stained with hematoxylin and eosin (H&E). The slides were scanned at 40X magnification on a Scanscope CS (Aperio, Vista, CA) slide scanner, and the digitized images hosted for viewing on a secure, regulatory-compliant website using proprietary ImageScope (Aperio, Vista, CA) software. In the present study, we elected to evaluate only patients from the control group of the trial since the specific effect of vedolizumab, a highly selective inhibitor of gut lymphocyte trafficking, on inflammatory responses in the mucosa, is currently unknown.
Statistical analyses and sample size considerations:

Descriptive statistics were used to assess the patient characteristics. Intra-rater agreement was defined as the correlation between two measurements on the same biopsy image made by the same pathologist at two different time points. Inter-rater agreement was defined as the correlation between two measurements on the same biopsy image made by two different pathologists. Intra- and inter-rater agreement was estimated for each histologic index using a two-way random-effects model incorporating interaction between images and raters. The resulting intra- and inter-rater class correlation coefficients (ICC) is equivalent to the weighted kappa coefficient in the setting of multiple raters in clinical trials of ordinal scales and can be applied to binary (qualitative) data as well (10-15).

Two-sided 95% confidence intervals (CIs) were obtained using methods discussed by Gilder et al (16) for balanced models and extended by Ye et al (17) for unbalanced models. The strength of agreement was evaluated according to the criteria of Landis and Koch whereby ICCs of <0.00, 0.00-0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80, 0.81-1.00 indicate “poor”, “slight”, “fair”, “moderate”, “substantial”, and, “almost perfect” agreement, respectively (18). The same approach was used for sub-group analysis of the images where no quality issues were identified. Outlying images were statistically identified using case-deletion diagnostics for mixed models (19).

As ICCs are defined by variance components, images that had the largest impact on estimates of variance components were identified as outliers in the
estimation of ICCs. Outlier images were used as examples in the consensus process. Correlations between VAS and GS and MRS were measured using Pearson’s correlation coefficient (r) with 95% CI.

We evaluated intra- and inter-rater agreement on the GS, MRS and VAS using a design in which our 5 pathologists made three independent measurements of 50 biopsy images. This sample size was sufficient without consideration of the triplicate images (20). Specifically, assuming a true ICC of 0.75, the study would have had an 83% chance of obtaining a one-sided 95% lower confidence bound for the ICC of 0.6, the “substantial” agreement criterion.

Prior to initiation of the RCT, the Research Ethics Board at the University of Western Ontario granted ethics approval in concordance with internationally accepted ethical guidelines. The biopsy slides analyzed in the present study were obtained from a clinical trial that complied with all applicable regulatory requirement(s)$$. The consent of study participants included the use of the collected data for other research purposes, and thus additional consent for the present study was not obtained. All biopsy data were de-identified and the pathologists were blinded to clinical information. All pathologist participants provided written informed consent.

__________________________

$ The ethics approval document is presented in the appendix section.
RESULTS:

Study population:

Table 1 shows the demographic and clinical characteristics of the patients who provided the histologic samples (1 sample per patient). One of the 50 selected digital images was found to be unsuitable for examination; therefore, the analysis is based on the remaining 49 images. The patient characteristics were generally representative of participants in induction trials of treatment for mild to moderately active UC (21). As shown in Figure 1, study patients had mild to moderate clinical and endoscopic disease activity according to the UCCS (scale range: 1-12) and MBS (scale range: 0-4), and the full spectrum of histologic disease activity according to the VAS (scale range 0-100).

Intra- and inter-rater agreement:

Intra-rater ICCs (95% CI) for total GS (grades 0 to 5), MRS and VAS scores were 0.82 (0.77, 0.87), 0.71 (0.64, 0.78) and 0.79 (0.73, 0.85), respectively, indicating “substantial” to “almost perfect” agreement. Inter-rater ICCs (95% CI) for the total GS (as grades 0 to 5), MRS and VAS scores were 0.56 (0.45, 0.68), 0.48 (0.38, 0.61) and 0.61 (0.51, 0.72) indicating “moderate” to “substantial” agreement (Table 2). When agreement for GS was measured on a collapsed categorical scale between 1 and 3 (category 1 = grades 0 or 1 “inactive or mildly active”, category 2 = grades 2 or 3 “moderately active” and category 3 = grades 4 or 5 “severely active with epithelial involvement”), intra-rater ICC was 0.766 (0.708, 0.827) and inter-rater ICC was 0.508 (0.397, 0.629) indicating “substantial” and “moderate” intra- and inter-rater agreement, respectively.
Alternatively, when agreement for GS was measured on a continuous scale from 0 to 22, intra-rater ICC was 0.842 (0.796, 0.886) and inter-rater ICC was 0.602 (0.495, 0.711) indicating “almost perfect” and “substantial” intra- and inter-rater agreement, respectively. The results of analyses of the individual index items follow.

**Architectural features**

Individual item ICCs for the assessment of architectural features are summarized in Table 3. The highest intra-rater agreement was observed for crypt architectural distortion: 0.85 (0.81, 0.89) and the lowest intra-rater agreement was patchiness: 0.48 (0.40, 0.58). The highest inter-rater agreement was observed for crypt architectural irregularities according to the MRS criteria: 0.72 (0.63, 0.80) and the lowest inter-rater agreement was with patchiness: 0.20 (0.11, 0.31).

**Acute inflammation**

Individual item ICCs for the assessment of acute inflammation are summarized in Table 4. The highest intra-rater agreement was observed for changes according to the Chicago index: 0.76 (0.70, 0.82) and the lowest intra-rater agreement was observed for assessment of crypt abscesses: 0.55 (0.47, 0.64). The highest inter-rater agreement was observed for detection of neutrophils: 0.52 (0.41, 0.64) while lowest inter-rater agreement was observed for the presence of lamina propria eosinophils according to the GS criteria: 0.26 (0.18, 0.37).
**Chronic inflammation**

Individual item ICCs for the assessment of chronic inflammation are summarized in Table 5. The highest intra-rater agreement was observed for detection of a chronic inflammatory infiltrate according to the GS: 0.81 (0.75, 0.86) whereas the lowest intra-rater agreement was observed for assessment of basal plasmacytosis: 0.78 (0.71, 0.84). The highest inter-rater agreement was observed for detection of a chronic inflammatory infiltrate according to both GS and MRS: 0.64 (0.54, 0.74) and the lowest inter-rater agreement was observed for assessment of basal plasmacytosis: 0.63 (0.53, 0.73).

**Epithelial injury**

Individual item ICCs for the assessment of epithelial injury are summarized in Table 6. The highest intra-rater agreement was also observed for detection of erosions or ulcerations according to the GS: 0.78 (0.73, 0.84), whereas the lowest intra-rater agreement was observed for identification of granulomas: 0.49 (0.41, 0.57). The highest inter-rater agreement was observed for detection of erosions or ulcerations according to the GS: 0.56 (0.45, 0.67) and surface epithelial integrity according to MRS: 0.56 (0.45, 0.67). The lowest inter-rater agreement was observed for identification of granulomas: 0.01 (0.00, 0.08).
**Correlation of Histologic Indices with Global VAS**

The intent of these analyses was to assess the validity of scores against a global measure of disease activity. The correlation between the VAS and MRS was $r=0.624$ (95% CI: 0.545, 0.688). For GS, correlation was measured in three different ways: on a 6-grade ordinal scale was $r=0.61$ (95% CI: 0.50, 0.67), on a continuum score scale (total cumulative score of 22) was $r=0.66$ (95% CI: 0.57, 0.72) and on a categorical scale between 0 and 3 (inactive as grades 0 or 1, mildly active as grades 2 or 3 and severely active with epithelial involvement as grades 4 or 5) was $r=0.58$ (95% CI: 0.48, 0.64).

**Disagreement and the Consensus process:**

Seventeen (34.7%) of the 49 biopsy images accounted for the majority of the disagreement based on influential statistics in mixed models performed to identify outliers. The most common reasons for disagreement based on the findings of the consensus process were differences in the interpretations of index item definitions including: artefact, granulation tissue, crypt destruction, crypt distortion, basal plasmacytosis, lamina propria neutrophils, and in the approaches to scoring slides of suboptimal and poor quality. Methods to standardize the interpretation of the items with the greatest disagreement were developed during the consensus process and are summarized in Table 7.

**Quality assessment:**

The pathologists identified 213 of 734 (29%) digital slides as being of “sub-optimal” quality. The leading explanations for sub-optimal image quality were: over staining (16%), inadequate sampling (8%), poor orientation (1%),
inability to focus adequately (1%) and other (5%). A total of 74 slides (10%) were deemed of poor quality and excluded in a subgroup analysis of intra- and inter-rater agreement.

**Results of sub-group analysis excluding poor quality images:**

Analysis excluding the digital slides rated as poor quality showed intra-rater ICCs of 0.85 (0.80, 0.89), 0.72 (0.65, 0.79) and 0.81 (0.76, 0.87) and inter-rater ICCs (95% CI) of 0.58 (0.47, 0.69), 0.49 (0.38, 0.61) and 0.60 (0.50, 0.71) and for GS, MRS and VAS, respectively. Individual ICC’s for items that assess architecture, acute inflammation, chronic inflammation, and epithelial injury are summarized in Table 8.

**DISCUSSION:**

In this large-scale evaluation of histologic indices of disease activity in UC we demonstrated “substantial” to “almost perfect” intra-rater agreement for both total GS and MRS scores and, with a few exceptions, for the individual items that constitute these instruments. GS showed improved agreement when considered as a continuous outcome measure where a cumulative score between 0 and 22 is generated based on individual item scoring. In contrast, inter-rater agreement was considerably lower for the GS, MRS and VAS. These differences in intra- and inter-rater agreement are expected, because observers are more likely to agree with themselves than with each other. However, the differences were greater than those observed in two recent studies of identical design that evaluated central endoscopic scoring in UC and CD (22, 23). We hypothesize that this discordance is due to variations in reader interpretation of item
definitions. Nevertheless, if an index lacks reliability then its ability to reproduce the same results is poor and it cannot be used as a validated outcome measure (24).

High inter-rater agreement was observed for items assessing acute inflammation including superficially located neutrophils compared to neutrophils in the lamina propria using both GS and MRS. Inter-rater agreement for the assessment of neutrophils in the lamina propria might be improved by standardizing the quality of sectioning and staining of the biopsies. Acute inflammation can be assessed by the presence or absence of neutrophils scattered in the epithelium or in the colonic crypts as cryptitis or crypt abscesses. Features that suggest chronic inflammation, such as basal plasmacytosis, may have significant prognostic implications (25) and be important to detect. All features of chronic inflammation showed high intra- and inter-rater ICCs according to GS and MRS. Epithelial injury may be considered a marker of severity but may also be confused with artefact that occurs when the endoscopic biopsy sample is acquired or when the slide is prepared. Our results showed some variation in agreement between items in this category. Crypt distortion according to GS and MRS demonstrated high agreement between central pathologists, and central histologic assessment can therefore be considered reliable for diagnosing IBD. Architectural distortion is a major criterion used to diagnose IBD but is not a marker of activity. As endpoints for clinical trials and clinical practice, acute and chronic inflammatory changes are more relevant than architectural changes and contribute to decision-making regarding treatment and
prognosis. A possible study limitation is that we only evaluated biopsies from the recto-sigmoid. Although we know of no data to address this issue, it is possible that regional differences in colonic histopathology might exist that would preclude generalizing of our results to biopsies taken from other locations.

Acquisition and preparation of biopsy samples and tissue sections are technically difficult processes that have potential for artefact and variation in quality. However, based on a subgroup analysis limited to slides with optimal quality, lower quality slides (10% of the total sample) did not seem to significantly contribute to the observed disagreement among pathologists. Because assessment of histologic features is interpretive rather than quantitative inherent variation in assessment exists. These concerns were expressed by the pathologists involved in this study through a systematic survey process, which led to generation of a new set of more detailed rules for future central histologic evaluation of UC. Future studies that incorporate standardized definitions, questionnaires and improved slide quality may help to objectively identify the best scoring system for use as an endpoint in clinical trials. Whether this instrument is a revised version of an existing index or a newly developed index that fulfills the criteria of validity requires further assessment.

Many of the individual items evaluated in this agreement study showed high intra- and inter-rater agreement. These included: crypt architectural distortion (MRS), detection of neutrophils (Chicago index), chronic inflammatory changes (GS), detection of chronic inflammatory changes (GS and MRS), erosions or ulcerations (GS), surface epithelial integrity (MRS) and may be the
best candidates for inclusion in a new disease activity index providing that they are responsive to change. Items that were not reliable (patchiness, assessment of surface neutrophils, detection of lamina propria eosinophils (GS), basal plasmacytosis, identifying granulomas) may continue to be problematic unless they can be improved through re-definition, training and increased sample quality. Basal plasmacytosis, in particular, requires properly oriented biopsy sections. The consensus process provided us with potential approaches to improve inter-rater agreement including standardizing definitions of microscopic changes, quality of slides and approach to slide assessment. An evaluative index that lacks reliability will not be a valid outcome measure (24). Therefore, identification of a reliable and responsive histologic disease activity index suitable for use, as a centrally read histologic endpoint in clinical trials of drug development for UC remains a research priority. This index would provide an objective measurement of response to therapy that directly measures underlying inflammation and which might be predictive of long-term clinical outcomes. We speculate that such an index could also potentially be useful in clinical practice. The strong correlation observed between VAS with GS and MRS suggests that these two indices are potentially valid and can be optimized in the future.

In conclusion, we found “substantial” to “almost perfect” intra-rater agreement among pathologists in the assessment of disease activity in UC using the GS and MRS, but only “moderate” inter-rater agreement. These findings suggest that pathologists are highly reliable in their own assessments of UC histologic disease activity using existing indices, but not as reliable when
assessments are compared across pathologists. To be useful as an outcome measure in clinical trials, a UC histologic index must be reliable and responsive; these properties are being evaluated in continuing studies that will ultimately define the optimal histologic index. Results of the consensus process helped us to characterize the most important sources of disagreement and generate rules that may potentially be used to improve inter-rater agreement. These rules may, in turn, be used for revising the existing instruments or creating a new instrument.

COMPETING INTERESTS:

The following authors are employees or are affiliated with Robarts Clinical Trials, Inc.: MHM, BGF, WJS, GD’H, RK, CB, KK, LMS, KAB, JKM, MKV, KG, MAV, RP, CL, MP, LWS, GYZ, BGL. The following authors have no competing interests: DKD, RR, NH, MS.
Figure 1. The following histograms illustrate the range of disease severity of the study population as assessed by a) histologic activity (VAS), b) endoscopic activity (MBS) and c) clinical activity (UCCS)

a)

![Histogram a) histologic activity (VAS)](image1)

b)

![Histogram b) endoscopic activity (MBS)](image2)
c)

* VAS, visual analogue scale; UCCS, ulcerative colitis clinical score; MBS, modified Baron score
Figure 2. Correlation between the global visual analogue scale (VAS) scores and overall scores of a) the modified Riley score and b) the Geboes score
**Table 1.** Patient demographics and clinical characteristics (N=49)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>40.2 ± 12.9</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>27 (55)</td>
</tr>
<tr>
<td>White race, n (%)</td>
<td>48 (98)</td>
</tr>
<tr>
<td>Time since diagnosis, <em>months</em></td>
<td>81.9 ± 79.9</td>
</tr>
<tr>
<td>Time since symptoms, <em>months</em></td>
<td>98.2 ± 79.6</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td>1(2)</td>
</tr>
<tr>
<td>Mesalamine use, n (%)</td>
<td>40 (81.6)</td>
</tr>
<tr>
<td>Ulcerative colitis clinical score</td>
<td>7 ± 1.7</td>
</tr>
<tr>
<td>Daily stool frequency score</td>
<td>2.2 ± 0.9</td>
</tr>
<tr>
<td>Rectal bleeding score</td>
<td>1.3 ± 0.9</td>
</tr>
<tr>
<td>Assessment by patient</td>
<td>1.2 ± 0.8</td>
</tr>
<tr>
<td>Score on Inflammatory bowel disease questionnaire</td>
<td>145 ± 30.4</td>
</tr>
<tr>
<td>Modified Baron score</td>
<td>2.7 ± 0.8</td>
</tr>
<tr>
<td>Modified Riley histologic score**</td>
<td>4 (1-7)</td>
</tr>
<tr>
<td>Hemoglobin concentration, g/dL</td>
<td>132.4 ± 17.1</td>
</tr>
<tr>
<td>White-cell count, X 10⁻³/mm³</td>
<td>8.5 ± 2.2</td>
</tr>
</tbody>
</table>

*Plus-minus values are means ± standard deviation

**Median (interquartile range)**
Table 2. Overall agreement among 5 pathologists for the Geboes, modified Riley and the visual analogue scale (N=49)

<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image</td>
<td>Reader</td>
</tr>
<tr>
<td>GS (0-5)</td>
<td>1.321</td>
<td>0.1712</td>
</tr>
<tr>
<td>MRS (0-7)</td>
<td>3.678</td>
<td>0.7388</td>
</tr>
<tr>
<td>VAS</td>
<td>3.566</td>
<td>40.93</td>
</tr>
</tbody>
</table>

* Values represent ICCs

GS, Geboes score; MRS, modified Riley score; VAS, visual analogue scale; ICC, interclass correlation coefficient
Table 3. Agreement for items assessing architecture

<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image Reader</td>
<td>Digital Image x Reader</td>
</tr>
<tr>
<td>GS – structural</td>
<td>0.4200</td>
<td>0.0078</td>
</tr>
<tr>
<td>MRS crypt architectural irregularities</td>
<td>0.5156</td>
<td>0.0067</td>
</tr>
<tr>
<td>Percent crypt distortion</td>
<td>0.5765</td>
<td>0.0988</td>
</tr>
<tr>
<td>Crypt density</td>
<td>0.3859</td>
<td>0.0652</td>
</tr>
<tr>
<td>Surface contour</td>
<td>0.2605</td>
<td>0.1635</td>
</tr>
<tr>
<td>Lamina propria fibrosis</td>
<td>0.1001</td>
<td>0.1657</td>
</tr>
<tr>
<td>Patchiness</td>
<td>0.0230</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

* Values represent ICCs

GS, Geboes score; MRS, modified Riley score; ICC, interclass correlation coefficient
<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image Reader</td>
<td>Digital Image \times Reader</td>
</tr>
<tr>
<td>GS – neutrophils in epithelium</td>
<td>0.4334</td>
<td>0.1335</td>
</tr>
<tr>
<td>MRS – neutrophils in epithelium</td>
<td>0.2833</td>
<td>0.0865</td>
</tr>
<tr>
<td>GS – lamina propria eosinophils</td>
<td>0.1997</td>
<td>0.2009</td>
</tr>
<tr>
<td>GS and MRS – neutrophils in lamina propria</td>
<td>0.2219</td>
<td>0.0621</td>
</tr>
<tr>
<td>MRS – acute inflammatory cell infiltrate</td>
<td>0.2374</td>
<td>0.0640</td>
</tr>
<tr>
<td>MRS – crypt abscesses</td>
<td>0.0908</td>
<td>0.0264</td>
</tr>
<tr>
<td>Chicago IEG</td>
<td>0.9749</td>
<td>0.1477</td>
</tr>
<tr>
<td>Surface neutrophils</td>
<td>0.1726</td>
<td>0.0499</td>
</tr>
<tr>
<td>Neutrophils</td>
<td>0.6475</td>
<td>0.1023</td>
</tr>
<tr>
<td>Neutrophils in lamina propria</td>
<td>0.1276</td>
<td>0.0393</td>
</tr>
</tbody>
</table>

* Values represent ICCs

GS, Geboes score; MRS, modified Riley score; ICC, interclass correlation coefficient; IEG, increased epithelial granulocytes
### Table 5. Agreement for items assessing chronic inflammation

<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th></th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image Reader</td>
<td>Digital Image x Reader</td>
<td>Residual</td>
</tr>
<tr>
<td>GS – chronic inflammatory infiltrate</td>
<td>0.4921</td>
<td>0.0610</td>
<td>0.0685</td>
</tr>
<tr>
<td>MRS – chronic inflammatory cell infiltrate</td>
<td>0.4934</td>
<td>0.0601</td>
<td>0.0674</td>
</tr>
<tr>
<td>Basal plasmacytosis</td>
<td>1.0140</td>
<td>0.0817</td>
<td>0.1486</td>
</tr>
</tbody>
</table>

* Values represent ICCs

**GS, Geboes score; MRS, modified Riley score; ICC, interclass correlation coefficient**
Table 6. Agreement for items assessing epithelial injury (EI)

<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image Reader</td>
<td>Digital Image x Reader</td>
</tr>
<tr>
<td>GS – crypt destruction</td>
<td>0.3777</td>
<td>0.1112</td>
</tr>
<tr>
<td>GS – erosion or ulceration</td>
<td>1.2580</td>
<td>0.2011</td>
</tr>
<tr>
<td>MRS – erosion or ulceration</td>
<td>0.0959</td>
<td>0.0359</td>
</tr>
<tr>
<td>MRS – surface epithelial integrity</td>
<td>0.6289</td>
<td>0.0835</td>
</tr>
<tr>
<td>MRS – mucin depletion</td>
<td>0.4629</td>
<td>0.0000</td>
</tr>
<tr>
<td>Surface epithelium</td>
<td>0.8620</td>
<td>0.3476</td>
</tr>
<tr>
<td>Harpaz Index EI</td>
<td>0.6356</td>
<td>0.1509</td>
</tr>
<tr>
<td>Granulomas</td>
<td>0.0017</td>
<td>0.0059</td>
</tr>
</tbody>
</table>

* Values represent ICCs

GS, Geboes score; MRS, modified Riley score; ICC, interclass correlation coefficient
Table 7. Results from the consensus process: recommendations for minimizing disagreement among readers

<table>
<thead>
<tr>
<th>Area of disagreement</th>
<th>Recommendations</th>
</tr>
</thead>
</table>
| **Distinguishing epithelial damage from artefact**                                   | **Approach:**  
• Use epithelial restitution.  
• Rely on adjacent epithelial cells.  
• Rely on recovering epithelium, which is flattened to cuboidal, and evaluate areas that are stripped.  
• Require the visualization of epithelial injury and regeneration.  
• If there are fibrin and neutrophils then score this area as erosion. |
| **Assessment of basal plasmacytosis and architecture in the presence of granulation tissue** | **Approach:**  
• Determine the presence of unequivocal ulceration.  
• Identify granulation tissue, which implies that an ulcer had been present. |
| **Disconcordant pieces of the same sample**                                          | **Approach:**  
• Score the worst biopsy piece on the slide when calculating the total score.  
• For individual items that require additional scoring, score the worst areas on the slide. |
| **Distinguishing crypt destruction from damage or artefact**                         | **Definition of crypt destruction:**  
• Disappearance of crypt-lining cells and presence of a neutrophilic infiltrate. |
|                                                                                      | **Approach:**  
• Provide central readers with photos depicting the upper and lower ends of the spectrum of crypt destruction. |
| **Crypt distortion**                                                                 | **Definition of crypt shortening:**  
• An abnormal gap between the base of the crypt and the muscularis mucosae in perpendicular sections. |
| **Basal plasmacytosis**                                                             | **Definition of basal plasmacytosis:**  
• Plasma cells between the muscularis mucosae and the crypt base and/or the presence of plasma cells in the lower lamina propria between the crypts with loss |
of the normal top-heavy gradient.

**Approach:**
- *Basal plasmacytosis cannot be assessed in poorly oriented (transverse) sections.*
- *A score item qualifier (ex, tangentially oriented) should be added to indicate an inability to assess.*

**Lamina propria neutrophils**

**Definition:**
- *Greater than one neutrophil in the lamina propria or abnormal epithelium.*

**Approach:**
- *High quality slides (appropriately stained and sectioned) are required to observe lamina propria neutrophils and neutrophils in the epithelium.*

**Quality Samples**
- *Consider use of disposable versus reusable forceps (26).*
- *Consider use of oval elongated versus oval fenestrated cups. Samples obtained with oval elongated cups are generally deeper and those with oval fenestrated cups are generally larger. Fenestrated cups could potentially induce greater tissue damage.*

**Sections**
- *Samples should be 4–5 micron thick.*
- *Knives used for sectioning should be sharp (new) to avoid “chatter” and improve section quality to allow the identification of neutrophils and eosinophils.*

**Staining**
- *Use of haematoxylin and eosin should be balanced to improve the identification of neutrophils and eosinophils.*

**Orientation**
- *The use of “perpendicular sections” would address some of the problems associated with the assessment of basal plasmacytosis and crypt distortion. This requires orientation of the samples following acquisition, which can be both difficult and time consuming, particularly in patients with active disease involving the submucosa.*
Table 8. Subgroup analysis excluding 74 slides deemed as poor quality by the pathologists

<table>
<thead>
<tr>
<th></th>
<th>Variance Components</th>
<th>Reliability (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Digital Image</td>
<td>Reader</td>
</tr>
<tr>
<td><strong>Architecture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS – structural</td>
<td>0.4458</td>
<td>0.0039</td>
</tr>
<tr>
<td>MRS – crypt architectural irregularities</td>
<td>0.5289</td>
<td>0.0042</td>
</tr>
<tr>
<td>Percent crypt distortion</td>
<td>0.5856</td>
<td>0.0939</td>
</tr>
<tr>
<td>Crypt density</td>
<td>0.3874</td>
<td>0.0570</td>
</tr>
<tr>
<td>Surface contour</td>
<td>0.2538</td>
<td>0.1670</td>
</tr>
<tr>
<td>Lamina propria fibrosis</td>
<td>0.0951</td>
<td>0.1674</td>
</tr>
<tr>
<td>Patchiness</td>
<td>0.0239</td>
<td>0.0029</td>
</tr>
<tr>
<td>Chronic Inflammation</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>GS – chronic inflammatory infiltrate</td>
<td>0.5062</td>
<td>0.0570</td>
</tr>
<tr>
<td>MRS – chronic inflammatory cell infiltrate</td>
<td>0.5082</td>
<td>0.0565</td>
</tr>
<tr>
<td>Basal plasmacytosis</td>
<td>1.0610</td>
<td>0.0819</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>0.4306</td>
<td>0.2874</td>
</tr>
<tr>
<td></td>
<td>0.1368</td>
<td>0.0871</td>
</tr>
<tr>
<td></td>
<td>0.0894</td>
<td>0.0702</td>
</tr>
<tr>
<td></td>
<td>0.2433</td>
<td>0.1719</td>
</tr>
<tr>
<td></td>
<td>0.48 (0.37, 0.60)</td>
<td>0.47 (0.36, 0.59)</td>
</tr>
<tr>
<td></td>
<td>0.73 (0.66, 0.80)</td>
<td>0.72 (0.65, 0.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epithelial Injury</td>
<td>GS – crypt destruction</td>
<td>GS – erosion or ulceration</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td></td>
<td>0.3808</td>
<td>0.1212</td>
</tr>
<tr>
<td></td>
<td>1.264</td>
<td>0.2047</td>
</tr>
<tr>
<td></td>
<td>0.0952</td>
<td>0.0360</td>
</tr>
<tr>
<td></td>
<td>0.6327</td>
<td>0.0796</td>
</tr>
<tr>
<td></td>
<td>0.4732</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>0.8750</td>
<td>0.3579</td>
</tr>
<tr>
<td></td>
<td>0.6227</td>
<td>0.1412</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0070</td>
</tr>
<tr>
<td>Geboes, Modified Riley, and VAS scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS (0-5)</td>
<td>1.380</td>
<td>0.1642</td>
</tr>
<tr>
<td>MRS (0-7)</td>
<td>3.713</td>
<td>0.7357</td>
</tr>
<tr>
<td>VAS</td>
<td>363.8</td>
<td>54.42</td>
</tr>
</tbody>
</table>
REFERENCES:


18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174


Preface to Chapter 3:

This chapter summarizes conclusions of the previous chapters, provides study limitations, future recommendations based on the totality of evidence at present, and presents our future plans for this research endeavour.
Chapter 3: Summary, Strengths, limitations, recommendations and future plans.

Summary:

The introduction to this thesis consists of a systematic review, which explored the methodological development of histologic indices of disease activity in UC. We identified 18 different indices that have been used as outcome measures, however only two of these were developed using appropriate methodologies. Furthermore, assessment of critical validity criteria such as reliability and responsiveness were either lacking or incomplete for both of these instruments. Based on these findings, the reliability of the GS and MRS was assessed using digitized images. Biopsy samples were taken from patients with the full spectrum of histologic disease activity, according to the VAS. Our results indicated “substantial” to “almost perfect” inter-rater agreement but only “moderate” intra-rater agreement among the readers.

Several reasons for disagreement among pathologists were identified using a consensus process that involved 10 pathologists with expertise in IBD histopathology. The most convincing reasons for disagreement is the lack of standardised definitions. We could not conclude that the GS and/or MRS were reliable indices for the evaluation of disease activity in UC. In the future, we will conduct a reliability study in which we apply the rules that we identified in our study, such as standardized definitions and optimal quality histology slides.
Strengths:

There are several strengths to our studies. The systematic review we conducted included search of the main medical databases and followed strictly the PRISMA guidelines. The histologic slide readings performed in the agreement study were done by a group of clinical researchers with substantive knowledge in the area of IBD and a history of clinical trial execution and outcome measure development. Study efficiency was increased because the histologic slides were obtained from a previously conducted multi-center randomized controlled trial, and because central reading allowed the scoring of histology slides by expert readers who were thousands of miles away from each other. Our research protocol included a training phase for pathologists to maximize their reading capability. Finally, the histologic slides used for scoring were taken from UC patients with the full spectrum of histologic disease activity and mild to moderate clinical and endoscopic disease activity that are likely representative of UC patients seen in clinical practice, supporting the generalizability of our results.

Another strength to our study is conducting an ad-hoc consensus process to investigate the unexpected large discordance between intra- and inter-rater agreements, as it was critical to understand the sources of this variability. The digital slide quality standards that were agreed upon in the consensus process can serve as future guidelines that can potentially minimize the problem of poor quality slides.
Limitations:

We acknowledge several limitations to our studies. The literature review we conducted explored four major databases but did not include the use of MESH terms for its PubMed search. The process of study selection, however, was undertaken using a PRISMA algorithm. Based on our assessment of the studies available from the literature, a meta-analysis was not possible. Moreover, the absence of quality assessment tools for studies of index development prevented us from performing formal quality evaluations of the studies we included. Nevertheless, we do not feel that such evaluations would have changed our selection of indices since only two indices were of sufficient quality to include in our agreement study.

The agreement study limitations include the potential for selection bias, as readers were a convenience sample of pathologists with special training in gastroenterology and expertise in IBD and not randomly selected. In “real life” clinical practice; generally trained rather than IBD trained pathologists are more likely to score histology samples for patients with IBD and resources similar to those utilized in this study such as central reading, time and funding would not be readily available. Information bias may have resulted from the poor quality of some of the slides.

Recommendations:

Until a validated, reliable and responsive histologic scoring system is available to evaluate disease activity in UC, partially validated indices, such as the GS or MRS, can provide preliminary indicators of disease activity and/or
progression. Central reading is a novel method of evaluating images for clinical endpoint and can be used to assess histologic disease activity in UC. Standardization of definitions and measures of slide quality may improve future use of currently available indices, but formal testing is still needed to verify this suggestion.

**Future Plans:**

We plan to conduct an integrated agreement/responsiveness clinical study using CIMS to evaluate the GS and the MRS for their inter-rater reliability in the evaluation of histologic disease activity in UC using standard definitions and measures of slide quality. This is part of our long-term commitment towards developing and fully validating a histologic disease activity index that can serve as an endpoint in clinical trials and clinical practice.
Bibliography:


Appendix:

Search Strategies:

**MEDLINE - Search Strategy**
1. ulcerative colitis.mp. or exp Colitis, Ulcerative/
2. histol*.mp. or exp Pathology, Clinical/
3. exp Immunohistochemistry/ or immunohisto*.mp.
4. exp Pathology/ or patholog*.mp.
5. exp Biopsy/ or biops*.mp.
6. 2 or 3 or 4 or 5
7. (index or index* or indice or indice* or scale* or scali* or score or score* or scori* or riley or Gebo*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]
8. 1 and 6 and 7

**EMBASE - Search Strategy**
1. histol*.mp. or exp histology/
2. exp immunohistochemistry/ or immunohisto*.mp.
3. exp pathology/ or patholog*.mp.
4. exp biopsy/ or biops*.mp.
5. 1 or 2 or 3 or 4
6. ulcerative colitis.mp. or exp ulcerative colitis/
7. (index or index* or indice or indice* or scale* or scali* or score or score* or scori* or riley or Gebo*).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
8. 5 and 6 and 7
PubMed - Search Strategy
1. Search (colitis and ulcerat*)
2. Search (histolog* OR patholog* OR immunohisto* OR biops*)
3. Search (index OR index* OR indice OR indice* OR scale* OR scali* OR score OR score* OR scori* OR riley OR Gebo*)
4. Search (#1 AND #2 AND #3)

Cochrane Library (CENTRAL) - Search Strategy
1. (colitis and ulcerat*)
2. histol* or pathol* or immunohisto* or biops*
3. index or index* or indice or indice* or scale or scale* or scali* or score or score* or scori* or riley or Gebo*
4. #1 and #2 and #3

IBD/FBD Specialized Register - Search Strategy
1. ulcerat* AND (histol* or pathol* or immunohisto* or biops*)
Ethics Approval:

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Brian Fegan
File Number: 102713
Review Level: Delegated
Approved Local Adult Participants: 50
Approved Local Minor Participants: 0
Protocol Title: Inter-observer and Intra-observer Agreement in the Histological Evaluation of Ulcerative Colitis Using a Central Reader Based Image Management System
Department & Institution: Schulich School of Medicine and Dentistry, Medicine Dept of Robarts Research Institute
Sponsor:
Ethics Approval Date: June 12, 2012 Expiry Date: June 30, 2014
Documents Reviewed & Approved & Documents Received for Information:

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Comments</th>
<th>Version Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western University Protocol</td>
<td></td>
<td>Version 1 2012/05/25</td>
</tr>
</tbody>
</table>

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices. Consolidated Guidelines, and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REBs as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB’s periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000949.

*Signature*

Ethics Officer to Contact for Further Information

<table>
<thead>
<tr>
<th>Name</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jason Holland</td>
<td>519-661-3036</td>
<td><a href="mailto:jholland@uwyo.ca">jholland@uwyo.ca</a></td>
</tr>
<tr>
<td>Ross Kelly</td>
<td>519-661-3036</td>
<td><a href="mailto:rkelly@uwyo.ca">rkelly@uwyo.ca</a></td>
</tr>
<tr>
<td>Should Walton</td>
<td>519-661-3036</td>
<td><a href="mailto:shouldwalton@uwyo.ca">shouldwalton@uwyo.ca</a></td>
</tr>
</tbody>
</table>

This is an official document. Please retain the original in your files.

The University of Western Ontario
Office of Research Ethics
Support Services Building Room 5150 • London, Ontario • CANADA • N6G 1G9
P: 519-661-3036 • F: 519-850-2466 • ethics@uwyo.ca • www.uwyo.ca/research/ethics

114