Validation of the “Inflammatory Bowel Disease - Distribution, Chronicity, Activity (IBD-DCA) Score” for Ulcerative Colitis and Crohn’s disease

Corinna Lang-Schwarz1, 26, Miriam Angeloni2, 26, Abbas Agaimy2, Raja Atreya3, 25, Christoph Becker3, 25, Theresa Dregelies1, Silvio Danese4, Jean-François Fléjou5, Nikolaus Gaßler6, Heike I. Grabsch7, 8, Arndt Hartmann2, Kateřina Kamarádová9, Anja A. Kühl10, 25, Gregory Y. Lauwers11, Alessandro Lugli12, Iris Nagtegaal13, Markus F. Neurath3, 25, Georg Oberhuber14, 15, Laurent Peyrin-Biroulet16, Timo Rath3, Robert Riddell17, Carlos A. Rubio18, Kieran Sheahan19, Britta Siegmund20, 25, Herbert Tilg21, Vincenzo Villanacci22, Maria Westerhoff23, Fulvia Ferrazzi2, 24, 27, and Michael Vieth1, 2, 27

1 Institute of Pathology, Klinikum Bayreuth GmbH, Bayreuth, Germany; 2 Institute of Pathology, Friedrich-Alexander-University, Erlangen, Germany; 3 Medical Clinic 1, Department of Medicine & Deutsches Zentrum Immuntherapie DZI, University Hospital, Friedrich-Alexander-University, Erlangen, Germany; 4 Department of Gastroenterology, IBD Centre, Humanitas Research Hospital, Via A. Manzoni 56, 20089 Rozzano, Milan, Italy; 5 Pathology Department, Saint-Antoine Hospital, Pierre et Marie Curie University, Paris, France; 6 Institute for Legal Medicine, Section Pathology, University Hospital, Jena, Germany; 7 Department of Pathology, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands; 8 Pathology and Data Analytics, Leeds Institute of Medical Research at St James’s, University of Leeds, Leeds, UK; 9 The Fingerland

© The Author(s) 2021. Published by Oxford University Press on behalf of European Crohn’s and Colitis Organisation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Department of Pathology, Charles University Faculty of Medicine and University Hospital, Hradec Králové, Czech Republic; 10 Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, iPATH.Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany; 11 Department of Pathology, Moffitt Cancer Center, Tampa, FL, USA; 12 Institute of Pathology, University of Bern, Bern, Switzerland; 13 Department of Pathology, Radboud University Medical Centre, Nijmegen, The Netherlands; 14 INNPATH, Institute of Pathology, Tirol Kliniken, Innsbruck, Austria; 15 Patho im Zentrum, St. Pölten, Austria; 16 Department of Gastroenterology, Nancy University Hospital, Vandoeuvre, France & Inserm U1256, Lorraine University, Vandoeuvre, France; 17 Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 18 Department of Pathology, Karolinska Institute and University Hospital, Stockholm, Sweden; 19 Department of Pathology & Centre for Colorectal Disease, St Vincent’s University Hospital & University College, Dublin, Ireland; 20 Medical Department (Gastroenterology, Infectiology, Rheumatology), Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany; 21 Department of Internal Medicine I, Gastroenterology, Hepatology, Endocrinology and Metabolism, Medical University Innsbruck, Innsbruck, Austria; 22 Department of Pathology, Spedali Civili Brescia, Italy; 23 Department of Pathology, University of Michigan, Ann Arbor, MI, USA; 24 Department of Nephropathology, Institute of Pathology, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany; 25 The Transregio 241 IBDome Consortium
26 Corinna Lang-Schwarz and Miriam Angeloni contributed equally to the work and share first authorship.

27 Fulvia Ferrazzi and Michael Vieth contributed equally to the work and share last authorship.

**Corresponding author:**

Michael Vieth, MD, Institute of Pathology, Klinikum Bayreuth GmbH, Preuschwitzer Str. 101, 95445 Bayreuth, Germany

and

Institute of Pathology, Friedrich-Alexander-University Krankenhausstr. 8-10, 91054 Erlangen, Germany

Phone: +49 921 400 5602

Fax: ++49 921 400 88 5630

e-mail: vieth.lkpathol@uni-bayreuth.de
ABSTRACT

Background and aims

Histological scoring plays a key role in the assessment of disease activity in ulcerative colitis (UC) and is also important in Crohn’s disease (CD). Currently, there is no common scoring available for UC and CD. We aimed to validate the Inflammatory Bowel Disease (IBD) – Distribution (D), Chronicity (C), Activity (A) score (IBD-DCA score) for histological disease activity assessment in IBD.

Methods

Inter- and intra-rater reliability were assessed by 16 observers on biopsy specimen from 59 patients with UC and 25 patients with CD. Construct validity and responsiveness to treatment were retrospectively evaluated on a second cohort of 30 patients.

Results

Inter-rater reliability was moderate to good for the UC cohort (intraclass correlation coefficients (ICCs) = 0.645, 0.623, 0.767 for D, C and A, respectively) and at best moderate for the CD cohort (ICC = 0.690, 0.303, 0.733 for D, C and A, respectively). Intra-rater agreement ranged from good to excellent in both cohorts. Correlation with the Nancy Histological Index (NHI) was moderate and strong with the Simplified
Geboes Score (SGS) and a Visual Analog Scale (VAS). Large effect sizes (ES) were obtained for all three parameters. External responsiveness analysis revealed correlated changes between IBD-DCA score and NHI, SGS and VAS.

**Conclusions**

The IBD-DCA score is a simple histological activity score for UC and CD, agreed and validated by a large group of IBD specialists. It provides reliable information on treatment response. Therefore, it has potential value for use in routine diagnostics as well as clinical studies.

Key words: histological index, inflammatory bowel disease, IBD-DCA
INTRODUCTION

Recently, the importance of histological activity scoring in predicting clinical outcomes of patients with inflammatory bowel diseases (IBD) has become apparent. As such, several metaanalyses provide evidence that histological activity scoring can outperform endoscopical activity assessment in ulcerative colitis (UC) in predicting clinical endpoints such as the occurrence of flares, the need for corticosteroid use, and hospitalization for acute severe UC.\(^1\)\(^-\)\(^5\) In Crohn’s disease (CD) with isolated terminal ileum involvement and clinical remission, histologic healing was associated with decreased risk of clinical relapse, medication escalation, and need for corticosteroid use.\(^6\)

Furthermore, first evidence suggests that histological activity is superior in predicting the development of dysplasia and carcinoma compared to endoscopic assessment in UC.\(^7\)\(^,\)\(^8\) Histologic mucosal healing has emerged as the new ultimate treatment goal in UC and CD as it is associated with improved clinical outcome, prolonged remission, fewer hospitalisations and decreased surgery.\(^1\)\(^,\)\(^2\)\(^,\)\(^8\)\(^,\)\(^9\)\(^-\)\(^24\) However, the most suitable histologic target feature to define histologic mucosal healing, is yet to be found. Recent studies suggest that neutrophils might play a key role in this issue while other studies focus on architectural distortion as distinctive feature between histologic quiescent disease and true histologic normalization.\(^3\)\(^,\)\(^25\)\(^-\)\(^28\)

Therefore, the ideal histological score should not only be able to assess disease activity but also restoration of chronic inflammation to normal to distinguish between quiescent colitis and histologic architectural normalization. The utility of histological activity scoring is such that, in 2016, the Food and Drug Administration
(FDA) of the United States Department of Health and Human Services recommended that histological activity scoring in UC should be carried out in parallel with endoscopical assessment.\textsuperscript{29}

Unfortunately, there are many IBD scoring systems, some have never been fully validated, others are too complicated to practical use and many do not fulfill the currently accepted standards for index development.\textsuperscript{30, 31} Detailed reviews of as many as 30 histological scoring indices for UC and 13 for CD have been published by The Cochrane Collaboration, highlighting their respective advantages and disadvantages.\textsuperscript{30, 31}

Given the urgent clinical need for a usable and standardized histological scoring system in UC and CD, we aimed to develop and validate a simple histological activity scoring index for idiopathic IBD that could be used for clinical trials and routine daily pathology practice and is, at the same time, easy to calculate. In accordance with the existing scores, the new score is not meant to establish a diagnosis of UC or CD but to assess the amount and severeness of active and chronic changes in already known IBD.

The Inflammatory Bowel Disease - Distribution, Chronicity, Activity (IBD-DCA) score was initially developed during the International Consensus Conference on Inflammatory Bowel Disease, held in Erlangen, Germany, from 8th to 10th January 2020, with participants from 12 countries.\textsuperscript{32} The aim of this study was to validate the new score.
METHODS

Phase 1: Development of the Inflammatory Bowel Disease – Distribution (D), Chronicity (C), Activity (A) (IBD-DCA) score

The IBD-DCA score was proposed during the Consensus Conference. A detailed description regarding how it was developed has already been published. In brief, the score consists of three main parameters which also constitute the name of the new index and the order in which the score should be assessed:

- D for assessment of the distribution of overall active or chronic changes in the IBD colon biopsy, regardless of whether they are epithelial, architectural or inflammatory,
- C for assessment of features of chronic injury (architectural distortion or chronic inflammation),
- A for assessment of activity features (neutrophils).

For further substratification of the parameters D, C and A, additional items shown to have high inter- or intra-rater reliability in already existing scores in the literature were adopted and included in the new index (table 1).

An overview of the new IBD-DCA score is shown in table 2.

A score of D0 implies that C and A are also 0 (normal). Lymphoid aggregates or lymphoid follicles are part of the normal mucosa and do not qualify for D1 or D2.
If there is only one biopsy, the tissue area of this single biopsy is 100%. If there are more than one biopsy from one container, the tissue area of these biopsies together is 100%.

Scoring should be done for each container separately. However, in case of equal results for all sample sites, the IBD-DCA score might be reported for all biopsies at the end of the report. In this case, it must be stated explicitly that the scoring results were equal for all containers. Detailed recommendations regarding optimal biopsy sampling (including number of biopsies and sites) in order to maximize diagnostic information have also currently been published by our group.\(^{32}\)

Figure 1 shows histological examples for each possible parameter.

Figure 2 shows an example for assessment of the IBD-DCA score.

**Phase 2: Validation of the IBD-DCA Score**

**Reliability and blinding**

To test reliability of using the IBD-DCA score and to validate, whether acceptable agreement could be consistently reached among pathologists, digital images of Hematoxylin and Eosin (H&E) stained biopsies from known UC and CD cases were provided to 16 pathologists. Each pathologist evaluated the biopsies with the IBD-DCA score independently and blinded to patient data as well as clinical and endoscopical features. The pathologists represented different centers within Europe,
the United States and Canada. Fourteen observers attended the face to face meeting, 2 observers did not.

For assessment of intra-observer reliability, 8 pathologists scored the entire case series twice with a 3-month washout period in between.

The study set consisted of H&E-stained virtual slides of biopsies from patients with CD (n=25) and UC (n=59) selected randomly from the DC Pathos database (DC Systeme, Heiligenhaus, Germany, https://www.dc-systeme.de/) from the Institute of Pathology, Klinikum Bayreuth. The study set represented the full spectrum of disease activity from histological normalization to severe, ulcerative inflammation. The CD cases included biopsies from colon, terminal ileum, as well as stomach. All slides had been digitalized at the Institute of Pathology, Klinikum Bayreuth, Germany, using a NanoZoomer S360 scanner (Hamamatsu, Herrsching am Ammersee, Germany). The participants were given access to the slides online via a password protected platform using NDPView 2 (Hamamatsu, Herrsching am Ammersee, Germany). Each pathologist scored every slide for parameters D, C and A separately according to table 2. In addition, participants were asked to record whether an „A2“ was due to crypt abscesses, erosion or ulceration. These individual items of parameter A2 were finally not included into the IBD-DCA score due to lack of further informative impact.

Feasibility

For assessment of feasibility, a pathologist from a center in Europe and a pathologist from the United States independently measured the time (in seconds) required to assess the IBD-DCA score for each slide on the virtual scanned slides during the
first reading. The measured time includes the time until the slides open and are visible until the scoring was done.

Clinical responsiveness to treatment and construct validity

In addition to the digital slide assessment, the ability of the IBD-DCA score to predict clinical responsiveness to treatment was assessed using cases from a well-characterized cohort of patients by two pathologists (MV and CLS). Similar to the evaluation performed in the development of the Nancy Histological index (NHI), responsiveness was retrospectively assessed in 30 patients with UC using two sets of biopsy specimens from each patient, taken at two different time points during treatment. Patients were diagnosed with UC between 2014 and 2020 at the Institute of Pathology at Klinikum Bayreuth GmbH, Bayreuth, Germany. The median time interval between the first and second biopsy was 13 months (3 to 67 months). Information about treatment was available for 22 patients, most of them receiving combinations of oral and local treatment. Among them, 12 received oral Mesalazine (5-ASA) with or without additional topical therapy, three received oral corticosteroids in combination with Mesalazine and topical therapy, two were treated with Azathioprine, three received TNFα-antibody therapy, one was treated with Vedolizumab and one with the JAK-inhibitor Tofacitinib.

All patients showed a change in histological disease activity between the two time points (referred to as „baseline-condition“ versus „follow up-condition“). Among them:

- n=5 ranged from severe activity (DX, CX, A2) to histologic normalization (D0, C0, A0),
- n=15 ranged from moderate activity (DX, CX, A2) to histologic normalization (D0, C0, A0),

- n=5 ranged from moderate activity (DX, CX, A2) to mild activity (DX, CX, A1) and

- n=5 ranged from mild activity (DX, CX, A1) to histologic normalization (D0, C0, A0).

As the IBD-DCA score summarizes severe and moderate activity in its feature A2 due to its two-tiered design, cases with a change in disease activity from severe to moderate were not included. The Mayo Endoscopic Subscore (MES) was also available for each biopsy set. The MES had been assessed by different gastroenterologists with special interest in IBD during routine endoscopy. The slides were retrieved from the archives of the Institute of Pathology, Klinikum Bayreuth, Germany, after a search in the institutional database. A total of 60 slides (two slides per patient) were randomized and the observers were blinded to clinical data and visit number.

Each slide was first scored by the two observers using the IBD-DCA score.

In order to compare the IBD-DCA-Score with other established scoring systems, the two observers subsequently scored the same slides using the NHI and the Simplified Geboes Score (SGS) with a minimum of one month washout between each scoring. The NHI and the SGS were chosen for comparisons with the IBD-DCA score as the two observers were familiar with those two scores from their participation in clinical trials. In another separate reading, the pathologists again scored the slides evaluating disease activity on a 100 mm Visual Analog Scale (VAS) with ten step intervals from 0 (normal mucosa) to 10 (most severe disease.
activity). In addition, for each slide, the MES was retrieved from the biopsy submission form for comparison of histological changes in disease activity with the MES.\textsuperscript{42}

In summary, 60 slides of 30 patients were scored four times by the two pathologists in 4 different readings, with at least one month time between each reading to exclude a recall bias.

As the different indices cannot be readily converted one to another, conversions between them were established as shown in figure 3.

The construct validity of the IBD-DCA Score was evaluated in terms of correlation between the developed score and the NHI, the SGS, the VAS (as assessed by MV and CLS for responsiveness analyses) as well as the MES.

Statistics

All analyses were performed using the R statistical framework v. 3.6.0.\textsuperscript{43} The 16 pathologists were given identifiers from one to 16. Intra- and inter-rater reliability were assessed separately for items D, C and A within the two cohorts (UC and CD) by calculating the Intraclass Correlation coefficient (ICC).\textsuperscript{44} For inter-rater reliability, ICC estimates and their 95% confidence intervals (CI) were calculated using the R package \textit{irr} v.0.84.1 and a single-rating, absolute agreement, 2-way random effects model.\textsuperscript{45, 46} Analysis for intra-rater reliability was performed using the R package \textit{psych} v.1.9.12.31 and a single-rating, absolute agreement, 2-way mixed-effects
model. ICC values < 0.5, 0.5-0.75, 0.75-0.9 and > 0.9 indicated „poor“, „moderate“, „good“ and „excellent“ agreement, respectively.

The construct validity was quantified separately for baseline and follow-up conditions through the pairwise Kendall correlation coefficients ($\tau_B$) between the IBD-DCA score and the other indices (NHI, SGS, VAS and MES). The analysis was performed using the R packages *Kendall v.2.2* and *NSM3 v.1.14*. In order to perform correlation analyses, score values were re-mapped into different multi-grade score systems (supplemental tables 1-3) derived from the correspondence between indices shown in figure 3.

Construct validity and responsiveness analyses were conducted using the scores from observer MV, after confirming a good inter-rater reliability between CLS and MV for IBD-DCA score as well as for the other indices (results not shown).

Responsiveness analysis was conducted in terms of internal and external responsiveness. Internal responsiveness evaluates the ability of the IBD-DCA score to predict changes in disease activity between baseline and follow-up condition. Effect Size (ES) statistics were used to estimate the magnitude of change. ES was calculated as $ES = \frac{Z_w}{\sqrt{N}}$, where $Z_w$ is the z-score calculated with a Wilcoxon signed rank test for paired samples and N is the number of paired samples, using the R package *coin v.1.3-1*. ES values 0.1 - 0.3, 0.3 - 0.5 and > 0.5 represent small, moderate and large changes in the measure, respectively. External responsiveness evaluates whether changes in the IBD-DCA score correlate with changes in the other scoring systems (NHI, SGS, VAS and MES). To this aim,
pairwise Kendall correlation between score differences (baseline minus follow-up) was computed.

**Ethical statement**

The ethics committee of Friedrich-Alexander-University Erlangen-Nuremberg, Germany approved the study (study number: 175_20 Bc).

**RESULTS**

**Reliability**

The scoring of the UC cohort showed moderate inter-rater reliability for parameter D (ICC 0.645, 95% confidence interval (CI): 0.554-0.737), poor to moderate agreement for parameter C (ICC 0.568, 95% CI: 0.468-0.673) and moderate to good for parameter A (ICC 0.748, 95% CI: 0.671-0.82). The scoring of the CD cohort showed an inter-rater agreement from moderate to good for parameters D (ICC 0.655, 95% CI 0.515-0.801) and A (ICC 0.644; 95% CI 0.504-0.792) and poor for parameter C (ICC 0.303, 95% CI 0.183-0.496, supplemental table 4).

Pairwise inter-rater reliability analysis between the 16 raters showed the presence of outliers within the two cohorts, i.e. 5 raters who had poor pairwise agreement with other raters. After having clarified potential misunderstandings of scoring terminology (under consideration of table 2), outliers were asked to re-score the parameters (D, C, A) for which they had obtained a poor agreement with other raters.
After re-scoring an improvement in pairwise inter-rater agreement was observed (supplemental figure 1 for UC and supplemental figure 2 for CD) and ICC estimates of the inter-rater reliability across the 16 raters improved (table 3).

To assess intra-rater reliability, eight pathologists re-scored the slides a second time. For each parameter individual intra-rater agreements were summarized by calculating the median (range) ICC. For the UC cohort, an intra-rater agreement of moderate to excellent was reached for parameters D (median ICC 0.894) and C (median ICC 0.798) and of good to excellent for parameter A (median ICC 0.909). For the CD cohort, an intra-rater agreement of moderate to excellent was reached for parameter D (median ICC 0.854), of poor to excellent for parameter C (median ICC 0.714) and of good to excellent for parameter A (median ICC 0.888). Intra-rater results for both cohorts are shown in table 4. Median ICC and range for intra-rater agreement in both cohorts are expressed separately for each histological item (D, C, A).

Feasibility

The median (range) time required for IBD-DCA assessment for both observers was 20.5 seconds for the CD cases (7 - 151 seconds) and 26.4 seconds for UC (4.8 - 300 seconds).
Construct validity

Construct validity was evaluated separately for baseline and follow-up conditions relying on Kendall correlation coefficient ($\tau_B$). For the baseline condition the IBD-DCA score showed a moderate association with the NHI ($\tau_B = 0.595$) and a good association with the SGS ($\tau_B = 0.792$) as well as with the VAS ($\tau_B = 0.896$). For the follow-up condition an almost perfect pairwise association of the IBD-DCA score with the NHI ($\tau_B = 1$), the SGS ($\tau_B = 0.963$) and the VAS ($\tau_B = 0.994$) was obtained. In both baseline and follow-up conditions there was no association with the MES (table 5).

When comparing the individual MES grades and the IBD-DCA score in terms of number of matches/mismatches after having converted them according to the 3-tiered scoring system shown in supplemental table 3, it was possible to observe that in the baseline condition the majority of biopsies was assigned to IBD-DCA grades 1 and 2 irrespectively of their MES grade whereas in follow-up condition almost all biopsies were assigned to IBD-DCA grade 0 in accordance with the MES. Correlations between the individual MES grades with the IBD-DCA score are available as supplemental table 5.

Responsiveness

In internal responsiveness analysis all three histological parameters (D, C and A) showed a large magnitude of change (ES = 0.635, 1.09, 1.229 for D, C and A, respectively) (Table 6).
Changes in histopathological scores showed good degrees of correlation with each other, whereas correlation between histopathological scores and Mayo endoscopic subscore was much lower (Table 7).

**DISCUSSION**

The number of existing scoring systems for assessment of histological activity in ulcerative colitis and Crohn’s disease seems – at first glance - large enough to provide ‘the perfect index’ to every pathologist and corresponding clinician. To date, pathologists are either free in their choice of index and even whether to use them at all or make this choice jointly with their endoscopists.33-35, 52

Nevertheless, when considered more closely, the existing indices show limitations. The main limitation concerning the indices for CD is that none of them has been fully validated to date.31 Concerning UC, existing indices are very heterogenous in their complexity of assessment algorithms as well as their content of assessed items.30, 34 The Nancy Histological Index (NHI) and the Robarts histopathology index (RHI) have undergone the most validation for UC so far, but both include inflammatory features only, as architectural features were thought unlikely to be responsive to change following therapy.31, 33

Histological mucosal healing is not well defined to date. In their currently published position paper, the European Crohn’s & Colitis Organisation define histological remission in its strictest way as return to normal.54 Therefore, crypt architectural distortion might be one of the new key features in this issue as it
distinguishes between quiescent UC (which has architectural distortion) and true histological normalization (which looks like normal colon). 25, 26

This is especially strengthened by the findings of Christensen et al., who showed increased odds of relapse-free survival for histologic normalization in comparison to endoscopic healing or histologic quiescence on a large cohort of 646 patients. 3 In a recently published systematic review and meta-analysis including 28 studies with 2806 patients (2677 with UC and 129 with CD), crypt architectural irregularities were also one of the individual features that predicted relapse as were basal plasmacytosis, neutrophilic infiltrations and mucin depletion. 5 Concerning CD, the role of histology in activity assessment is not definitely clear yet. However, according to Christensen et al, histologic healing has also shown to be superior in predicting clinical outcomes in ileal CD than endoscopic healing. 6

As activity in UC patients should be assessed by endoscopy in conjunction with histology according to current FDA recommendations, the participants of the International Consensus Conference on Activity Scoring in IBD developed a histological score for both UC and CD that could easily be implemented into routine daily practice and is able to distinguish between histological remission and histologic normalization.

A major strength of the proposed new IBD-DCA score relies in its good inter- and intra-rater reliability, which was assessed in – to the best of our knowledge – the largest group of pathologists to date. Without special training in scoring, inter-rater reliability revealed moderate ICC estimates for all three parameters D (distribution), C (chronicity) and A (activity) within the UC cohort and moderate ICC estimates for
parameters D and A within the CD cohort. These values further improved when 4 out of the 5 initial “outlier” raters performed a second round of scoring. One of these “outliers” did not attend the face to face meeting which could partially explain the initial poor agreement with the other raters. However, it is worth noting that the results of observer number 11, who also did not attend the meeting, were consistent with those of the other raters from the very beginning, emphasizing the simplicity of the score.

Overall, the IBD-DCA score reached comparable ICC estimates to those of other published indices, despite including 4 to 5 times more observers.³³⁻³⁷, ³⁵

To the best of our knowledge and according to Mosli et al., none of the established scoring indices assessed feasibility so far, nor did the latest developed Robarts histopathology index (RHI).³⁰⁻³⁴ In this study, feasibility assessment was done on whole slide images (virtual slides), demonstrating that the IBD-DCA score is applicable in digital pathology. Scoring of virtual slides is a pillar of digital pathology and will be probably used more frequently in the future.

Another advantage of the IBD-DCA score compared to other scoring systems is its simplicity. It is configured to intuitively follow the usual practice of a pathologist assessing a bowel biopsy specimen from low to high power magnification. Features of chronic injury as well as active inflammatory findings are – apart from normal – only divided into two levels of severity, creating a two-tiered system for chronic as well as active inflammation. The advantage of a two-tiered system is strengthened by the findings of Lemmens et al.⁵⁵ In their correlation analysis between endoscopic and histological scores some scores were infrequently used. This was especially true
for the middle grades. The fact that inter-rater agreement in the IBD-DCA score for parameter A was good for UC, but moderate for the stratification of A2 into its special features further confirms this proposed concept. Erosions and ulcers have also been summarized as „mucosal breaks“ in other gastrointestinal diseases due to lack of negative clinical impact.56

Our study had some limitations. The main limitation is that the validation was performed retrospectively on slides from routine work. Further prospective validation in additional datasets, preferably from a randomised controlled trial is clearly necessary. Although the CD cohort data are promising, there is also clearly a need for further prospective validation on larger cohorts for the upper and lower gastrointestinal tract including a study set for responsibility analysis to prove potential applicability of the IBD-DCA score for CD as the role of histological activity assessment in CD is not yet definitely clear due to the discontinous and transmural nature of the disease.

In this study, we presented the IBD-DCA score that has been developed with international consensus and validated in its interobserver agreement by a large group of pathologists from Europe, the US and Canada.

Although further studies are clearly necessary, our findings open new avenues for the clinical use of the IBD-DCA score for routine use in histological assessment of IBD activity.
FUNDING INFORMATION

This work was supported by the Deutsche Forschungsgemeinschaft TRR241 projects INF, A03, C02, C03 and C04 (grant number 375876048).

The Deutsche Forschungsgemeinschaft funds the Heisenberg Professorship of Raja Atreja (grant number 316866639).

ACKNOWLEDGEMENTS

The participants thank Tilman Schulz, MD, Institute of Pathology, Klinikum Bayreuth GmbH, Bayreuth, Germany, for providing us with the schemes in figure 2.
REFERENCES


21. Feagins LA, Melton SD, Iqbal R, Dunbar KB, Spechler SJ. Clinical implications of histologic abnormalities in colonic biopsy specimens from patients with ulcerative colitis in clinical remission. Inflamm Bowel Dis 2013;18:1477–82. DOI:10.1097/MIB.0b013e318281f4ae

predicts sustained clinical remission in patients with early-stage Crohn's disease.

Gastroenterology 2010;138(2):463-8; quiz e10-1.
DOI:10.1053/j.gastro.2009.09.056

DOI:10.1002/ibd.22977


DOI:10.3109/00365529409090475


44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.


the TOUCHSTONE randomised controlled trial. Gut 2019;68:1162-8. DOI:10.1136/gutjnl-2018-316702


### TABLES

Table 1. Histological features used to construct the IBD-DCA score.

<table>
<thead>
<tr>
<th>Item</th>
<th>ICC (95% CI)</th>
<th>References</th>
<th>Corresponding parameter in IBD-DCA score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic inflammatory infiltrate</td>
<td>0.750 (0.640-1)</td>
<td>NHI(^{33})</td>
<td>C</td>
</tr>
<tr>
<td>Basal plasmacytosis</td>
<td>0.63 (0.48-0.74)</td>
<td>Mosli et al(^{35})</td>
<td>C 1 and C 2</td>
</tr>
<tr>
<td>Crypt architectural distortion</td>
<td>0.72 (0.59-0.80)</td>
<td>Mosli et al for MRS</td>
<td>C 1</td>
</tr>
<tr>
<td>Acute inflammatory infiltrate</td>
<td>0.772 (0.704-0.940)</td>
<td>NHI(^{33})</td>
<td>A 1</td>
</tr>
<tr>
<td>Lamina propria neutrophils</td>
<td>0.61 (0.48-0.69)</td>
<td>Mosli et al and</td>
<td>A 1</td>
</tr>
<tr>
<td>Neutrophils in epithelium</td>
<td>0.74 (0.68-0.80)</td>
<td>Bressenot et al for GS(^{34, 37, 38})</td>
<td>A 1</td>
</tr>
<tr>
<td>Erosion</td>
<td>0.79 (0.66-0.86)</td>
<td>RHI(^{34})</td>
<td>A 2</td>
</tr>
<tr>
<td></td>
<td>0.82 (0.77-0.88)</td>
<td>Bressenot et al for</td>
<td></td>
</tr>
<tr>
<td>Ulceration</td>
<td>GS(^{37, 38})</td>
<td>0.865 (0.750-1)</td>
<td>NHI(^{33})</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>0.79 (0.66-0.86)</td>
<td>RHI(^{34})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.82 (0.77-0.88)</td>
<td>Bressenot et al for</td>
<td>A 2</td>
</tr>
<tr>
<td></td>
<td>0.90 (0.79-0.97)</td>
<td>GS and for</td>
<td>Gramlich Index(^{37-39})</td>
</tr>
</tbody>
</table>

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval; NHI, Nancy Histological Index; RHI, Robarts histopathology index; MRS, Modified Riley Score; GS, Geboes Score; RI, Riley Index.
Table 2: Components of the new IBD-DCA score

<table>
<thead>
<tr>
<th>Variable</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distribution (D)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>1</td>
<td>&lt; 50% of tissue affected per same biopsy site</td>
</tr>
<tr>
<td>2</td>
<td>&gt; 50% of tissue affected per same biopsy site</td>
</tr>
<tr>
<td><strong>Chronic features (C)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>1</td>
<td>Crypt distortion and/or mild lymphoplasmacytosis</td>
</tr>
<tr>
<td>2</td>
<td>Marked lymphoplasmacytosis (and/or marked basal plasmacytosis)</td>
</tr>
<tr>
<td><strong>Activity features (A)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>1</td>
<td>Two or more neutrophils in lamina propria in one high power field (HPF) and/or intraepithelial neutrophils (any number)</td>
</tr>
<tr>
<td>2</td>
<td>Crypt abscesses, erosions, ulcers</td>
</tr>
</tbody>
</table>
Table 3. Intraclass Correlation coefficient estimate (with 95% CI) for inter-rater agreement in ulcerative colitis (UC) and Crohn’s disease (CD) cohorts after re-scoring.

<table>
<thead>
<tr>
<th></th>
<th>ICC (95% CI)</th>
<th>CD (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D - Distribution</strong></td>
<td>0.645 (0.554 - 0.737)</td>
<td>0.69 (0.556 - 0.824)</td>
</tr>
<tr>
<td><strong>C - Chronicity</strong></td>
<td>0.623 (0.532 - 0.717)</td>
<td>0.303 (0.183 - 0.496)</td>
</tr>
<tr>
<td><strong>A - Activity</strong></td>
<td>0.767 (0.695 - 0.835)</td>
<td>0.733 (0.604 - 0.852)</td>
</tr>
</tbody>
</table>
Table 4. Intra-rater agreement for both cohorts.

<table>
<thead>
<tr>
<th></th>
<th>UC</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median ICC</td>
<td>ICC Range</td>
</tr>
<tr>
<td>D - Distribution</td>
<td>0.894</td>
<td>0.745 – 1</td>
</tr>
<tr>
<td>C - Chronicity</td>
<td>0.798</td>
<td>0.706 – 1</td>
</tr>
<tr>
<td>A - Activity</td>
<td>0.909</td>
<td>0.884 - 0.986</td>
</tr>
</tbody>
</table>
Table 5. Estimates (with 95% CIs) of pairwise Kendall correlation coefficient ($\tau_B$) between IBD-DCA and other compared indexes in baseline and follow-up conditions.

<table>
<thead>
<tr>
<th>Compared index</th>
<th>Baseline condition</th>
<th>Follow up condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau_B$ (95% CI)</td>
<td>Two-sided p-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHI</td>
<td>$0.595 (0.418-0.773)$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1 (0.804-1.196)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt; 1E-10$</td>
</tr>
<tr>
<td>SGS</td>
<td>$0.792 (0.611-0.972)$</td>
<td>$2.38E-05$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.963 (0.827-1.1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt; 1E-10$</td>
</tr>
<tr>
<td>VAS</td>
<td>$0.896 (0.721-1.071)$</td>
<td>$1.19E-06$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.994 (0.859-1.13)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt; 1E-10$</td>
</tr>
<tr>
<td>MES</td>
<td>$0.01 (-0.162-0.181)$</td>
<td>$0.978$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.111 (-0.039-0.26)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.573$</td>
</tr>
</tbody>
</table>

Abbreviations: NHI, Nancy Histological Index; SGS, Simplified Geboes Score; VAS, Visual Analog Scale; MES, Mayo Endoscopic Subscore.
Table 6. Effect sizes for the histological items of the DCA-score.

<table>
<thead>
<tr>
<th>Effect Size (ES)</th>
<th>Z-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D- Distribution</td>
<td>0.635</td>
<td>3.4765</td>
</tr>
<tr>
<td>C - Chronicity</td>
<td>1.09</td>
<td>5.972</td>
</tr>
<tr>
<td>A - Activity</td>
<td>1.229</td>
<td>6.731</td>
</tr>
</tbody>
</table>
Table 7. Kendall correlation coefficient (95% CI) between score changes in histopathological and endoscopic scores.

<table>
<thead>
<tr>
<th></th>
<th>NHI</th>
<th>SGS</th>
<th>VAS</th>
<th>MES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DCA</strong></td>
<td>0.753</td>
<td>0.89</td>
<td>0.953</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.574 to 0.933)</td>
<td>(0.722 to 1.058)</td>
<td>(0.813 to 1.093)</td>
<td>(-0.102 to 0.322)</td>
</tr>
<tr>
<td><strong>NHI</strong></td>
<td>0.565</td>
<td>0.739</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.338 to 0.792)</td>
<td>(0.571 to 0.908)</td>
<td>(0.064 to 0.513)</td>
<td></td>
</tr>
<tr>
<td><strong>SGS</strong></td>
<td>0.841</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.671 to 1.011)</td>
<td>(-0.121 to 0.321)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VAS</strong></td>
<td></td>
<td></td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.075 to 0.465)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: NHI, Nancy Histological Index; SGS, Simplified Geboes Score; VAS, Visual Analog Scale; MES, Mayo Endoscopic Subscore.
Figure 1 Hematoxylin and Eosin (H&E)-stained histological examples for all possible parameters of the IBD-DCA score. a – D0, normal mucosa (and C0 and A0, magnification 8.02x), b – D1, less than 50% of biopsies affected (magnification 3.3x), c – D2, ≥50% of biopsies affected (magnification 3.61x), d – C0 and A0, normal mucosa (magnification 20x), e – C1, crypt architectural distortion (magnification 22x), f – C2, architectural distortion and marked lamina propria lymphoplasmacytosis including basal lymphoplasmacytosis (magnification 12.6x), g – A1, intraepithelial neutrophils (white arrows, magnification 35.7x), h-i – A2, crypt abscesses (h) and ulcer (i) (magnification 33.8x and 16.9x, respectively).
1. **Distribution: D**  
Assess parameter D as amount of overall affected tissue in scanning magnification (2.5-4x).

Example shows four biopsies, affected by inflammatory and architectural changes in >50% of tissue, resulting in "D2".

2. **Chronicity: C**  
Assess parameter C in magnification 4 to 10x.

Example shows architectural distortion as well as a particularly prominent bandlike (lympho-) plasmacytosis corresponding to "C2".

3. **Activity: A**  
Assess parameter A in higher magnification.

Example shows a cluster of neutrophilic granulocytes in the tunica propria as well as some granulocytes in the crypt epithelium resulting in "A1".

Summary IBD-DCA score for shown example is: D2 C2 A1.
Figure 3: Conversions of IBD-DCA-Score for correlation analyses versus NHI, SGS, VAS and MES. Correlating scoring values are indicated in same colours. MES grades 0 to 1 both refer to normal mucosa respectively endoscopic remission.

<table>
<thead>
<tr>
<th>IBD-DCA</th>
<th>NHI</th>
<th>SGS</th>
<th>VAS</th>
<th>MES</th>
</tr>
</thead>
<tbody>
<tr>
<td>D0 and C0 and A0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0-1</td>
</tr>
<tr>
<td>C1 and A0</td>
<td></td>
<td>0.1-1.1</td>
<td>1-2</td>
<td></td>
</tr>
<tr>
<td>C2 and A0</td>
<td>1</td>
<td>1.2</td>
<td>3-4</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>3</td>
<td>2B.1-4.0</td>
<td>5-7</td>
<td>2</td>
</tr>
<tr>
<td>A2</td>
<td>4</td>
<td>4.1-4.4</td>
<td>8-10</td>
<td>3</td>
</tr>
</tbody>
</table>

Abbreviations: NHI Nancy Histological Index, SGS Simplified Geboes Score, VAS Visual Analog Scale, MES Mayo Endoscopic Subscore
COMPETING INTERESTS STATEMENT

Michael Vieth reports lecture fees from Falk, Shire, Lilly, Malesci, Pentax, Olympus and AstraZeneca.

Raja Atreya has served as a speaker, or consultant, or received research grants from AbbVie, Biogen, Boehringer Ingelheim, Celgene, Celltrion, Dr. Falk Pharma, Ferring, InDex Pharmaceuticals, Janssen-Cilag, MSD Sharp & Dome, Pfizer, Roche Pharma, Samsung Bioepis, Takeda.

Laurent Peyrin-Biroulet reports personal fees from Galapagos, AbbVie, Janssen, Genentech, Ferring, Tillots, Pharmacosmos, Celltrion, Takeda, Boehringer Ingelheim, Pfizer, Index Pharmaceuticals, Sandoz, Celgene, Biogen, Samsung Bioepis, Alma, Sterna, Nestle, Inotrem, Enterome, Allergan, MSD, Roche, Arena, Gilead, Hikma, Amgen, BMS, Vifor, Norgine; Mylan, Lilly, Fresenius Kabi, Oppilan Pharma, Sublimity Therapeutics, Applied Molecular Transport, OSE Immunotherapeutics, Enthera, Theravance; grants from Abbvie, MSD, Takeda; stock options: CTMA.

Gregory Y. Lauwers is a consultant for Alimentiv.
Britta Siegmund has served as Consultant for Abbvie, BMS, Boehringer, Celgene, Falk, Janssen, Lilly, Pfizer, Prometheus, Takeda and has received speaker’s fees from Abbvie, CED Service GmbH, Falk, Ferring, Janssen, Novartis, Pfizer, and Takeda as a representative of Charité – Universitätsmedizin Berlin.


AUTHORS’ CONTRIBUTIONS

Conception and design of the study: Corinna Lang-Schwarz, Michael Vieth, Markus F. Neurath, Arndt Hartmann

Acquisition of data: Corinna Lang-Schwarz, Abbas Agaimy, Jean-François Fléjou, Arndt Hartmann, Kateřina Kamarádová, Anja A. Kühl, Gregory Y. Lauwers, Alessandro Lugli, Iris Nagtegaal, Georg Oberhuber, Robert Riddell, Carlos A. Rubio, Kieran Sheahan, Vincenzo Villanacci, Maria Westerhoff, Michael Vieth

Article drafting: Corinna Lang-Schwarz, Maria Westerhoff

Analysis and interpretation of data: Miriam Angeloni, Fulvia Ferrazzi
Critical revision of the article for important intellectual content: Raja Atreya, Christoph Becker, Theresa Dregelies, Silvio Danese, Markus F. Neurath, Laurent Peyrin-Biroulet, Timo Rath, Robert Riddell, Britta Siegmund, Herbert Tilg, Maria Westerhoff

Final approval of the submitted version: all authors

DATA AVAILABILITY STATEMENT

Additional data, as far as not published, are available on demand via e-mail from the corresponding author.