Accepted Manuscript

Development and Validation a Magnetic Resonance Index for Assessing Fistulas in Patients With Crohn's Disease

Pieter Hindryckx, Vipul Jairath, Guangyong Zou, Brian G. Feagan, William J. Sandborn, Jaap Stoker, Reena Khanna, Larry Stitt, Tanja van Viegen, Lisa M. Shackelton, Stuart A. Taylor, Cynthia Santillan, Banafsche Mearadji, Geert D'Haens, Marie-Paule Richard, Julian Panes, Jordi Rimola

PII: S0016-5085(19)41120-7
DOI: https://doi.org/10.1053/j.gastro.2019.07.027
Reference: YGAST 62778

To appear in: *Gastroenterology*
Accepted Date: 16 July 2019

Please cite this article as: Hindryckx P, Jairath V, Zou G, Feagan BG, Sandborn WJ, Stoker J, Khanna R, Stitt L, van Viegen T, Shackelton LM, Taylor SA, Santillan C, Mearadji B, D'Haens G, Richard M-P, Panes J, Rimola J, Development and Validation a Magnetic Resonance Index for Assessing Fistulas in Patients With Crohn's Disease, *Gastroenterology* (2019), doi: https://doi.org/10.1053/j.gastro.2019.07.027.

ORIGINAL ARTICLE

Title: Development and Validation a Magnetic Resonance Index for Assessing Fistulas in Patients With Crohn's Disease

Short title: *A Novel MRI Index for Fistulizing CD*

Pieter Hindryckx,[1,2] Vipul Jairath,[1,3,4] Guangyong Zou,[1,4] Brian G. Feagan, [1,3,4] William J. Sandborn,[1,5] Jaap Stoker,[6] Reena Khanna[1,3], Larry Stitt,[1,4] Tanja van Viegen,[1] Lisa M. Shackelton,[1] Stuart A. Taylor,[7] Cynthia Santillan,[8] Banafsche Mearadji,[6] Geert D'Haens,[1,9] Marie-Paule Richard,[10] Julian Panes,[11] Jordi Rimola[12]

Affiliations: [1]Robarts Clinical Trials, Inc., London, Ontario, Canada; [2]Department of Gastroenterology, University of Ghent, Ghent, Belgium; [3]Department of Medicine, Division of Gastroenterology, University of Western Ontario, London, Ontario, Canada; [4]Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada; [5]Division of Gastroenterology, University of California San Diego, La Jolla, California, USA; [6]Department of Radiology and Nuclear Medicine, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; [7]Centre for Medical Imaging, University College London, London, UK; [8]Department of Radiology, University of California San Diego, La Jolla, California, USA; [9]Inflammatory Bowel Disease Centre, Academic Medical Centre, Amsterdam, The Netherlands; [10]Tigenix, SAU, Madrid, Spain; [11]Department of Gastroenterology, Hospital Clínic de Barcelona, IDIBAPS, CIBERehd, Barcelona, Spain; [12]Department of Radiology, Hospital Clínic de Barcelona, Barcelona, Spain.

Grant support: None

Abbreviations: MRI, magnetic resonance imaging; CD, Crohn's disease; ICC, intraclass correlation coefficient; VAS, visual analogue scale; MAGNIFI-CD, Magnetic Resonance Novel Index for Fistula Imaging in Crohn's Disease; VAI, Van Assche Index; mVAI, modified Van Assche Index; CI, confidence interval; RCT, randomized controlled trial; ANOVA, analysis of variance; PDAI, Perianal Disease Activity Index; CDAI,

Crohn's Disease Activity Index; IBDQ, Inflammatory Bowel Disease Questionnaire; CRP, C-reactive protein; SD, standard deviation; ROC, receiver operating curve

Correspondence to: Dr. Vipul Jairath, Associate Professor of Medicine, Departments of Medicine (Division of Gastroenterology), Epidemiology and Biostatistics, University of Western Ontario 1151 Richmond St., London, Ontario, Canada, N6A 5B6. Phone: 519-685-8500; fax: 519-663-3658; email: vjairath@uwo.ca.

Disclosures: PH has received consulting fees from Abbvie and Takeda, and speaker's fees from Ferring, Falk Pharma, Vifor Pharma, Tillotts Pharma, Chiesi, Takeda and Abbvie; **VJ** has received consulting fees from AbbVie, Eli Lilly, GlaxoSmithKline, Arena pharmaceuticals, Genentech, Pendopharm, Sandoz, Merck, Takeda, Janssen, Robarts Clinical Trials, Topivert, Celltrion, and speaker's fees from Takeda, Janssen, Shire, Ferring, Abbvie, Pfizer; **BGF** has received grant/research support from AbbVie Inc., Amgen Inc., AstraZeneca/MedImmune Ltd., Atlantic Pharmaceuticals Ltd., Boehringer-Ingelheim, Celgene Corporation, Celltech, Genentech Inc/Hoffmann-La Roche Ltd., Gilead Sciences Inc., GlaxoSmithKline (GSK), Janssen Research & Development LLC., Pfizer Inc., Receptos Inc./Celgene International, Sanofi, Santarus Inc., Takeda Development Center Americas Inc., Tillotts Pharma AG and UCB; consulting fees from Abbott/AbbVie, Akebia Therapeutics, Allergan, Amgen, Applied Molecular Transport Inc., Aptevo Therapeutics, Astra Zeneca, Atlantic Pharma, Avir Pharma, Biogen Idec, BioMx Israel, Boehringer-Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan/Biogen, EnGene, Ferring Pharma, Roche/Genentech, Galapagos, GiCare Pharma, Gilead, Gossamer Pharma, GSK, Inception IBD Inc, JnJ/Janssen, Kyowa Kakko Kirin Co Ltd., Lexicon, Lilly, Lycera BioTech, Merck, Mesoblast Pharma, Millennium, Nestle, Nextbiotix, Novonordisk, Pfizer, Prometheus Therapeutics and Diagnostics, Progenity, Protagonist, Receptos, Salix Pharma, Shire, Sienna Biologics, Sigmoid Pharma, Sterna Biologicals, Synergy Pharma Inc., Takeda, Teva Pharma, TiGenix, Tillotts, UCB Pharma, Vertex Pharma, Vivelix Pharma, VHsquared Ltd. and Zyngenia; speakers bureau fees from Abbott/AbbVie, JnJ/Janssen, Lilly, Takeda, Tillotts and UCB Pharma; is a scientific advisory board member for Abbott/AbbVie, Allergan, Amgen, Astra Zeneca, Atlantic Pharma, Avaxia Biologics Inc., Boehringer-Ingelheim, Bristol-Myers

Squibb, Celgene, Centocor Inc., Elan/Biogen, Galapagos, Genentech/Roche, JnJ/Janssen, Merck, Nestle, Novartis, Novonordisk, Pfizer, Prometheus Laboratories, Protagonist, Salix Pharma, Sterna Biologicals, Takeda, Teva, TiGenix, Tillotts Pharma AG and UCB Pharma; and is the Senior Scientific Officer of Robarts Clinical Trials Inc.; **WJS** has received research grants from Atlantic Healthcare Limited, Amgen, Genentech, Gilead Sciences, Abbvie, Janssen, Takeda, Lilly, Celgene/Receptos; consulting fees from Abbvie, Allergan, Amgen, Boehringer Ingelheim, Celgene, Conatus, Cosmo, Escalier Biosciences, Ferring, Genentech, Gilead, Janssen, Lilly, Miraca Life Sciences, Nivalis Therapeutics, Novartis Nutrition Science Partners, Oppilan Pharma, Otsuka, Paul Hastings, Pfizer, Precision IBD, Progenity, Prometheus Laboratories, Ritter Pharmaceuticals, Robarts Clinical Trials Inc., Salix, Shire, Seres Therapeutics, Sigmoid Biotechnologies, Takeda, Tigenix, Tillotts Pharma, UCB Pharma and Vivelix; and stock options from Ritter Pharmaceuticals, Oppilan Pharma, Escalier Biosciences, Precision IBD and Progenity; **JS** has received consulting fees from Robarts Clinical Trials, Inc.; **RK** has received consulting fees from AbbVie, Encycle, Janssen, Pfizer, Takeda and Robarts Clinical Trials Inc.; and speaker fees from AbbVie, Janssen, Shire and Takeda; **LS** is an employee of Robarts Clinical Trials, Inc.; **TVV** is an employee of Robarts Clinical Trials, Inc.; **LMS** has received consulting fees from Robarts Clinical Trials, Inc., **SAT** has received consulting fees from Robarts Clinical Trials, Inc.; **CS** has received consulting fees from Robarts Clinical Trials, Inc.; **BM** has received consulting fees from Robarts Clinical Trials, Inc.; **GZ** is an employee of Robarts Clinical Trials, Inc.; **G'DH** has received consulting fees from Robarts Clinical Trials; personal fees from Ablynx, Amakem, Amgen, AM Pharma, Boehringer-Ingelheim, Bristol Myers Squibb, Cosmo, Celgene, Celtrion, Covidien, Engene, Galapagos, Medimetrics, Mundipharma, Mitsubishi, Novonordisk, Pfizer, Receptos, Salix, Sandoz, Setpoint, Shire, Teva, Tigenix, Topivert, Versant and Vifor; grants and personal fees from Abbvie, Ferring, Glaxo Smith Kline, Jansen Biologics, Hospira, Merck Sharp Dome, Prometheus Labs, Robarts Clinical Trials, Takeda and Tillotts; and grants from Dr Falk Pharma and Photopill; **M-PR** is a former employee of Tigenix; **JP** has received consulting fees from Abbvie, Arena, Boehringer-Ingelheim, Celgene, Genentech, GoodGut, GSK, Janssen, MSP, Nestlé, Oppilan, Pfizer, Progenity, Takeda,

Theravance, TiGenix and lecture fees from Abbvie, Biogen, Janssen, MSD, Pfizer, Takeda; **JR** has received consulting fees from Robarts Clinical Trials, Inc., Takeda and Tigenix, grant/research support from AbbVie and Genentech, and speaker/honoraria fees from AbbVie, MSD, and Takeda.

Author contributions: Study concept and design: PH, VJ, BGF, WJS, GZ, JP, JR; Acquisition of data: TVV, M-PR, LS; Analysis and interpretation of data, PH, VJ, BGF, JS, LS, TVV, LMS, SAT, CS, BM, GZ, JP, JR; Drafting of manuscript: PH, VJ, LMS; Critical revision of the manuscript for important intellectual content: all authors; Statistical analysis: LS, GZ; Study supervision: PH, VJ, TVV, JR.

Abstract:

Background & Aims: There is no validated magnetic resonance imaging (MRI) index for assessment of perianal fistulas in patients with Crohn's disease (CD). We developed and internally validated a new instrument.

Methods: We used paired baseline and week 24 MRI scans from 160 participants in a randomized placebo-controlled trial of stem cell therapy for patients with perianal fistulizing CD. Four radiologists scored disease activity using index items identified during previous studies and exploratory items. Reliability was assessed using intraclass correlation coefficients. We developed an index using backward elimination linear regression analysis in which potential independent variables were items having intraclass correlation coefficients of at least 0.4 and the dependent variable was perianal fistulizing disease activity, measured on a 100 mm visual analogue scale. The final model was internally validated using the .632 bootstrap method to correct model optimism and quantify calibration accuracy. We evaluated responsiveness of the index by assessing longitudinal validity and estimating standardized effect sizes.

Results: We developed the magnetic resonance novel index for fistula imaging in CD (MAGNIFI-CD) using 6 items. The optimism-corrected R2 of the model was 0.71, which was comparable to R2 for the original sample (0.74). The calibration slope for the model was 0.98. Compared with the original and modified versions of the Van Assche index, the MAGNIFI-CD had improved operating characteristics. Estimates of intraclass correlation coefficients for MAGNIFI-CD, the modified Van Assche index, and Van Assche index were 0.85 (95% CI, 0.77–0.90), 0.81 (95% CI, 0.74–0.86), and 0.81 (95% CI, 0.71–0.86) for intra-rater reliability, and 0.74 (95% CI,0.63–0.80), 0.67 (95% CI, 0.55–0.75), and 0.68 (95% CI, 0.56–0.77) for inter-rater reliability. Corresponding standardized effect size estimates were 1.02 (95% CI, 0.65–1.39), 0.84 (95% CI, 0.48–1.21), and 0.68 (95% CI, 0.33–1.03).

Conclusions: We developed an index, called the MAGNIFI-CD, based on 6 items. It assesses MRI data and determines perianal fistulizing CD activity with improved operating characteristics compared to previous indices. This index may be used as an outcome measure in clinical trials comparing treatment effects in patients with perianal fistulizing CD. Although the performance of the MAGNIFI-CD indicates its stability and reasonable external validity, external validation is needed.

KEY WORDS: ICCs, VAS, healing, remission

INTRODUCTION

Perianal fistulas affect approximately one-third of patients with Crohn's disease (CD) during their lifetime and are associated with significant morbidity and impaired quality of life.[1, 2] A multidisciplinary initiative of clinicians, patients and patient-support organizations in the United Kingdom identified perianal CD as one of the top ten research priorities for the inflammatory bowel diseases.[3] Despite progress in drug development for luminal CD, few randomized controlled trials (RCTs) have specifically addressed fistula healing. A meta-analysis of medical therapies for fistulizing CD published in 2018 found moderate quality evidence from RCTs to support the efficacy of tumor necrosis factor antagonists, ustekinumab, and mesenchymal stem cell therapy for induction of fistula remission.[4] Interest in drug development for treatment of fistulizing perianal CD has been recently bolstered by promising outcomes observed in a clinical trial of stem cell therapy,[5] and further studies of this modality are underway (clinicaltrials.gov NCT03279081).

Although magnetic resonance imaging (MRI) is considered the gold standard for assessment of perianal fistulas,[6] the lack of a validated outcome measure to assess disease activity and response to therapy is a barrier to efficient drug development for perianal fistulizing CD. The Van Assche Index (VAI) is the most frequently used MRI index for assessment of perianal fistulas,[7] however this instrument was developed without using standard procedures for item identification and selection, and evaluation of reliability and responsiveness. In addition, the VAI has been only partially validated.[8,9] We previously developed standardized scoring conventions for the existing component items of the VAI through a formal expert consensus process, assessed

these and novel items proposed during the consensus process, and ultimately developed a 5-item (extension, hyperintensity on T2-weighted images, rectal wall involvement, inflammatory mass, dominant feature of primary tract and extension) modified version of the VAI (mVAI) that had numerically higher intraclass correlation coefficients (ICCs) for intra- and inter-rater reliability compared to the VAI.[9] In addition, further refinements to items considered during the development of the mVAI were proposed during a second post-reliability expert consensus meeting that included recommendations to 1) definitively exclude assessment of the presence of an ano/rectovaginal tract, 2) re-categorize number of fistulas as "single, unbranched" and "complex," with complex fistulas including both single, branched and multiple tracts, 3) replace rectal wall involvement with "presence or absence of proctitis," defined as increased wall thickness and size of mesorectal lymph nodes, creeping fat, and increased perimural T2 signal and enhancement,[10] and 4) modify the description of hyperintensity of primary tract and extensions on post-contrast fat saturated T1-weighted images to rate the most severe lesion by comparing signal intensity with nearby, in-plane vessels.[9]

We developed a new index using 160 MRI scans collected during the conduct of a phase 3, randomized placebo-controlled clinical trial of adipose derived stem cell therapy for fistulizing perianal CD.[5] We first re-assessed the reliability of the component items refined during the development of the mVAI, and the reliability of exploratory items proposed by the current study radiologists. Items with acceptable reliability were used to develop a refined MRI index, MAGNIFI-CD. The operating properties of MAGNIFI-CD were then explored and compared to the mVAI and the VAI.

METHODS

Study population

We used MRI sequences and necessary clinical covariates from the ADMIRE CD study, a phase 3 randomized placebo-controlled trial of a local injectable stem cell therapy (darvadstrocel [formerly Cx601]) for complex perianal fistulas in patients with CD.[5] The inclusion criteria, patient characteristics/demographics, and clinical data from the ADMIRE CD trial have been previously described and no post-hoc analyses related either to efficacy or safety are part of this study. All MRI sequences used in the present study were prospectively re-read for the purposes of this study by four radiologists who were not involved in the original RCT and were masked to clinical information, treatment assignment and timepoint.

The original trial[5] randomly allocated 107 patients to active treatment consisting of a single injection of 120 million cells distributed into the tissue adjacent to all fistula tracts and internal openings, and 105 patients to control treatment consisting of a single injection of saline solution. The present study used scans (independent of treatment assignment in ADMIRE CD) from 160 patients for whom both baseline and Week 24 scans were available.

Ethical considerations

All patient information used in this study was de-identified with respect to the originating study, patient identification number and investigational site. The informed consent obtained during the original clinical trial complied with International Conference on Harmonization Good Clinical Practice and all applicable regulatory requirements. Approval to re-read the MRI scans independently for the purposes of this study was

granted by the Western University Health Science Research Ethics Board (file number 108869).

Pelvic MRI

Pelvic MRI scans in the ADMIRE CD study were acquired with either 1.5-T or 3.0-T scanners (depending on site facilities), according to a standardized protocol for perianal fistulizing disease (**Supplementary Table 1**) and included T2-weighted sequences in three planes of space with fat saturation in the axial plane, and T1-weighted sequences with fat saturation. Post-contrast sequences in the axial plane were performed after IV administration of 0.1 mL/kg gadolinium contrast agent. As the imaging parameters and pulse sequences listed in **Supplementary Table 1** may be field strength dependent, appropriate pulse parameters were included to adjust for these potential differences. Use of the same equipment and parameters was requested for all study sequences and images

Reading of the MRI sequences

All MRI scans were reviewed for consistent image quality, type, and use of contrast prior to assessment of disease activity by four expert abdominal radiologists (ST, CS, JR, BM), selected for specific expertise in the interpretation of MR images in fistulizing CD, trained on scoring of the original and mVAI, and blinded to clinical information and study time point. Items assessed included those generated during the development of the mVAI[9] and exploratory items proposed by the study radiologists (number of external and internal openings and length of fistula tract) (**Table 1**), as well as a global measure of perianal fistulizing disease activity (100 mm visual analogue scale [VAS]; where 0 mm = no disease and 100 mm = worst disease encountered). Scans that lacked

sufficient overall quality, or those that had missing VAS or index items scores were excluded from further analysis.

Datasets used for assessment of reliability, responsiveness and index development

For assessment of index reliability, each of the four central readers independently evaluated disease activity in Week 24 scans (N=40) that were randomly selected using computer-generated scan numbers from patients in both treatment groups, twice, in random (computer-generated) order on separate occasions, in the absence of any clinical information. Week 24 images were used for assessing reliability, since they are more likely to represent a spectrum of potential disease activity.

All paired (baseline and Week 24) scans with adequate quality were used to evaluate responsiveness. Each of the four radiologists read approximately one-fourth of the paired images (range 35–42 pairs). In the case of Week 24 images that had already been read twice by all four readers for reliability, a single read was chosen using a computer-generated randomly selected reader and using the first of the two reliability reads.

Index development was based on Week 24 scans with adequate quality using the selection process described above.

Statistical methods

Scoring of the VAI and mVAI

The original VAI items were not directly assessed but were obtained by recoding the mVAI items (**Table 1**) as follows: 1) number of fistula tracts (VAI; single [unbranched or branched], multiple): number of external and internal openings was determined when a

complex tract (mVAI) was identified by the central reader. The presence of one external opening and more than one internal opening was interpreted as a "single, branched" tract for calculation of the VAI, whereas the presence of two or more external openings was interpreted as "multiple" tracts; 2) location (VAI; extra- or intersphincteric, transsphincteric, suprasphincteric): a submucosal location on the mVAI was assigned a score of zero, extrassphincteric and intersphincteric locations were each assigned a score of 1 and a suprasphincteric location was assigned a score of 3 for calculation of the VAI; 3) extension (VAI; infralevatoric, supralevatoric): a horseshoe configuration on the mVAI was not considered an extension; 4) collections (VAI; cavities > 3 mm): small, medium, and large collections identified for the mVAI item "inflammatory mass" were interpreted as collections "present" for the VAI; 5) proctitis (VAI; rectal wall involvement [normal, thickened]): the presence of "proctitis" identified for the mVAI (as recommended by the RAND experts who participated in the study described by Samaan *et al;*[9] scored as 0 when absent and 2 for present) was interpreted as "thickened" for the "rectal wall involvement" item and assigned a score of 2 for calculation of the VAI.

Assessment of reliability

As the data were either ordinal or continuous, reliability was quantified using intraclass correlation coefficients (ICC), which are equivalent to quadratically weighted kappa for ordinal data. Point estimates for ICCs were obtained using a 2-way random effects analysis of variance model with interaction for the MR indices, each index component, and the 100 mm VAS global measure of perianal fistulizing disease activity. The advantage of this model is that it can simultaneously estimate intra- and inter-rater reliability.[11] Specifically, the model contains subject, rater and their interaction terms.

The inter-rater ICC is then defined as the proportion of total variability in observed measurements accounted for by the subject-to-subject variability, and the intra-rater ICC is defined as the correlation between two randomly selected measurements on the same subject for a randomly selected rater. To avoid the assumption that the data are normally distributed and to account for the fact that the observations within an image are not independent, we used the cluster nonparametric percentile bootstrap method with 2000 replicates with replacement at the level of image to obtain the associated two-sided 95% confidence intervals (CI) for intra- and inter-rater ICCs.[12] The degree of reliability was interpreted according to the benchmarks proposed by Landis and Koch whereby ICCs of <0.0, 0.0–0.20, 0.21–0.4, 0.41–0.6, 0.61-0.8, and >0.81 constitute 'poor,' 'slight,' 'fair,' 'moderate,' 'substantial' and 'almost perfect' reliability.[13] It should be noted, however, that other benchmarks have also been proposed for the interpretation of ICC.[14]

Development of the novel index

Items from the mVAI and exploratory items identified in the present study were considered as candidate items for the new index. Fistula location was not considered based on expert opinion of the study radiologists that it is unlikely to change with effective treatment. For index development, items with at least a 'moderate' level of reliability (i.e. ICC ≥0.40) were selected as candidate items and the VAS global measure of perianal fistulizing disease activity was used as the anchor. Exploratory bivariate analyses between the VAS score and each of the selected items were performed first to guide the coding of each item. Specifically, we prespecified that items would be coded as continuous if a linear relationship was demonstrated between a

change in the item score and a change in the VAS score. If a linear relationship was not evident, the bivariate relationships were used to collapse item levels.

A full model was then obtained that incorporated all items, re-scaled where deemed necessary, followed by a step-down model building approach with P = 0.05 used as the criterion for item retention. Residuals from the final model were subjected to diagnostics examination (see **Supplementary Figure 1**), which suggests no obvious violations of the key assumptions for linear regression analysis, i.e., homogeneity of residual variance, normality of residual distribution, and the lack of outliers. The stability of the final model was assessed and calibrated using the .632 bootstrap method with 2000 replicates.[15] Briefly, we obtained the optimism-corrected performance (e.g., R^2 and calibration intercept and slope) as the difference between the apparent performance in the study sample and the average optimism, where the average optimism was obtained using the bootstrap method with 2000 replications. Within each replication, the bootstrap sample was used to develop a model using backward elimination, with apparent performance estimated. This model was then tested on data points in the original sample that were not included in the bootstrap sample to obtain the test performance. The optimism for each replication was obtained by the difference between the apparent performance and test performance. This process was repeated 2000 times to obtain the average optimism. On average, 63.2% of the subjects were used to build the model, which was then tested on the remaining 36.8% of the subjects.

The coefficients from the final model were standardized by dividing by the smallest coefficient and then rounded to allow simple calculation of the new index. The

operating characteristics of the new index were evaluated in terms of reliability and responsiveness using methods described elsewhere in this section.

Evaluation of responsiveness

Responsiveness of the MR indices was evaluated from two perspectives as suggested by the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist.[16] We first evaluated longitudinal validity[17] by estimating correlation coefficients and their associated 95% CIs between change scores for the MR indices, the VAS and other measures of disease activity including the perianal disease activity index (PDAI), the Crohn's Disease Activity Index (CDAI), the Inflammatory Bowel Disease Questionnaire (IBDQ), and C-reactive protein (CRP) concentrations. We then evaluated responsiveness by quantifying the ability of the indices to detect a meaningful change, defined as an improvement in the VAS of one-half of the baseline standard deviation (SD).[18] Results were reported in terms of the standardized effect size (mean difference divided by standard deviation) and the associated 95% CI. Interpretation of correlations and standardized effect sizes was done according to benchmarks set by Cohen.[19] Standardized effect sizes were compared using methods for areas under correlated receiver operating curves (ROC), as it can be shown that the area under the ROC curve is a one-to-one function of standardized effect size[20] when data are normal.[21]

Justification of sample size

Sample size calculation for reliability was based on the 1-way random effects model,[22] which tends to provide more conservative estimates compared to those based on 2-way models. Assuming a true ICC of 0.7, evaluation of 40 MRI scans 2 times by 4 central

readers would yield an approximately 80% chance of obtaining a lower bound for the 2-sided 95% CI for an ICC greater than 0.5. We did not do formal sample estimation for index development as there is no simple procedure for this purpose, However, using the rule of thumb of 10 observations per predictor, this study had sufficient sample size for 16 predictors. Sample size justification for responsiveness was based on estimation of standardized effect size. Specifically, assuming 30% of patients experienced meaningful change scores, a sample size of 160 would result in an 80% chance that the lower limit of the 95% CI would exclude 0.

Statistical analyses were conducted with SAS (version 9.4) and rms package in R (version 3.5.2) software.[23]

RESULTS

MRI sequences

Of the 212 patients randomized in the ADMIRE CD trial, 186 had MRIs available at both baseline and Week 24. For the purpose of this post hoc study which required rescoring of all paired images, 11 pairs were excluded due to image quality (baseline only [n=2], Week 24 only [n=6], or both time points [n=3]), and 15 pairs were excluded due to missing VAS global perianal fistulizing disease activity or index item scores, leaving a total of 160 patients (86%) with MRI pairs available for analysis. Baseline clinical characteristics of the patients whose scans were read and used for subsequent analyses described below are provided in **Supplementary Tables 2 and 3.**

Assessment of reliability

Intraclass correlation coefficients (95% CI) for intra- and inter-rater reliability for the VAI, mVAI, the exploratory items, and the VAS are shown in **Table 2** (intra-rater reliability for

each reader is shown in **Supplementary Table 4**). For the VAI and mVAI, intra-rater reliability was almost perfect (0.81 [0.71–0.86] and 0.81 [0.74–0.86], respectively) and inter-rater reliability was substantial (0.68 [0.56–0.77] and 0.67 [0.55–0.75], respectively). Almost perfect intra- and inter-rater reliability was observed for the VAS-based assessment (0.89 [0.83–0.92] and 0.82 [0.75–0.86], respectively).

All component items considered in the development of the mVAI had moderate to substantial intra-rater reliability and moderate inter-rater reliability. This was also true for the original VAI items, with the exception of "number of fistula tracts" and "extension," for which only moderate intra-rater and fair inter-rater reliability were observed.

For the exploratory items, intra-rater reliability was substantial, and inter-rater reliability was moderate for "length of fistula tract," whereas assessment of the "number of external openings" and "number of internal openings" were associated with moderate intra-rater and fair inter-rater reliability.

New index development for perianal fistulizing Crohn's disease: MAGNIFI-CD

The mean (SD) Week 24 VAS global perianal fistulizing disease activity score was 42.4 (21.6) and the distribution of the VAS scores for the 160 patients whose images were used for index development is shown in Figure 1a. Eight candidate items for the new index had inter-rater ICC estimates ≥ 0.4, including number of fistula tracts, hyperintensity of primary tract on T2-weighted images, hyperintensity of primary tract on post-contrast T1-weighted images, proctitis, dominant feature of primary tract and extensions, length of fistula tract, fistula extension and inflammatory mass.

Coding of each item in the multivariable linear regression model was guided by bivariate relationships between scores on the VAS and each item (Figure 1b; the

corresponding frequency distribution of investigated items is shown in **Supplementary Table 5**). Specifically, there were linear relationships between increments in VAS scores and the number of fistula tracts, hyperintensity of primary tract on T2-weighted images, and length of the fistula tract. For hyperintensity of the primary tract on post-contrast T1-weighted images both scores of 0 and 1 had VAS scores of approximately 20, justifying the collapsing of this item from three original categories (0 = absent, 1 = mild, 3 = pronounced) to two categories (0 = absent/mild and 1 = pronounced) for model development. For the extension item (4 original categories; 0 = absent, 1 = infralevatoric, 2 = horseshoe configuration, 3 = supralevatoric), categories 1 and 3 were collapsed, leaving 3 categories for this item (0 = absent, 1 = horseshoe configuration, 2 = infralevatoric/supralevatoric). For the inflammatory mass item, the original categories of focal and diffuse were switched (such that 0 = absent, 1 = focal, 2 = diffuse, 3 = small collections, 4 = medium collections, 5 = large collections). Switching of the categories for extension and inflammatory mass were discussed with clinical experts. The bivariate relationships between the VAS score and each item after the re-scaling process are shown in Figure 1c (the corresponding frequency distribution and univariable linear regression models of the investigated items after re-scaling are shown in **Supplementary Tables 6 and 7**).

A backwards step-down procedure with the .632 bootstrap method of 2000 samples removed the items hyperintensity of primary tract on T2- weighted images and proctitis from the initial model, leaving 6 items in the final model for predicting the VAS **(Table 3)**. The R^2 was minimally changed from the full model that included 8 items to the final model that included 6 items. Residual diagnostic plots did not indicate severe

violation of model assumptions (results not shown). The optimism-corrected R^2 (0.71) was comparable with that based on the original sample (0.74). Furthermore, the optimism-corrected calibration slope was 0.98, and close to an ideal value of 1.0. We therefore conclude that the final model with 6 independent variables is very stable (Figure 1d), suggesting the final model has reasonable external validity.

Using the standardized coefficients from Table 3, MAGNIFI-CD can be calculated as:

MAGNIFI-CD =
- 3 x number of fistula tracts (3 levels)
- 2 x hyperintensity of primary tract on post-contrast T1-weighted images (2 levels)
- 2 x Dominant feature (3 levels)
- 2 x Fistula length (3 levels)
- 2 x Extension (3 levels)
- 1 x Inflammatory mass (6 levels)

The total MAGNIFI-CD score ranges from 0 (no disease activity) to 25 (severe disease activity). The mean (SD) MAGNIFI-CD score based on the 160 patients used for the development of the new index was 13.6 (5.6). Intra-class correlation coefficients (95% CI) for the new index scores were consistent with nearly perfect intra-rater (0.85, 0.77–0.90) and substantial inter-rater reliability (0.74, 0.63–0.80).

Assessment of responsiveness

Longitudinal validity

Correlations among the changes in the MRI index scores and those of the PDAI, CDAI, IBDQ, and CRP concentrations irrespective of treatment assignment are shown in **Table 4**. Changes in the MRI index scores were highly correlated with each other and the VAS (correlation coefficients ≥ 0.60). However, the correlations among changes in

the other outcomes and changes in the MRI index scores were small (all correlation coefficients ≤ 0.20).

Standardized effect size estimates

When meaningful change was defined as a decrease in the VAS of one-half of the baseline SD, 56/160 patients (35.0%) were considered improved or changed, 77/160 (48.1%) were considered unchanged while 27/160 (16.9%) worsened. The standardized effect size (95% CI), according to this definition of change was 0.68 (0.32 to 1.03) for the original VAI, 0.84 (0.48–1.21) for the mVAI, and 1.02 (0.65–1.39) for MAGNIFI-CD. Although the difference in standardized effect size between MAGNIFI-CD and the mVAI was not statistically significant (P = 0.29), the standardized effect size for MAGNIFI-CD was statistically superior to the standardized effect size for the VAI (P = 0.045).

For patients whose VAS changed more than one-half of the SD, the mean (SD) MAGNIFI-CD scores at baseline and Week 24 were 15.3 (4.9) and 11.4 (5.9), respectively. The values for patients whose VAS changed less than one-half of the SD were 14.8 (5.4) and 14.4 (5.2), respectively.

DISCUSSION

Identification of new treatments for fistulizing perianal CD is an unmet medical need. An important barrier to progress in this area is the lack of a well-validated MRI disease activity index. In previous studies we developed standardized scoring conventions for the original items of the VAI and assessed their reliability in a convenience sample of MR images.[9] Modifications were made to improve item reliability and new candidate items were identified through a formal consensus process. A modified VAI was created using a mixed model approach, and final recommendations

for further improvement were suggested for future evaluation. In the present study, we retested the reliability of all of the items assessed during the derivation of the mVAI, additional exploratory items, and, furthermore, evaluated the reliability and responsiveness of the resulting indices using MRIs from a randomized, placebo-controlled trial of a stem cell therapy of known efficacy.[5] Both the VAI and the mVAI were associated with almost perfect intra-rater, and substantial inter-rater reliability, and ICCs that were nearly identical to those observed in the derivation cohort.[9] These observations in an independent dataset confirm the validity of our original findings.

Items of sufficient reliability were considered candidate items for incorporation into a new MRI activity index. Using the VAS as the dependent variable, application of a backward step-down regression procedure yielded a new index consisting of 6 items with a total score ranging from 0–25. Three items that met the criterion for inclusion in further index development based on their inter-rater reliability (ICC ≥0.4; **Table 2**), were not included in the final model. Fistula location was not included as a candidate index item at the recommendation of the study radiologists. This item will not change following administration of an effective therapy and fibrous tracts may persist despite lesion healing. Both proctitis and hyperintensity of T2-weighted images were eliminated during the backwards step-down procedure. Although proctitis has prognostic implications for the likelihood of response to anti-inflammatory therapy,[24] expert opinion is that it has little value for defining the severity of fistulas.

The final index has improved inter-rater reliability and responsiveness compared to both the VAI and the mVAI. MAGNIFI-CD incorporates clinically important items relevant to the burden of perianal disease inflammation, such as number and the length

of fistulas and, in contrast to the mVAI, does not include assessment of “proctitis,” a finding that is not directly related to the severity of the disease. Thus MAGNIFI-CD is likely to be a more efficient measure of disease activity that can be used to detect early efficacy signals and dose-response relationships in early drug development, a setting where use of the binary endpoint of fistula “remission” is relatively statistically inefficient. However, further work is needed to evaluate this possibility, as well as the operating properties of the new index in order to determine cut-points for minimal clinically important change.

Of note, we did not find meaningful correlation between changes in the MR indices and changes in other clinical measures of disease activity. The lack of correlation with the CDAI is not surprising given that luminal activity is distinct from fistula activity. Indeed, the patient population in the ADMIRE CD study had non-active or mildly active luminal CD (CDAI ≤220) despite the presence of complex perianal fistulas. Furthermore, although the CDAI items “stool frequency” and “abdominal pain” add significant weight to the total index score, they are not likely to influence the pathological process in fistulizing CD. Although the PDAI was developed to measure clinical disease activity in patients with fistulizing disease, it was not developed using methodology that conformed to accepted principles of evaluative index development. In this regard, the lack of correlation between changes in the MR indices and the PDAI is analogous to the poor correlation that exists between endoscopic disease activity and the CDAI. In the latter case, a clinical (symptom and sign-based) measure of disease activity has poor correlation with the pathological process of inflammation and is therefore a relatively poor measure of treatment efficacy. This observation has important implications for drug

development and selection of clinical trial outcomes, as well as for regulatory assessment of new drugs.

Our study had several strengths. This work is the result of an international collaboration between a team of experts in gastrointestinal radiology, inflammatory bowel disease specialists, and biostatistics. Through collaboration with industry partners, we were able to utilize MR images from the largest randomized study to date of an effective therapy for CD patients with perianal fistulas.

Our study has some limitations. First, the dataset excluded patients with simple fistulas, rectovaginal fistulas, rectal/anal stenosis and/or severe proctitis (defined as the presence of ulcerations). These restrictions may limit the generalizability of the index in these subgroups. Second, the participating radiologists were highly experienced in pelvic MRI and intensively trained on item scoring conventions, thus our results may not be generalizable to other settings. Third, the new index was not independently evaluated in a second data set. However, to our knowledge there is no other known placebo-controlled dataset from a RCT of an effective therapy for perianal fistulizing CD that has incorporated MRI assessments. Fourth, the images were limited specifically to an intervention of stem cell therapy injection, and thus the operating properties of the new index should also be assessed in studies of other treatment types. Despite these limitations we believe this work addressed an important unmet need in perianal CD for a validated MRI index for the assessment of perianal fistula activity. Use of MAGNIFI-CD in future clinical trials will simultaneously facilitate research and development of novel therapies for a patient group with large unmet need and burden of illness, while providing additional datasets for external validation of the index.

REFERENCES

1. Schwartz DA, Loftus EV, Jr., Tremaine WJ, et al. The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. Gastroenterology 2002;122:875–880.
2. Mahadev S, Young JM, Selby W, et al. Quality of life in perianal Crohn's disease: what do patients consider important? Dis Colon Rectum 2011;54:579–585.
3. Hart AL, Lomer M, Verjee A, et al. What are the top 10 research questions in the treatment of inflammatory bowel disease? A priority setting partnership with the James Lind Alliance. J Crohns Colitis 2017;11:204–211.
4. Lee MJ, Parker CE, Taylor SR, et al. Efficacy of medical therapies for fistulizing Crohn's disease: Systematic review and meta-analysis. Clin Gastroenterol Hepatol 2018; doi: 10.1016/j.cgh.2018.01.030. [Epub ahead of print].
5. Panes J, Garcia-Olmo D, Van Assche G, et al. Expanded allogeneic adipose-derived mesenchymal stem cells (Cx601) for complex perianal fistulas in Crohn's disease: a phase 3 randomised, double-blind controlled trial. Lancet 2016;388:1281–1290.
6. Gecse KB, Sebastian S, Hertogh G, et al. Results of the fifth scientific workshop of the ECCO [II]: Clinical aspects of perianal fistulising Crohn's disease-the unmet needs. J Crohns Colitis 2016;10:758–765.

7. Van Assche G, Vanbeckevoort D, Bielen D, et al. Magnetic resonance imaging of the effects of infliximab on perianal fistulizing Crohn's disease. Am J Gastroenterol 2003;98:332–339.

8. Horsthuis K, Ziech ML, Bipat S, et al. Evaluation of an MRI-based score of disease activity in perianal fistulizing Crohn's disease. Clin Imaging 2011;35:360–365.

9. Samaan MA, Puylaert CAJ, Levesque BG, et al. The development of a magnetic resonance imaging index for fistulising Crohn's disease. Aliment Pharmacol Ther 2017;46:516–528.

10. Tutein Nolthenius CJ, Bipat S, Mearadji B, et al. MRI characteristics of proctitis in Crohn's disease on perianal MRI. Abdom Radiol (NY) 2016t;41:1918–1930.

11. Eliasziw M, Young SL, Woodbury MG, et al. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Phy Ther 1994;74:777–788.

12. Field C, Welsh AH. Bootstrapping clustered data. J Roy Statist Soc Ser B Methodological 2007;69:369–390.

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

14. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–290.

15. Efron B and R Tibshirani. Improvements on cross-validation: The .632+ bootstrap method. J Am Stat Assoc 1997; 92: 548–560.

16. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–745.

17. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–468.

18. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–592.

19. Cohen J. A power primer. Psychol Bull 1992;112:155–159.

20. Delong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–845.

21. Zou GY and Yue L. Using confidence intervals to compare several correlated areas under the receiver operating characteristic curves. Stat Med 2013; 32:5077–5090.

22. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012;31:3972–3981.

23. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

24. Panés J and Rimola J. Perianal fistulizing Crohn's disease: pathogenesis, diagnosis and therapy. Nat Rev Gastroenterol Hepatol 2017;14:652–664.

FIGURE LEGENDS

Figure 1a. Distribution of the VAS scores for the 160 patients used in the development of the MAGNIFI-CD.

Figure 1b Univariable summaries of VAS scores as stratified by levels of candidate items. The figure shows the VAS scores for the items evaluated according to each of their levels. These were used to guide the number of levels for regression analysis. For example, a linear relationship is present for the number of fistula tracts, so three levels were included whereas only two levels were appropriate for hyperintensity of the primary tract on post-contrast T1-weighted images.

Figure 1c Univariable summaries of VAS scores as stratified by re-scaled levels of candidate items. The figure shows the VAS scores for the items evaluated according to each of their levels after re-scaling. For example, hyperintensity of the primary tract on post-contrast T1-weighted images was collapsed into two levels by combining the first two levels

Figure 1d. Calibration plot of actual versus predicted VAS using the final model with six variables (number of fistula tracts, hyperintensity of primary tract on post-contrast T1-weighted images, dominant feature, fistula length, extension, and inflammatory mass). The 45^{o} line shows perfect (Ideal) prediction. The model performance as assessed by the derivation sample is shown by the dotted line (Apparent). The model performance as assessed by bootstrap validation with 2000 replications is shown by the dashed line (Bias-corrected). The closeness between the Apparent and Bias-corrected plots suggests stability for model performance in data sets other than that used to derive the model.

Table 1. MRI items assessed for reliability

Item	Category	Definition
Number of fistula tracts	None	No tracts visible
	Single, unbranched	Single internal opening leading to a single fistula tract (internal opening defined as discontinuation of anal mucosa or closest proximity of tract to anal mucosa)
	Complex	Either a single internal opening leading to more than one fistula tract or multiple internal openings
Location (scored for the most severe/predominant fistula tract)	Submucosal	Tract lies superficial to the internal sphincter
	Intersphincteric	Tract extends through the internal sphincter to the intersphincteric plane then to the perineal skin
	Transsphincteric	Tract extends via the internal and external anal sphincter (or puborectal muscle) into the ischiorectal fossa then to the perineal skin

	Extrasphincteric	Tract extends through the ischiorectal fossa upwards and through the levator ani muscles to the rectal wall completely outside the sphincter mechanism
	Suprasphincteric	Tract extends via intersphincteric space, then tracts superiorly to above the puborectalis muscle (i.e., above the anorectal junction) before curving downward through the levator muscle lateral to the external anal sphincter and puborectal muscle into the ischiorectal fossa then to the perineal skin
Extension (score the most severe)	Absent	No extension
	Infralevatoric	Extends upward in the ischioanal fossa but remains below the levator ani muscle
	Horseshoe configuration	Extends into the intersphincteric space on both sides of the midline
	Supralevatoric	Any extension in the supralevatoric space (i.e. above where the levator plate is connected to the

		anorectum)
Hyperintensity of primary tract or extensions on fat saturated T2-weighted images (rate the most severe lesion by comparing signal intensity with nearby, in plane vessels)	Absent	No hyperintensity visible, only scar tissue
	Mild	Slight increase in signal intensity but less than nearby, in-plane vessels
	Pronounced	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels

Hyperintensity of primary tract or extensions on post-contrast fat saturated T1-weighted images (rate the most severe lesion by comparing signal intensity with nearby, in-plane vessels)	Absent	No hyperintensity visible
	Mild	Slight increase in signal intensity but less than nearby, in-plane vessels
	Pronounced	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels
Presence of proctitis	Absent	Normal appearance of rectal wall
	Present	Increased wall thickness and size of mesorectal lymph nodes (>5 mm), creeping fat, increased perimural T2 signal and enhancement.

Inflammatory mass	Absent	No inflammatory mass
	Diffuse	Diffuse inflammation of surrounding tissues
	Focal	Lesion > 3 mm in diameter on T2-weighted images (but does not include linear tracts with diameter > 3mm) with diffuse enhancement on T1-weighted post contrast images (i.e., granulation tissue)
	Small collection	Circumscribed cavity 3-10 mm in diameter (but does not include linear tracts with diameter >3 mm). Hyperintense appearance on fat saturated T2-weighted images with rim enhancement on T1-weighted post-contrast images
	Medium collection	As defined above except diameter measures 11-20 mm
	Large collection	As defined above except diameter measures >20 mm

Dominant feature of primary tract and extensions	Predominantly fibrous	> 50% of tract has a fibrotic appearance (i.e., hypointense on fat saturated T2-weighted images)
	Predominantly filled with granulation tissue	> 50% of tract is filled with granulation tissue (i.e., hyperintense on fat saturated T2-weighted images with enhancement of contents and wall on T1-weighted post-contrast images)
	Predominantly filled with fluid or pus	> 50% of tract is filled with fluid or pus (i.e., hyperintense on fat saturated T2-weighted images with no enhancement of contents on fat saturated post-contrast T1-weighted images [though lining of tract may enhance])
Internal openings[1]	0	
	1	
	2	
	> 2	

External openings[1]	0
	1
	2
	3
	> 3
Length of fistula tract[1]	< 2.5 cm
	2.5 cm to 5 cm
	> 5 cm

[1]Exploratory item proposed by the study radiologists.

Table 2. Variance components and estimates of intraclass correlation coefficients (ICC) for the original and modified van Assche indices and the VAS based on assessment of 40 scans (each read twice by 4 readers who were blinded to patient clinical information).

	Variance components					Intraclass correlation coefficients (95% bootstrap CI)	
	Slide	**Reader**	**Slide X Reader**	**Residual**	**Total**	Intra-rater	Inter-rater
Modified van Assche index							
Extension	0.553	0.030	0.202	0.446	1.211	0.64 (0.48–0.77)	0.45 (0.31–0.58)
Hyperintensity of primary tract on T2-weighted images	0.205	0.000	0.066	0.125	0.396	0.68 (0.51–0.79)	0.52 (0.34–0.64)
Proctitis	0.068	0.004	0.026	0.047	0.145	0.68 (0.45–0.84)	0.47 (0.13–0.69)
Inflammatory mass	0.684	0.001	0.520	0.317	0.906	0.79 (0.58–0.89)	0.45 (0.20–0.62)
Dominant feature	0.179	0.010	0.069	0.081	0.339	0.76 (0.63–0.84)	0.53 (0.36–0.65)
Total mVAI	10.860	0.108	2.111	3.044	16.153	0.81 (0.74–0.86)	0.67 (0.55–0.75)
Other items explored during derivation of the modified van Assche index							
Number of fistula tracts	0.115	0.032	0.020	0.118	0.285	0.59 (0.48–0.69)	0.40 (0.27–0.52)
Location (modified definition)	0.363	0.041	0.190	0.186	0.780	0.76 (0.67–0.83)	0.47 (0.32–0.58)
Hyperintensity of primary tract on post-contrast T1-weighted images	0.145	0.004	0.034	0.115	0.298	0.61 (0.47–0.73)	0.49 (0.34–0.61)

	Variance components					Intraclass correlation coefficients (95% bootstrap CI)	
	Slide	**Reader**	**Slide X Reader**	**Residual**	**Total**	Intra-rater	Inter-rater
Original VAI[1]							
Number of fistula tracts	0.241	0.047	0.110	0.290	0.688	0.58 (0.45–0.70)	0.35 (0.22–0.48)
Location	0.196	0.012	0.077	0.136	0.421	0.68 (0.56–0.77)	0.47 (0.34–0.56)
Extension	0.216	0.038	0.061	0.241	0.556	0.57 (0.39–0.71)	0.39 (0.26–0.51)
Hyperintensity of primary tract on T2-weighted images	0.205	0.000	0.066	0.125	0.396	0.68 (0.51–0.79)	0.52 (0.34–0.64)
Collections	0.053	0.001	0.036	0.007	0.097	0.93 (0.83–1.00)	0.55 (0.05–0.79)
Rectal wall involvement	0.068	0.004	0.026	0.047	0.145	0.68 (0.45–0.84)	0.47 (0.13–0.69)
Total VAI	12.63	0.163	2.106	3.541	18.440	0.81 (0.71–0.86)	0.68 (0.56–0.77)
Exploratory items							
Length of fistula tract	0.275	0.000	0.033	0.145	0.453	0.68 (0.54–0.78)	0.61 (0.45–0.71)
Number of external openings	0.1251	0.005	0.039	0.177	0.346	0.49 (0.35–0.61)	0.36 (0.23–0.49)
Number of internal openings	0.064	0.008	0.059	0.161	0.292	0.45 (0.29–0.58)	0.22 (0.07–0.35)
Visual Analogue Scale (VAS)	406.9	13.90	22.16	56.07	499.03	0.89 (0.83–0.92)	0.82 (0.75–0.86)

[1]Re-coding of the mVAI items was used to derive ICC estimates for the original VAI items as described in the Methods.

Table 3. Items and weights for the candidate index MAGNIFI-CD (developed based on 160 Week 24 scans read once by 4 readers [each reading 35 to 42 scans])

Item	Full model Coefficient (SE)	Final Model Coefficient (SE)	Standardized Coefficient[1]
Intercept	-4.84(3.51)	-3.22 (3.16)	
Number of fistula tracts 0 = None 1 = Single, unbranched 2 = Complex	8.80 (2.16)	9.48 (2.07)	3
Hyperintensity of primary tract on post-contrast T2-weighted images Absent Mild Pronounced	2.62 (2.51)	-----	
Hyperintensity of primary tract on post-contrast T1-weighted images 0 = Absent/Mild 1 = Pronounced	5.79 (2.93)	7.57 (2.31)	2
Dominant feature[2] 0 = Predominantly fibrous 1 = Predominantly granulation tissue 2 = Predominantly fluid/pus	7.23 (1.91)	7.80 (1.84)	2
Proctitis	1.07 (2.09)	-----	
Fistula length 0 = < 2.5 cm 1 = 2.5 cm to 5 cm 2 = > 5 cm	6.69 (1.47)	6.72 (1.46)	2
Extension 0 = Absent 1 = Horseshoe 2 = Infralevatoric/Supralevatoric	6.20 (1.22)	6.18 (1.20)	2
Inflammatory mass 0 = Absent 1 = Focal 2 = Diffuse 3 = Collections-Small 4 = Collections-Medium 5 = Collections-Large	3.12 (0.71)	3.15 (0.70)	1
R^2	0.744	0.742	

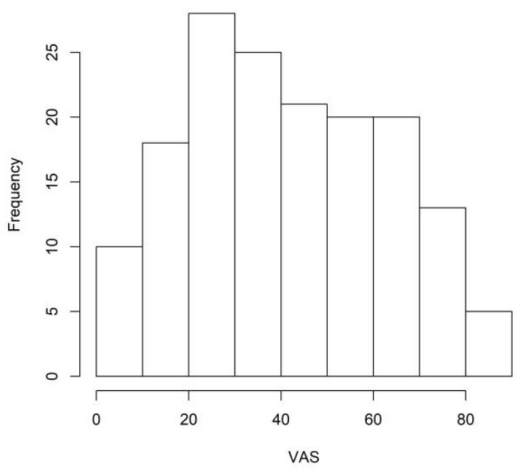
SE, standard error.
[1]Calculated by dividing the regression coefficient for each item by the smallest coefficient and rounding.
[2]If multiple fistula are present, the worst is rated.

Table 4. Estimates of the Pearson correlation coefficient (95% CI) amongst changes in MRI indices and other measures of disease activity[1]

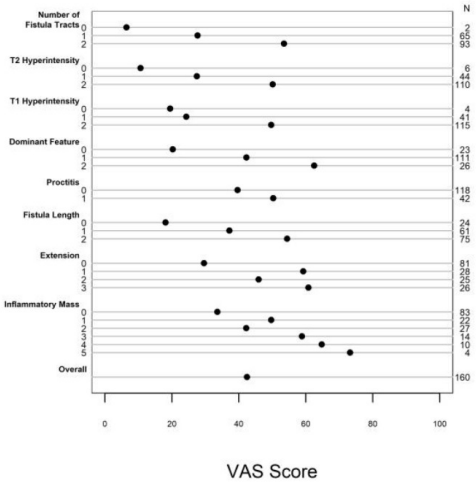
	VAI	mVAI	VAS	PDAI	IBDQ	CDAI	CRP
MAGNIFI-CD	0.65 (0.55, 0.74) n=153	0.75 (0.67, 0.81) n=153	0.69 (0.60, 0.76) n=155	0.14 (-0.02, 0.29) n=150	-0.09 (-0.25, 0.07) n=149	0.12 (-0.04, 0.28) n=147	0.08 (-0.10, 0.25) n=128
VAI		0.87 (0.82, 0.90) n=154	0.60 (0.49, 0.69) n=157	0.13 (-0.03, 0.29) n=152	-0.07 (-0.23, 0.09) n=151	0.20 (0.04, 0.35) n=149	0.06 (-0.11, 0.23) n=128
mVAI			0.63 (0.52, 0.72) n=154	0.12 (-0.04, 0.27) n=149	-0.08 (-0.24, 0.08) n=148	0.20 (0.04, 0.35) n=146	0.07 (-0.11, 0.24) n=126
VAS				0.28 (0.13, 0.42) n=155	-0.23 (-0.38, -0.08) n=154	0.24 (0.08, 0.38) n=152	0.12 (-0.05, 0.28) n=131
PDAI					-0.36 (-0.49, -0.21) n=151	0.24 (0.08, 0.39) n=150	-0.11 (-0.27, 0.07) n=129
IBDQ						-0.34 (-0.47, -0.19) n=151	-0.04 (-0.22, 0.13) n=127
CDAI							0.09 (-0.09, 0.26) n=127

[1]Correlations are based on the population used for index development and validation (N=160); numbers of patients included in each correlation analysis are based on the number of patients with data available for both outcomes.

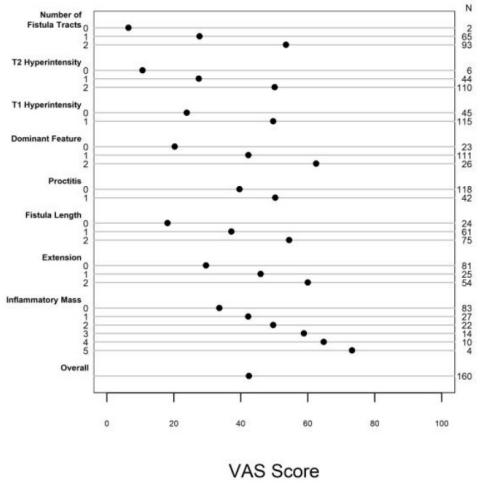
A.



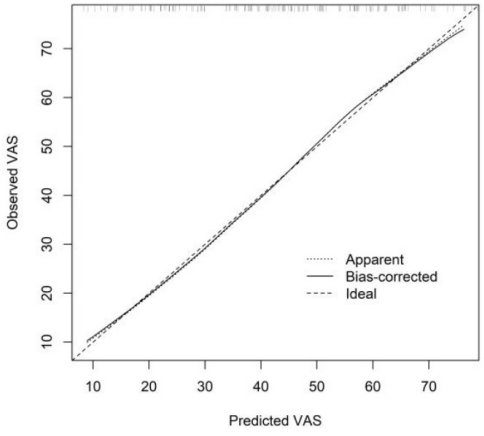
B.



C.



D.



WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT: We developed and validated magnetic resonance imaging (MRI) index for assessing perianal fistulas in patients with Crohn's disease (CD).

NEW FINDINGS: We developed an index, called the MAGNIFI-CD, based on 6 items. It assesses MRI data and determines perianal fistulizing CD activity with improved operating characteristics compared to previous indices.

LIMITATIONS: We validated the index only internally. External validation is necessary.

IMPACT: This index might be used as an outcome measure in clinical trials comparing treatment effects in patients with perianal fistulizing CD

LAY SUMMARY: We developed a new system to determine the severity of perianal fistulas in patients with Crohn's disease based on 6 factors collected during magnetic resonance imaging.

Supplementary Material

Hindryckx P et al. Development and Validation of MAGNIFI-CD: a Novel Magnetic Resonance Index for Fistula Imaging in Crohn's Disease

Figure 1. Diagnostics examination of residuals from the final model with six independent variables.

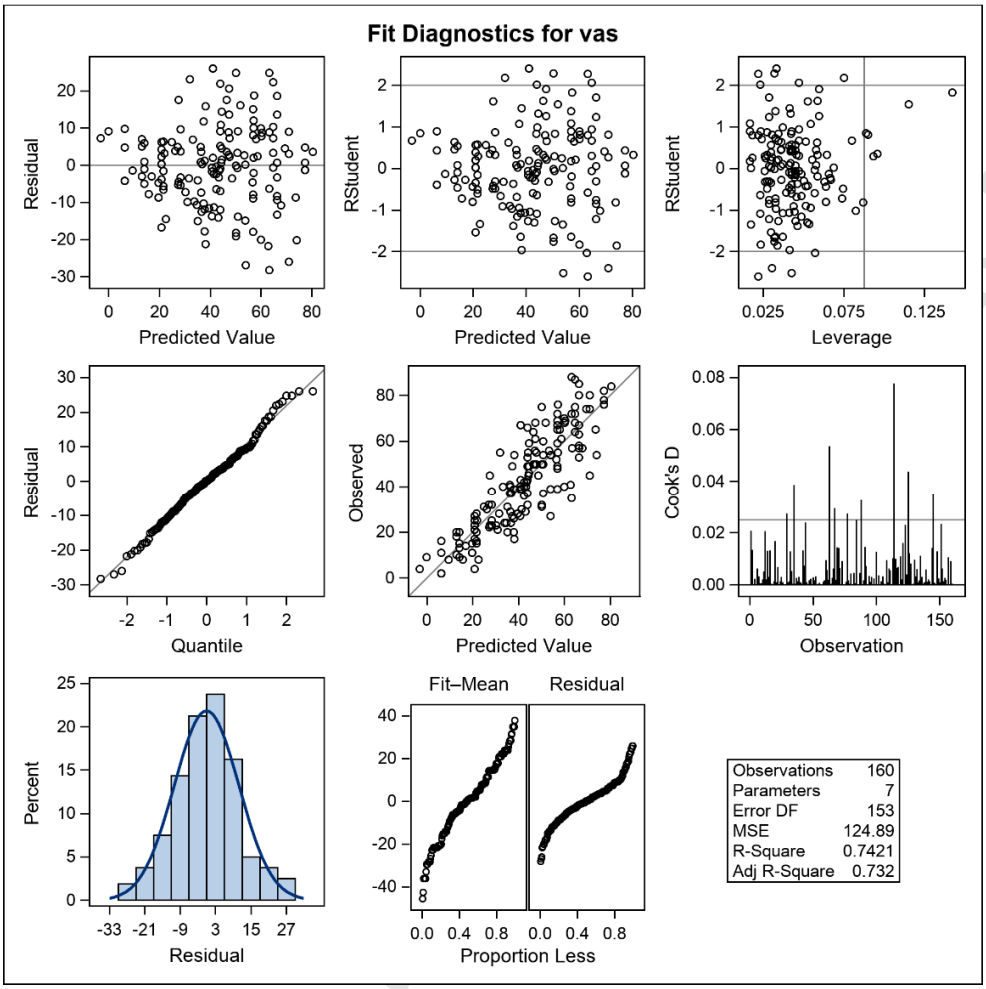


Table 1. Standard pelvic MRI acquisition protocol.

	Localizer	**Sagittal T2-weighted Fast Spin-Echo**	**Coronal T2-weighted Fast Spin-Echo**	**Axial T2-weighted Fast Spin-Echo**	**Axial T2-weighted Fast Spin-Echo with fat saturation**	**Axial T1-weighted Fast Spin-Echo with fat saturation plus IV contrast enhancement**
TE/TR (msec)	N/A	70/2500	70/2500	70/2500	85/4000	10.8-32.3/600
FOV (cm)	40	30	30	30	30	30
Slice Thickness (mm)	7	4	4	4	4	4
Distance Factor (mm)	5.0	0.4	0.4	0.4	0.4	0.4
Matrix	256x128	512x256	512x256	512x256	256x256	256x256
Frequency	N/A	SI	SI	RL	RL	RL

TE: echo time; TR: repetition time; FOV: field of view; N/A: Not applicable; SI = Superior inferior, RL = Right left

Table 2. Baseline and Week 24 clinical characteristics of the patients whose scans were used in the assessment of index reliability.

Characteristic	**Mean (SD)**	
	Baseline	Week 24
C-reactive protein, mg/L	9.1 (13.7)	9.0 (14.3)
Perianal Disease Activity Index score	6.1 (2.4)	3.3 (2.7)
Crohn's Disease Activity Index score	71.0 (46.5)	66.8 (52.6)
Inflammatory Bowel Disease Questionnaire Score	182.4 (28.5)	189.5 (32.2)

SD, standard deviation

Table 3. Baseline and Week 24 clinical characteristics of the patients whose scans were used in the assessment of index responsiveness.

Characteristic	**Mean (SD)**	
	Baseline	Week 24
C-reactive protein, mg/L	7.1 (10.5)	7.1 (11.5)
Perianal Disease Activity Index score	7.0 (2.8)	4.4 (3.6)
Crohn's Disease Activity Index score	92.9 (49.9)	89.7 (68.7)
Inflammatory Bowel Disease Questionnaire Score	169.9 (31.0)	177.3 (34.7)

SD, standard deviation

Table 4. Estimates (95% CI) of reader-specific intraclass correlation coefficients based on 40 scans.

	Reader			
	1	**2**	**3**	**4**
Modified van Assche index				
Extension	0.62 (0.37, 0.78)	0.55 (0.29, 0.73)	0.40 (0.11, 0.63)	0.87 (0.78, 0.93)
Hyperintensity of primary tract on T2-weighted images	0.69 (0.47, 0.83)	0.67 (0.46, 0.81)	0.78 (0.62, 0.88)	0.55 (0.28, 0.74)
Proctitis	0.78 (0.61, 0.88)	0.66 (0.44, 0.80)	0.63 (0.41, 0.79)	0.53 (0.26, 0.72)
Inflammatory mass	0.78 (0.60, 0.88)	0.89 (0.81, 0.94)	0.64 (0.41, 0.79)	0.87 (0.76, 0.93)
Dominant feature	0.76 (0.57, 0.87)	0.71 (0.50, 0.85)	0.63 (0.38, 0.79)	0.89 (0.79, 0.94)
Total mVAI	0.85 (0.72, 0.92)	0.86 (0.74, 0.93)	0.71 (0.50, 0.84)	0.80 (0.64, 0.89)
Other items explored during derivation of the modified van Assche index				
Number of fistula tracts	0.44 (0.14, 0.67)	0.48 (0.20, 0.69)	0.49 (0.22, 0.69)	0.68 (0.46, 0.82)
Location (modified definition)	0.63 (0.39, 0.79)	0.87 (0.77, 0.93)	0.69 (0.49, 0.82)	0.59 (0.33, 0.76)
Hyperintensity of primary tract on post-contrast T1-weighted images	0.51 (0.22, 0.72)	0.77 (0.59, 0.88)	0.69 (0.48, 0.83)	0.52 (0.22, 0.73)
Original VAI[1]				
Number of fistula tracts	0.41 (0.09, 0.65)	0.59 (0.34, 0.76)	0.49 (0.22, 0.69)	0.70 (0.50, 0.83)
Location	0.65 (0.42, 0.80)	0.74 (0.55, 0.85)	0.75 (0.57, 0.86)	0.41 (0.10, 0.64)
Extension	0.50 (0.20, 0.71)	0.45 (0.17, 0.67)	0.29 (-0.02, 0.55)	0.75 (0.57, 0.86)

	Reader			
	1	**2**	**3**	**4**
Hyperintensity of primary tract on T2-weighted images	0.69 (0.47, 0.83)	0.67 (0.46, 0.81)	0.78 (0.62, 0.88)	0.55 (0.28, 0.74)
Collections	0.90 (0.81, 0.94)	1.00	0.84 (0.73, 0.92)	1.00
Rectal wall involvement	0.78 (0.61, 0.88)	0.66 (0.44, 0.80)	0.63 (0.41, 0.79)	0.53 (0.26, 0.72)
Total VAI	0.83 (0.69, 0.91)	0.83 (0.70, 0.91)	0.79 (0.64, 0.89)	0.74 (0.55, 0.86)
Exploratory items				
Length of fistula tract	0.63 (0.39, 0.79)	0.52 (0.26, 0.72)	0.78 (0.62, 0.88)	0.78 (0.61, 0.88)
Number of external openings	0.56 (0.28, 0.74)	0.33 (0.02, 0.58)	0.48 (0.21, 0.69)	0.50 (0.21, 0.70)
Number of internal openings	0.35 (0.03, 0.61)	0.60 (0.35, 0.77)	0.29 (-0.02, 0.55)	0.46 (0.16, 0.67)
Visual Analogue Scale	0.90 (0.81, 0.94)	0.94 (0.90, 0.97)	0.77 (0.61, 0.87)	0.91 (0.84, 0.96)

Table 5. Distribution of investigated characteristics prior to item re-scaling (corresponding to Figure 1b)

Characteristic	n	Mean VAS (SD)
Number of fistula tracts		
0 None	2	6.5 (3.5)
1 Single, unbranched	65	27.7 (15.3)
2 Complex	93	53.5 (18.3)
Hyperintensity of primary tract on T2-weighted images		
0 Absent	6	10.7 (7.3)
1 Mild	44	27.5 (17.4)
2 Pronounced	110	50.2 (18.8)
Hyperintensity of primary tract on post-contrast T1-weighted images		
0 Absent	4	19.5 (23.8)
1 Mild	41	24.3 (16.9)
2 Pronounced	115	49.7 (18.6)
Dominant feature		
0 Predominantly fibrous	23	20.3 (18.4)
1 Predominantly granulation tissue	111	42.3 (18.5)
2 Predominantly fluid/pus	26	62.5 (17.0)
Proctitis		
0 No	118	39.6 (21.3)
1 Yes	42	50.3 (20.7)
Fistula length		
0 <2.5 cm	24	18.1 (11.9)
1 2.5 cm to 5 cm	61	37.2 (16.3)
2 > 5 cm	75	54.5 (19.5)
Extension		
0 Absent	81	29.6 (17.3)
1 Infralevatoric	28	59.3 (17.0)
2 Horseshoe configuration	25	46.0 (14.5)
3 Supralevatoric	26	60.8 (16.8)
Inflammatory mass		
0 Absent	83	33.6 (19.9)
1 Diffuse	22	49.7 (19.6)
2 Focal	27	42.3 (18.3)
3 Collections – small	14	58.9 (14.5)
4 Collections – medium	10	64.8 (14.0)
5 Collections – large	4	73.3 (8.5)
Overall	160	42.4 (21.6)

SD, standard deviation; VAS, visual analog score

Table 6. Distribution of investigated characteristics after item re-scaling (corresponding to Figure 1C)

Characteristic	n	Mean VAS (SD)
Number of fistula tracts (no re-scaling)		
0 None	2	6.5 (3.5)
1 Single, unbranched	65	27.7 (15.3)
2 Complex	93	53.5 (18.3)
Hyperintensity of primary tract on T2-weighted images (no re-scaling)		
0 Absent	6	10.7 (7.3)
1 Mild	44	27.5 (17.4)
2 Pronounced	110	50.2 (18.8)
Hyperintensity of primary tract on post-contrast T1-weighted images (re-scaled)		
0 Absent/Mild	45	23.9 (17.4)
1 Pronounced	115	49.7 (18.6)
Dominant feature (no re-scaling)		
0 Predominantly fibrous	23	20.3 (18.4)
1 Predominantly granulation tissue	111	42.3 (18.5)
2 Predominantly fluid/pus	26	62.5 (17.0)
Proctitis (no re-scaling)		
0 No	118	39.6 (21.3)
1 Yes	42	50.3 (20.7)
Fistula length (no re-scaling)		
0 <2.5 cm	24	18.1 (11.9)
1 2.5 cm to 5 cm	61	37.2 (16.3)
2 > 5 cm	75	54.5 (19.5)
Extension (re-scaled)		
0 Absent	81	29.6 (17.3)
1 Horseshoe configuration	25	46.0 (14.5)
2 Infralevatoric/Supralevatoric	54	60.0 (16.8)
Inflammatory mass (re-scaled)		
0 Absent	83	33.6 (19.9)
1 Focal	27	42.3 (18.3)
2 Diffuse	22	49.7 (19.6)
3 Collections – small	14	58.9 (14.5)
4 Collections – medium	10	64.8 (14.0)
5 Collections – large	4	73.3 (8.5)
Overall	160	42.4 (21.6)

SD, standard deviation; VAS, visual analog score

Table 7. Univariable linear regression models for candidate items after item re-scaling (N=160)

Candidate Item	Coefficient (SE)	R^2	P Value
Number of fistula tracts	25.5 (2.6)	.38	<.001
Hyperintensity of primary tract on T2-weighted images	21.5 (2.6)	.30	<.001
Hyperintensity of primary tract on post-contrast T1-weighted images	25.8 (3.2)	.29	<.001
Dominant feature	21.1 (2.6)	.29	<.001
Proctitis	10.7 (3.8)	.05	.006
Fistula length	17.9 (1.9)	.36	<.001
Extension	15.2 (1.5)	.41	<.001
Inflammatory mass	8.0 (1.0)	.27	<.001