TITLE: Reliability of MR Enterography Endpoints for Fibrostenosing Crohn Disease

ARTICLE TYPE: Original Research

SUMMARY STATEMENT: (29/30 words)

Stricture length, maximal stricture wall thickness, and maximal small bowel dilation on MR enterography can reliably describe Crohn disease stricture severity and serve as eligibility/efficacy criteria in clinical trials.

KEY RESULTS: (74/75 words)

- In a retrospective study of MR enterography exams from 2 independent samples of patients (N=99; N=51) with terminal ileal Crohn disease strictures, 10 features demonstrated at least moderate inter-rater reliability (intraclass correlation coefficient ≥0.41).
- Stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation were independently associated with stricture severity (regression coefficient range: 0.09-3.97, all P<.001).
- Three of 9 features evaluated demonstrated a small degree of responsiveness (area under the curve ≥0.56).

ABBREVIATIONS: AUC, area under the receiver operating curve; ICC, intraclass correlation coefficient; VAS, visual analog scale.

ABSTRACT (298/300 words)

Background: Clinical decision-making and drug development for fibrostenosing Crohn disease is constrained by a lack of imaging definitions, scoring conventions, and validated endpoints.

Purpose: Assess the reliability of MR enterography features to describe Crohn disease strictures and determine correlation with stricture severity.

Materials and Methods: A retrospective study of patients with symptomatic terminal ileal Crohn disease strictures who underwent MR enterography at tertiary care centers (Cleveland Clinic, 09/2013-11/2020; Mayo Clinic, 02/2008-03/2019) was conducted using convenience sampling. In Part 1 (Development), blinded and trained radiologists independently evaluated 26 MR enterography features from baseline and follow-up exams >6 months apart with no bowel resection performed between exams. Follow-up exams closest to 12 months post-baseline were selected. Reliability was assessed using the intraclass correlation coefficient (ICC). In Part 2 (Validation), after re-defining 5 features, reliability was re-estimated in an independent convenience sample using baseline exams. Multivariable linear regression analysis identified features with at least moderate inter-rater reliability (ICC≥0.41) that were independently associated with stricture severity.

Results: 99 (mean age, 40 years ± 14 [SD]; 50 males) and 51 (mean age, 45 years ± 16 [SD]; 35 females) patients were included in Part 1 and 2, respectively. In Part 1, 9 features had at least moderate inter-rater reliability. One additional feature demonstrated moderate reliability in Part 2. Stricture length (ICC=0.85 [95% CI 0.75-0.91] and 0.91 [0.75-0.96] in Part 1 and 2, respectively) and maximal associated small bowel dilation (ICC=0.74 [0.63-0.80] and 0.73 [0.58-0.87] in Part 1 and 2, respectively) had the highest inter-rater reliability. Stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation were independently (regression coefficients=0.09-3.97, *P*<.001) associated with stricture severity.

***Conclusion**:* MR enterography definitions and scoring conventions for reliably assessing features of Crohn disease strictures were developed and validated, and feature correlation with stricture severity was determined.

INTRODUCTION (266/400)

Symptomatic strictures, which occur in more than half of patients with Crohn disease, convey increased risk of disease-related complications and surgery (1-3). Although several validated endoscopic scoring indices exist for quantifying mucosal disease activity (4, 5), endoscopy is limited for assessing fibrostenosing Crohn disease. Specifically, it cannot evaluate transmural pathology, stricture length, penetrating complications, upstream bowel dilation, or the presence of proximal stricture and small bowel inflammation. Furthermore, in the case of multiple strictures, inability to pass the endoscope through the most distal stricture precludes the assessment of more proximal bowel segments (6). Alternative evaluative methods are required for clinical trials of anti-fibrotic therapy. This need is especially critical since multiple anti-fibrotic agents are under investigation (7).

Cross-sectional imaging with MR enterography or CT enterography is the preferred method for assessing fibrostenotic Crohn disease in clinical practice (8). Both modalities are highly sensitive and specific for identifying Crohn disease strictures (9-11). Although MR enterography is more expensive, less available, and more time consuming than CT enterography, advantages include superior soft tissue resolution and lack of ionizing radiation exposure (12).

Consensus recommendations define a stricture on CT/MR enterography as a luminal narrowing with unequivocal upstream dilation ≥3 cm (12). However, to our knowledge, this definition has not been validated. Moreover, guidance on how to describe stricture morphology (eg, how to measure a single stricture) and the reliability of key MR enterography-defined features (eg, luminal narrowing, bowel dilation, and wall thickening) is unknown.

The current study aimed to assess the reliability of MR enterography imaging features to describe Crohn disease strictures and determine their correlation with stricture severity.

MATERIALS AND METHODS (798/800)

Study design and patients

This retrospective analysis used data from a convenience sample of patients with stricturing Crohn disease who underwent standard-of-care MR enterography at Cleveland Clinic (Ohio, United States; 09/2013-11/2020) and Mayo Clinic (Minnesota, United States; 02/2008-03/2019). Institutional review board approval was obtained (Cleveland Clinic Foundation: 19-572; Mayo Clinic: 18-011580). Required written informed consent was waived. Alimentiv Inc. participated in study design; collection, analysis, and interpretation of data; report-writing; and the decision to submit for publication. Pfizer Inc. and the Leona M. and Harry B. Helmsley Charitable Trust provided financial support. All authors had control of the data and information submitted for publication.

Eligibility criteria included: (i) diagnosis of Crohn disease by a gastroenterologist, (ii) terminal ileal stricture within 15 cm of the ileocecal valve/ileocolonic anastomosis meeting CONSTRICT criteria (**Supplemental Table 1**) (13), and (iii) stricture symptoms within 1 month of baseline. Patients with penetrating disease (sinus tracts permitted), malignancy, or non-ileal/neo-terminal ileal strictures were excluded. Patients with >2 distal ileal strictures; end ileostomy with isolated subcutaneous stricture; diverting loop ileostomy proximal to the ileal stricture; total proctocolectomy with ileoanal/Kock pouch; distal ileal strictureplasty; no fat-saturated, gadolinium-enhanced imaging; and slice thickness ≥5 mm were also ineligible.

MR enterography image acquisition

The imaging protocol is described in **Supplemental Appendix 1.**

MR enterography stricture measurements and observations: Part 1 (Development)

Part 1 included patients with baseline and follow-up MR enterography >6 months apart and no bowel resection between exams. Follow-up MR enterography exams closest to 12 months post-baseline were selected. JGF and MEB (Part 2 central readers) ensured eligibility, image quality, and a spectrum of stricture severity (**Figure 1**).

Previously identified features were selected for evaluation (13) (**Table 1**). A Training Manual of definitions and scoring conventions was developed (**Supplemental Appendix 2**) (14). A single stricture was defined as (i) an area of continuous luminal narrowing; (ii) multiple areas of luminal narrowing connected by active inflammation, or (iii) multiple areas with luminal narrowing ≤3 cm apart. Maximal small bowel dilation could be observed proximal to the stricture, or within strictures containing multiple regions of luminal narrowing.

MR enterography exams were uploaded to a secure server. Images were de-identified and central readers were blinded to all patient information. A computer workstation captured annotations, measurements, and ratings (**Supplemental Appendix 3**).

Four radiologists (ECE, JPH, JLF, SPS, with 6-41 years of experience) completed a 1-day training session on the computer workstation before evaluation. Stricture severity was rated on a 100 mm visual analog scale (VAS) (0=completely normal; 100=worst disease ever seen) (**Supplemental Figure 1**).

To optimize memory extinction, MR enterography exams were randomly assessed during 2 interpretive reading sessions separated by ≥2 weeks. Baseline MR enterography exams were evaluated in both sessions, whereas one-quarter of follow-up exams were evaluated once. Only the most distal terminal ileal stricture was assessed. After central reading, JGF confirmed that the same stricture was evaluated at baseline and follow-up.

MR enterography stricture measurements and observations: Part 2 (Validation)

Revisions to the Training Manual were made to improve reliability estimates for select features (**Supplemental Appendix 2**). These features, and those with at least moderate inter-rater reliability, were re-evaluated by 4 trained radiologists (JGF, JLF, JMR, and MEB, with 8-40 years of experience) using an independent sample of MR enterography exams in Part 2. Reliability was re-estimated. A single MR enterography exam was required in Part 2.

Feature association with stricture severity and responsiveness testing

Multiple linear regression analysis evaluated the association between reliable features (intraclass correlation coefficient [ICC]≥0.41) and stricture severity, with stricture severity as the dependent and features as the independent variables.

Paired baseline and follow-up images were used to assess the responsiveness of reliable features (ie, the ability of features to detect change in stricture severity independent of therapy). Development of a multi-component index using reliable and responsive features was planned.

Statistical analysis

Summary statistics described baseline demographic characteristics. Reliability was quantified using the ICC. Point estimates for ICCs were estimated using a 2-way random effects analysis of variance model with interaction between MR enterography exams and central readers. Two-sided 95% CIs were estimated using the nonparametric percentile cluster bootstrap method. ICC estimates were interpreted as: <0.00=poor; 0.00-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=substantial; and >0.80=almost perfect (15).

In exploratory analyses, the within-patient SD for the most reliable features was calculated (**Supplemental Appendix 4**).

Responsiveness was defined as an improvement in the stricture severity VAS of at least one-half of the SD of the baseline VAS (16) and quantified using the area under the receiver operating curve (AUC) (0.56=small; 0.64=medium; 0.71=large) (17)(18, 19).

A model of stricture severity was developed using multivariable linear regression modelling (**Supplemental Appendix 4**).

Sample size calculation was estimated a priori based on the 1-way random effects model (20) (**Supplemental Appendix 4**).

Analyses were conducted by ZW and RC using SAS version 9.4 (SAS Institute).

RESULTS (977/1000)

Patient characteristics

A total of 3702 MR enterography exams were acquired from the 2 clinics for Part 1 (**Figure 1**) and Part 2 (**Supplemental Figure 2**). A total of 99 (mean age, 40 years ± 14 [SD]; 50 males, 49 females) and 51 patients (mean age, 45 years ± 16 [SD]; 16 males, 35 females) were included in Part 1 and 2, respectively (**Table 2**). In Part 1, median stricture length was 9.2 cm (interquartile range [IQR] 4.3, 16.3 cm), median stricture wall thickness was 9.1 mm (IQR 8.0, 10.8 mm), and median maximal associated small bowel dilation was 30.3 mm (IQR 26.8, 35.9 mm). In Part 2, median stricture length was 12.0 cm (interquartile range [IQR] 4.2, 19.3 cm), median stricture wall thickness was 8.5 mm (IQR 7.1, 10.0 mm), and median maximal associated small bowel dilation was 35.0 mm (IQR 30.5, 42.8 mm). **Supplemental Table 2** describes the stricture characteristics of the samples.

Assessment of reliability: Part 1

Intra- and inter-rater ICCs are summarized in **Table 3**. Intra-rater reliability estimates for individual central readers are provided in **Supplemental Table 3**.

In Part 1, there was almost perfect inter-rater reliability for assessment of stricture length (ICC 0.85 [95% CI: 0.75, 0.91]) and substantial reliability for maximal associated small bowel dilation (ICC 0.74 [95% CI: 0.63, 0.80]; **Figure 2 and 3**). MR enterography measurements with moderate inter-rater reliability included maximal stricture wall thickness (ICC 0.58 [95% CI: 0.48, 0.67]) and maximal luminal diameter within the stricture (ICC 0.50 [95% CI 0.37, 0.62]). All remaining measurements had slight to fair reliability (ICC range 0.09 [95% CI 0.04, 0.12]–0.29 [95% CI 0.19, 0.35]). Intra-rater reliability was at least moderate (ICC range 0.46 [95% CI 0.36, 0.55]–0.92 [95% CI 0.86, 0.96]) for all measurements except wall thickness of a normal-appearing ileal loop, which had fair inter-rater reliability (ICC 0.27 [95% CI 0.15, 0.48]).

MR enterography observations with moderate inter-rater reliability linked to inflammation included diffusion restriction (ICC 0.48 [95%CI 0.17, 0.80]) and mural hyperenhancement (ICC 0.44 [95% CI 0.36, 0.52]), and those reflective of fibrosis were enhancement pattern on the delayed gadolinium phase (ICC 0.48 [95% CI 0.35, 0.59]) and sacculation (ICC 0.46 [95% CI 0.36, 0.55]) (**Figure 4**). All remaining observations had slight to fair inter-rater reliability (ICC range 0.03 [95% CI 0.00, 0.08]–0.38 [95% CI 0.29, 0.46]). Intra-rater reliability was moderate to almost perfect for all observations (ICC range 0.45 [95% CI 0.34, 0.55]–0.67 [95% CI 0.55, 0.76]).

The mean stricture severity VAS score was 40.8 (SD 22.6); scores approximately followed a normal distribution (**Supplemental Figure 3**). Inter-rater reliability was moderate (ICC 0.51 [95% CI: 0.43, 0.58]) and intra-rater reliability was substantial (ICC 0.78 [95% CI: 0.72, 0.83]) (**Table 3**).

The within-patient SD for stricture length, maximal associated small bowel dilation, maximal stricture wall thickness, and maximal luminal diameter within the stricture was 4.9 cm, 5.0 mm, 6.4 mm, and 8.0 mm, respectively (**Figure 5**). For shorter strictures (≤5.0 cm) the within-patient SD for stricture length, maximal associated small bowel dilation, maximal stricture wall thickness, and maximal luminal diameter within the stricture was 2.7 cm, 6.0 mm, 2.6 mm, and 5.1 mm, respectively. For longer strictures (>5.0 cm) the within-patient SD for these features was 5.6 cm, 4.5 mm, 8.0 mm, and 8.9 mm, respectively.

Assessment of reliability: Part 2

The investigators theorized that modifying the Training Manual might improve the reliability of intramural T2 signal, luminal ulceration, perienteric fat edema/free fluid, perienteric fat stranding, and vasa recta distension. A second version of the Training Manual was developed, with perienteric fat edema/free fluid and perienteric fat stranding collapsed into 1 observation. The 4 re-defined features, and 9 features with at least moderate inter-rater reliability in Part 1,

were re-tested for reliability in Part 2 (**Table 3**). Intra-rater reliability estimates for individual central readers are provided in **Supplemental Table 4**.

Inter-rater reliability for stricture length remained almost perfect (ICC 0.91 [95% CI 0.75, 0.96]) (**Figure 6**). Maximal associated small bowel dilation continued to demonstrate substantial inter-rater reliability (ICC 0.73 [95% CI 0.58, 0.87]), and moderate inter-rater reliability was verified for maximal stricture wall thickness (ICC 0.57 [95% CI 0.43, 0.67]) and maximal luminal diameter (ICC 0.55 [95% CI 0.40, 0.68]). Moderate inter-rater reliability was confirmed for enhancement pattern at the delayed gadolinium phase, restricted diffusion, sacculation, and mural hyperenhancement (ICC range 0.42 [95% CI 0.27, 0.54]–0.52 [95% CI 0.38, 0.66]). There was a numerical improvement in the inter-rater ICCs for the 4 modified observations. Inter-rater reliability for vasa recta distention increased from fair to moderate after re-training (ICC 0.58 [95% CI 0.45, 0.69]).

The mean VAS of stricture severity score was 41.3 (SD 20.0), and the distribution of scores was close to normal (**Supplemental Figure 4**). Inter-rater reliability was substantial (ICC 0.68 [95% CI 0.57, 0.76]) and intra-rater reliability was almost perfect (ICC 0.83 [95% CI 0.75, 0.87]) (**Table 3**).

Relationship between reliable features and global stricture severity

Multivariable regression analysis using data from Part 1 demonstrated that global stricture severity could be predicted by stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation (regression coefficients: 0.09, 3.97, and 0.83, respectively; all $P<.001$; adjusted $R^2=0.37$) (**Table 4**). When data from Part 2 were used, the regression coefficients were 0.10, 3.11, 0.88 for stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation, respectively (all $P<.001$; adjusted $R^2=0.77$).The bootstrap internal validation resulted in an R^2 value of 0.72 and optimism-corrected calibration slope of 0.99, suggesting reasonable external performance (**Table 4**).

Responsiveness testing

Of the 9 features tested for responsiveness, 3 demonstrated a small degree of responsiveness (maximal associated small bowel dilation [AUC=0.59], enhancement pattern at delayed gadolinium phase [AUC=0.58], and sacculation [AUC=0.56]) while the remaining 6 were not responsive (AUC<0.56; **Supplemental Table 5**).

A candidate index using reliable features was not developed due to the lack of observed feature-level responsiveness.

DISCUSSION

Treatment of fibrostenosing Crohn disease is a critical unmet need, and drug development is hindered by a lack of valid clinical trial eligibility criteria and efficacy endpoints. Endoscopic and surgical assessment is often precluded in fibrostenosing Crohn disease, and symptoms poorly reflect inflammatory and obstructive severity. Histologic indices have been used to evaluate stricturing Crohn disease, however, existing instruments are heterogeneous (21) and remain largely unvalidated (22). Furthermore, histologic assessment cannot describe extent or degree of fibrostenosis. While interdisciplinary initiatives have proposed imaging-based stricture definitions (12, 13), measurement and scoring criteria remain unestablished. We sought to develop MR enterography definitions and scoring conventions and determine features that can be used as eligibility criteria and endpoints in clinical trials and clinical practice. Three MR enterography features were reliably measured/observed in 2 separate samples of symptomatic patients with small bowel strictures. Stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation showed moderate to almost perfect inter-rater agreement (ICC range 0.57 [95% CI 0.43, 0.67]–0.91 [95% CI 0.75-0.96]) and were independently associated with stricture severity (all P<.001).

A companion study (23) also identified stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation as reliable and strongly associated with stricture

severity on CT enterography using the same definitions. However, the 2 studies were designed and undertaken at different times, had slightly different eligibility criteria, and used non-overlapping sets of central readers. While we did not set out with the hypothesis that the same features would be identified for both modalities, this finding supports feature validity and suggests that MR enterography and CT enterography can be used interchangeably. Additionally, these 3 features have face validity with respect to the pathophysiology of fibrostenosing disease. Stricture wall thickness and small bowel dilation directly reflect the underlying disease process while stricture length reflects total disease burden. They are also established prognostic factors associated with surgical intervention (24-29). Using our conventions for defining stricture border, >95% of patients solely had a single stricture within 15 cm of the ileocecal valve/anastomosis. Incorporating this knowledge into clinical practice could result in more efficient and consistent stricture descriptions.

Maximal small bowel dilation is potentially useful as an eligibility criterion and efficacy endpoint. Our data suggest that regardless of stricture length, a ≥6 mm reduction in small bowel dilation is a reasonable threshold. While stricture wall thickness of ≥5.0 mm may be an appropriate for eligibility criterion, this feature is a suboptimal endpoint since it has a limited dynamic range and can reflect fibrosis, muscular hypertrophy, and/or inflammation. Improvement in stricture length (≥50% reduction) or reduction in length based on the within-patient SD may be a reasonable endpoint. Early therapeutic studies approved by regulatory agencies have used the imaging definitions validated in our study (NCT05843578; NCT05013385). It should be noted that feature reliability may be improved by side-by-side timepoint comparisons (30), and semi-automated (31) or artificial intelligence-based algorithms.

Meaningful responsiveness was not demonstrated for candidate features, and the VAS did not identify a sufficient number of responders. Lack of response may be explained by the inability of existing anti-inflammatory therapies to affect stricture morphology, or imaging having been performed during symptomatic exacerbations.

The development of effective anti-fibrotic therapies for stricturing Crohn disease will enable re-evaluation of feature responsiveness, and if appropriate items are identified, a multi-item evaluative index may be developed. In the interim, stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation should be used with a patient-reported outcome in clinical trials of anti-fibrotics. Given the regulatory requirement of a co-primary patient reported outcome, we only included patients with symptomatic Crohn disease strictures. The importance of using reliable, objective outcome measures for the assessment of anti-fibrotic therapies is underscored by the finding that symptoms do not consistently reflect disease state (32, 33).

Our work has limitations. First, this was a retrospective study of images from tertiary care centers. However, 2 independent patient samples were created to represent the patient populations in future clinical trials, thus our findings and are likely generalizable for this application (13). Second, a VAS served as the basis for responsiveness testing since an accepted reference standard for stricture severity is lacking. While the VAS has no independent clinical validity, we found it to be at least moderately reliable in the 2 separate patient samples as estimated by 2 panels of central readers. Moreover, there is a strong historical precedent for using the VAS to develop Crohn disease indices that have subsequently been validated and gained regulatory acceptance (5, 34, 35). Interestingly, the Crohn disease endoscopic index of severity – which was derived using the VAS – was used as a surrogate for endoscopic inflammatory severity during the development of the MR index of Crohn disease activity (11). Third, we were unable to develop and validate an evaluative index. This may not be possible given the wide spectrum of Crohn disease strictures, yet we remain hopeful. Fourth, central readers were experienced gastrointestinal subspecialists, and the majority were members of the Society of Abdominal Radiology Crohn Disease-Focused Panel. Finally, we did not investigate the association between MR enterography and transmural histopathologic features (eg, degree of fibrosis, inflammation, muscularis propria hypertrophy, and collagen content).

We demonstrated that MR enterography can be used to reliably evaluate features of Crohn disease symptomatic terminal ileal strictures when standardized definitions and scoring conventions were implemented, and stricture length, maximal stricture wall thickness, and maximal associated small bowel dilation were correlated with stricture severity. While further work is needed to establish responsiveness, these results provide initial guidance regarding eligibility criteria and efficacy endpoints for clinical trial design.

REFERENCES

1. Cosnes J, Cattan S, Blain A, et al. Long-term evolution of disease behavior of Crohn's disease. Inflamm Bowel Dis 2002;8:244-250. doi: 10.1097/00054725-200207000-00002

2. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. Gut 2001;49:777-782. doi: 10.1136/gut.49.6.777

3. Rieder F, Zimmermann EM, Remzi FH, Sandborn WJ. Crohn's disease complicated by strictures: a systematic review. Gut 2013;62:1072-1084. doi: 10.1136/gutjnl-2012-304353

4. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointestinal endoscopy 2004;60:505-512. doi: 10.1016/s0016-5107(04)01878-4

5. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983-989. doi: 10.1136/gut.30.7.983

6. Al-Bawardy B, Hansel SL, Fidler JL, Barlow JM, Bruining DH. Endoscopic and radiographic assessment of Crohn's disease. Gastroenterol Clin North Am 2017;46:493-513. doi: 10.1016/j.gtc.2017.05.005

7. Vogel J, da Luz Moreira A, Baker M, et al. CT enterography for Crohn's disease: accurate preoperative diagnostic imaging. Dis Colon Rectum 2007;50:1761-1769. doi: 10.1007/s10350-007-9005-6

8. Bettenworth D, Bokemeyer A, Baker M, et al. Assessment of Crohn's disease-associated small bowel strictures and fibrosis on cross-sectional imaging: a systematic review. Gut 2019;68:1115-1126. doi: 10.1136/gutjnl-2018-318081

9. Bruining DH, Siddiki HA, Fletcher JG, et al. Benefit of computed tomography enterography in Crohn's disease: effects on patient management and physician level of confidence. Inflamm Bowel Dis 2012;18:219-225. doi: 10.1002/ibd.21683

10. Fiorino G, Bonifacio C, Peyrin-Biroulet L, et al. Prospective comparison of computed tomography enterography and magnetic resonance enterography for assessment of disease activity and complications in ileocolonic Crohn's disease. Inflamm Bowel Dis 2011;17:1073-1080. doi: 10.1002/ibd.21533

11. Rimola J, Rodriguez S, Garcia-Bosch O, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. Gut 2009;58:1113-1120. doi: 10.1136/gut.2008.167957

12. Bruining DH, Zimmermann EM, Jr EVL, et al. Consensus recommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography in patients with small bowel Crohn’s disease. Radiology 2018;286:776-799. doi: 10.1148/radiol.2018171737

13. Rieder F, Bettenworth D, Ma C, et al. An expert consensus to standardise definitions, diagnosis and treatment targets for anti-fibrotic stricture therapies in Crohn's disease. Aliment Pharmacol Ther 2018;48:347-357. doi: 10.1111/apt.14853

14. U.S. Food and Drug Administration. Clinical performance assessment: considerations for computer-assisted detection devices applied to radiology images and radiology device data in premarket notification (510(k)) submissions. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-performance-assessment-considerations-computer-assisted-detection-devices-applied-radiology. Published 2022. Updated September 28, 2022. Accessed October 27, 2023.

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

16. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical care 2003;41:582-592. doi: 10.1097/01.MLR.0000062554.74615.4C

17. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39:897-906. doi: 10.1016/0021-9681(86)90038-x

18. Grissom RJ. Probability of the superior outcome of one treatment over another. J Appl Psychol 1994;79:314-316. doi: 10.1037/0021-9010.79.2.314
19. Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ, Harmon RJ. Measures of clinical significance. J Am Acad Child Adolesc Psychiatry 2003;42:1524-1529. doi: 10.1097/00004583-200312000-00022
20. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Statistics in medicine 2012;31:3972-3981. doi: 10.1002/sim.5466
21. Gordon IO, Bettenworth D, Bokemeyer A, et al. International consensus to standardise histopathological scoring for small bowel strictures in Crohn's disease. Gut 2022;71:479-486. doi: 10.1136/gutjnl-2021-324374
22. Gordon IO, Bettenworth D, Bokemeyer A, et al. Histopathology Scoring Systems of Stenosis Associated With Small Bowel Crohn's Disease: A Systematic Review. Gastroenterology 2020;158:137-150.e131. doi: 10.1053/j.gastro.2019.08.033
23. Rieder F, Ma C, Hanzel J, et al. Reliability of CT Enterography for Fibrostenosing Crohn Disease. Radiology (in press).
24. Bouhnik Y, Carbonnel F, Laharie D, et al. Efficacy of adalimumab in patients with Crohn's disease and symptomatic small bowel stricture: a multicentre, prospective, observational cohort (CREOLE) study. Gut 2018;67:53-60. doi: 10.1136/gutjnl-2016-312581
25. Bettenworth D, Gustavsson A, Atreja A, et al. A pooled analysis of efficacy, safety, and long-term outcome of endoscopic balloon dilation therapy for patients with stricturing Crohn's disease. Inflamm Bowel Dis 2017;23:133-142. doi: 10.1097/MIB.0000000000000988
26. El Ouali S, Baker ME, Lyu R, et al. Validation of stricture length, duration and obstructive symptoms as predictors for intervention in ileal stricturing Crohn's disease. United European Gastroenterol J 2022;10:958-972. doi: 10.1002/ueg2.12314
27. Maehata Y, Nagata Y, Moriyama T, et al. Risk of surgery in patients with stricturing type of Crohn's disease at the initial diagnosis: a single center experience. Intest Res 2019;17:357-364. doi: 10.5217/ir.2018.00107

28. Grass F, Fletcher JG, Alsughayer A, et al. Development of an objective model to define near-term risk of ileocecal resection in patients with terminal ileal Crohn's disease. Inflamm Bowel Dis 2019;25:1845-1853. doi: 10.1093/ibd/izz079
29. Inoue A, Bartlett DJ, Shahraki N, et al. Predicting risk of surgery in patients with small bowel Crohn's disease strictures using computed tomography and magnetic resonance enterography. Inflamm Bowel Dis 2022;28:1677-1686. doi: 10.1093/ibd/izab332
30. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-263.
31. Lin S, Lin X, Li X, Chen M, Mao R. Making qualitative intestinal stricture quantitative: embracing radiomics in IBD. Inflamm Bowel Dis 2019;26:743-745. doi: 10.1093/ibd/izz197
32. Bruining DH, Siddiki HA, Fletcher JG, Tremaine WJ, Sandborn WJ, Loftus EV, Jr. Prevalence of penetrating disease and extraintestinal manifestations of Crohn's disease detected with CT enterography. Inflamm Bowel Dis 2008;14:1701-1706. doi: 10.1002/ibd.20529
33. Jones J, Loftus EV, Jr., Panaccione R, et al. Relationships between disease activity and serum and fecal biomarkers in patients with Crohn's disease. Clin Gastroenterol Hepatol 2008;6:1218-1224. doi: 10.1016/j.cgh.2008.06.010
34. Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976;70:439-444.
35. Hindryckx P, Jairath V, Zou G, et al. Development and Validation of a Magnetic Resonance Index for Assessing Fistulas in Patients With Crohn's Disease. Gastroenterology 2019;157:1233-1244.e1235. doi: 10.1053/j.gastro.2019.07.027
36. Grand DJ, Guglielmo FF, Al-Hawary MM. MR enterography in Crohn's disease: current consensus on optimal imaging technique and future advances from the SAR Crohn's disease-focused panel. Abdom Imaging 2015;40:953-964. doi: 10.1007/s00261-015-0361-8

TABLES

Table 1. Definitions of imaging features for assessing Crohn disease-related strictures on MR enterography.

Feature	Definition/Measurement
MR enterography imaging measurements	
Stricture length	Length of stricture, measured in cm
Maximal associated small bowel dilation	Greatest dilation within stricture or proximal small bowel, measured in mm
Maximal stricture wall thickness	Maximally thickened area, measured in mm
Maximal luminal diameter within stricture	Greatest luminal diameter within the stricture, measured in mm
Minimum luminal diameter within stricture	Luminal diameter from inner wall to inner wall in the most narrowed region, measured in mm
Luminal diameter of normal appearing ileal loop	Luminal diameter of normal, adjacent, appropriately distended bowel loop, measured in mm
Stricture distance to ileocecal valve/anastomosis	Distance from the ileocecal valve/anastomosis to the distal end of stricture, measured in mm
Wall thickness of normal appearing ileal loop distended with enteric contrast	Wall thickness of normal, adjacent appropriately distended bowel loop, measured in mm
MR enterography imaging observations	
Stricture type	Measured as naïve, anastomotic
Pattern of enhancement: 7-minute delayed gadolinium phase	Measured as homogenous, not homogenous
Restricted diffusion (DWI)[a]	Measured as absent, present
Sacculation	Measured as absent, present
Mural hyperenhancement	Measured as absent, mild, moderate, severe
Perienteric fat edema or free fluid	Measured as absent, present
Percentage of stricture with luminal narrowing	Proportion of the stricture in which the luminal area is <50% of a normally distended ileal loop, measured as <50%, 50-75%, >75%
Vasa recta distension	Increased number and size of vasa recta that supply a bowel loop, measured as absent, present
Regional enlarged lymph nodes	Measured as absent, present
Fibrofatty proliferation	Measured as absent, present
Whiskering/serosal wall irregularity	Measured as absent, present
Intramural T2 signal on fat-suppressed images	Measured as absent, mild, moderate, severe
Luminal ulcerations	Breaks in the luminal wall that extend into the bowel wall, measured as absent, present
Perienteric fat stranding	Measured as absent, present
Intramural fat	Measured as absent, present
Pattern of enhancement: portal/enteric phase	Measured as portal, enteric
Intensity of delayed enhanced compared to portal	Measured as greater than portal, equal to portal, less than portal

Abbreviations: DWI, diffusion weighted imaging.

[a]Maximum B value = 800 sec/mm^2

Table 2. Patient characteristics for the development and validation groups at the time of the first MR enterography exam

Characteristic	Development group (N=99)[a]	Validation group (N=51)
Age, years		
Mean (SD)	39.9 (14.3)	45.2 (15.8)
Median (IQR)	38.0 (28.0, 51.0)	41.0 (34.0, 58.0)
Range	19, 71	20, 80
Age at symptom onset, years		
Mean (SD)	20.7 (9.6)	24.4 (9.7)
Median (IQR)	19.0 (15.0, 24.0)	23.0 (18.0, 27.0)
Age at Crohn disease **diagnosis, years**		
Mean (SD)	21.8 (10.3)	25.0 (10.2)
Median (IQR)	20.0 (15.0, 25.0)	23.0 (19.0, 28.0)
Sex, n (%)		
Male	50 (50.5)	16 (31.4)
Female	49 (49.5)	35 (68.6)
Race, n (%)[b]		
White	96 (97.0)	49 (96.1)
Black or African American	1 (1.0)	0 (0)
Not reported	2 (2.0)	2 (3.9)
Weight, kg		
Mean (SD)	73.8 (17.8)	74.1 (16.8)
Median (IQR)	72.0 (60.6, 86.2)	73.5 (62.4, 82.8)
Height, cm		
Mean (SD)	171.6 (9.2)	169.0 (9.6)
Median (IQR)	172.0 (165.0, 177.4)	170.2 (162.6, 175.3)
BMI[c]		
Mean (SD)	25 (5.6)	25.8 (4.6)
Median (IQR)	24.5 (21.3, 26.9)	25.6 (22.6, 28.2)
Smoking status, n (%)		
Never	70 (70.7)	26 (51.0)
Former	17 (17.2)	17 (33.3)
Current	11 (11.1)	8 (15.7)
Disease location, n (%)		
Colon	0 (0)	0 (0)
Ileum	99 (100)	51 (100)
Colon and ileum	54 (54.5)	26 (51.0)

Abbreviations: BMI, body mass index; ; IQR, interquartile range;.

[a]Data missing for 1 patient from the development group; see **Figure 1**.

[b]As recorded in chart history and self-reported.

[c]BMI was calculated as $kg/m^{2.}$

Table 3. Inter- and intra-rater reliability of radiographic features for assessing Crohn disease-related strictures[a]

	Development Group (n=99)[b]		Validation Group (n=51)	
Feature	Inter-rater ICC (95% CI)	Intra-rater ICC (95% CI)	Inter-rater ICC (95% CI)	Intra-rater ICC (95% CI)
MR enterography imaging measurements				
Stricture length	*0.85 (0.75, 0.91)*	0.92 (0.86, 0.96)	*0.91 (0.75, 0.96)*	0.96 (0.87, 0.99)
Maximal associated small bowel dilation	*0.74 (0.63, 0.80)*	0.79 (0.69, 0.84)	*0.73 (0.58, 0.87)*	0.86 (0.74, 0.93)
Maximal stricture wall thickness[c]	*0.58 (0.48, 0.67)*	0.70 (0.62, 0.76)	*0.57 (0.43, 0.67)*	0.71 (0.61, 0.79)
Maximal luminal diameter within stricture	*0.50 (0.37, 0.62)*	0.64 (0.54, 0.74)	*0.55 (0.40, 0.68)*	0.70 (0.57, 0.83)
Minimum luminal diameter within stricture[d]	0.29 (0.19, 0.35)	0.48 (0.36, 0.63)	NA	NA
Luminal diameter of normal-appearing ileal loop	0.29 (0.20, 0.38)	0.46 (0.36, 0.55)	NA	NA
Minimum luminal diameter within stricture[d]	0.29 (0.19, 0.35)	0.48 (0.36, 0.63)	NA	NA
Stricture distance to ileocecal valve/anastomosis[e]	0.21 (0.10, 0.31)	0.64 (0.37, 0.83)	NA	NA
Wall thickness of normal loop	0.09 (0.04, 0.12)	0.27 (0.15, 0.48)	NA	NA
MR enterography imaging observations				
Stricture type	*0.61 (0.51, 0.70)*	0.76 (0.68, 0.83)	*0.70 (0.57, 0.81)*	0.87 (0.81, 0.93)
Pattern of enhancement: delayed gadolinium phase	*0.48 (0.35, 0.59)*	0.67 (0.55, 0.76)	*0.44 (0.29, 0.56)*	0.69 (0.56, 0.80)
Restricted diffusion (DWI)	*0.48 (0.17, 0.80)*	0.54 (0.29, 0.86)	*0.48 (0.21, 0.69)*	0.58 (0.25, 0.78)
Sacculation	*0.46 (0.36, 0.55)*	0.63 (0.54, 0.71)	*0.52 (0.38, 0.66)*	0.74 (0.62, 0.84)
Mural hyperenhancement	*0.44 (0.36, 0.52)*	0.64 (0.55, 0.70)	*0.42 (0.27, 0.54)*	0.69 (0.57, 0.78)
Perienteric fat edema or free fluid[f,g]	0.38 (0.29, 0.46)	0.62 (0.54, 0.69)	NA[d]	NA[d]
Percentage of stricture with luminal narrowing	0.36 (0.20, 0.50)	0.46 (0.31, 0.60)	NA	NA
Vasa recta distension[f]	0.36 (0.27, 0.44)	0.58 (0.48, 0.67)	*0.58 (0.45, 0.69)*	0.67 (0.54, 0.78)
Regional enlarged lymph nodes	0.35 (0.26, 0.44)	0.57 (0.48, 0.66)	NA	NA
Fibrofatty proliferation	0.33 (0.25, 0.41)	0.61 (0.53, 0.68)	NA	NA
Whiskering/serosal wall irregularity	0.33 (0.24, 0.41)	0.53 (0.42, 0.63)	NA	NA

Intramural T2 signal on fat-suppressed images[f]	0.33 (0.23, 0.40)	0.56 (0.47, 0.63)	0.40 (0.30, 0.49)	0.76 (0.70, 0.80)
Luminal ulcerations[f]	0.27 (0.19, 0.34)	0.59 (0.51, 0.66)	0.35 (0.21, 0.48)	0.58 (0.47, 0.69)
Perienteric fat stranding[f,g]	0.25 (0.15, 0.35)	0.45 (0.34, 0.55)	0.37 (0.26, 0.48)	0.72 (0.62, 0.81)
Intramural fat	0.24 (0.13, 0.33)	0.49 (0.36, 0.60)	NA	NA
Pattern of enhancement: portal/enteric phase	0.09 (0.03, 0.16)	0.52 (0.42, 0.61)	NA	NA
Intensity of delayed gadolinium compared to portal/enteric	0.03 (0.00, 0.08)	0.60 (0.50, 0.69)	NA	NA
Visual analog scale[h]				
Stricture severity	*0.51 (0.43, 0.58)*	0.78 (0.72, 0.83)	*0.68 (0.57, 0.76)*	0.83 (0.75, 0.87)

Abbreviations: DWI, diffusion weighted imaging; NA, not applicable; ICC, intraclass correlation coefficient; VAS, visual analog scale.

[a]The interpretation of ICC estimates was guided by benchmarks proposed by Landis and Koch (<0.00="poor"; 0.00-0.20="slight"; 0.21-0.40="fair," 0.41-0.6="moderate," 0.61-0.80="substantial," and >0.80="almost perfect" reliability) (15).

[b]Data missing for 1 patient in the development group; see **Figure 1**.

[c]Maximal stricture wall thickness was calculated as the mean of 3 measurements.

[d]Relatively low ICC estimates were observed for *minimal luminal diameter of the stricture,* which can be explained by the narrow range of the total measurements in the study sample. The mean rating of this feature was 3.0 mm (SD: 1.2 mm). The 10th percentile was 1.3 mm and the 90th percentile was 5.0 mm.

[e]Relatively low ICC estimates were observed for *stricture distance to the ileocecal valve/anastomosis,* which can be explained by nearly all of the strictures being located adjacent to the ileocecal valve/anastomosis. For example, out of 200 baseline MR eneterography observations, the 4 central readers indicated that 180, 171, 187, and 193 strictures abutted the ileocecal valve/anastomosis, respectively.

[f]Features for which radiologist reader instructions/definitions were changed in Part 2 in an effort to improve agreement.

[g]Perienteric fat stranding and perienteric fat edema or free fluid were combined into a single feature in Part 2.

[h]Stricture severity was rated on a VAS where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen).

NA = Feature dropped from reader evaluation in the validation group due to poor reliability in the development group.

Italic text = Features for which the threshold of moderate reliability (ie, inter-rater ICC≥0.41) was met.

Table 4. Relationship between reliable features and global stricture severity.

Feature	Part 1 (Development)		Part 2 (Validation)	
	Regression coefficient (95% CI)	***P* value[a]**	**Regression coefficient (95% CI)**	***P* value[a]**
Stricture length	0.09 (0.07, 0.15)	<.0001	0.10 (0.08, 0.12)	<.0001
Maximal stricture wall thickness	3.97 (2.58, 5.35)	<.0001	3.11 (1.73, 4.49)	<.0001
Maximal associated small bowel dilation	0.83 (0.37, 1.28)	0.0006	0.88 (0.62, 1.15)	<.0001

[a]*P* values were calculated using Student's t-test.

FIGURE LEGENDS

Figure 1. Creation of the development group. Part 1 (development phase) included baseline and follow-up MR enterography exams from 100 patients with Crohn disease who had symptomatic terminal ileal strictures. Abbreviations: IPAA, ileal pouch-anal anastomosis; TI, terminal ileum.

[1]Stricture was defined by the inclusion criteria (3 cm maximal associated small bowel dilation or the inability of a colonoscope to pass).

[2]As part of the consensus review process 108 potential MR enterography exams were re-reviewed by 2 senior radiologists (JG and MEB) at their respective institutions to ensure that there was no potential adhesive disease, prior strictureplasty, and/or image quality issues. If issues were identified by the respective central readers, the exam was excluded. Eight patients were ultimately excluded.

[3]Strictureplasty change proximal to neo-TI stricture interfered with measurement of proximal dilation (as detected by central readers and confirmed by senior radiologists not participating in the imaging interpretation).

[4]One patient was excluded during independent assessment (most distal stricture measured at baseline and follow-up were not the same stricture).

Figure 2. Baseline MR enterography exam taken from a 26-year-old male depicting a long segment naïve stricture with marked inflammation and multiple areas of narrowing. Top row: stricture length measurements made in Part 1 by each central reader. The dotted lines indicate 3D spline tool annotations, which are superimposed onto the 2D image and correspond to the location of the stricture lumen on images that are outside of the imaging plane. Central readers

created markings along the lumen of the stricture from the proximal to the distal end. Bottom row: maximal associated small bowel dilation measurements made in Part 1 by each central reader.

Figure 3. Baseline MR enterography exam taken from a 39-year-old female with a short naïve stricture with inflammation involving the ileocecal valve and terminal ileum. Left box: coronal true-FISP images with fat saturation (the white arrows point to a single stricture that was assessed by all 4 central readers). Right box, top row: variability in pulse sequences selected for length measurements and annotations from each of the 4 central readers in Part 1. Right box, bottom row: variability in pulse sequences selected for maximum small bowel dilation measurements and annotations from 1 of the 4 central readers. Stricture length, maximal associated small bowel dilation, and visual analog scale evaluations made by each central reader are presented below the annotations. Stricture severity was rated on a visual analog scale where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen). Abbreviations: Max, maximum; VAS, visual analog scale.

Figure 4. Central reader interpretations of 4 MR enterography features in Part 1 using select images taken from a 26-year-old male: (A) coronal diffusion-weighted image, (B) coronal fat saturated post-gadolinium image, (C) coronal true-fast image with stead state precession (true-FISP) image (arrows point to sacculation; inset, coronal single-shot fast spin-echo image showing another sacculation), and (D) coronal single-shot fast spin-echo image with fat saturation.

Figure 5. Central reader measurements for (A) stricture length, (B) maximal associated small bowel dilation, (C) maximal stricture wall thickness, and (D) maximal luminal diameter for each patient in Part 1. The black dots indicate the mean measurement value for each patient. For each variable, patient values are displayed from the smallest to largest. The dotted horizontal line in (A) indicates a stricture length of 5.0 cm. The dotted horizontal line in (B) indicates maximum

small bowel dilation of 3.0 cm. Note that measurement variability in maximal associated small bowel dilation remains relatively constant.

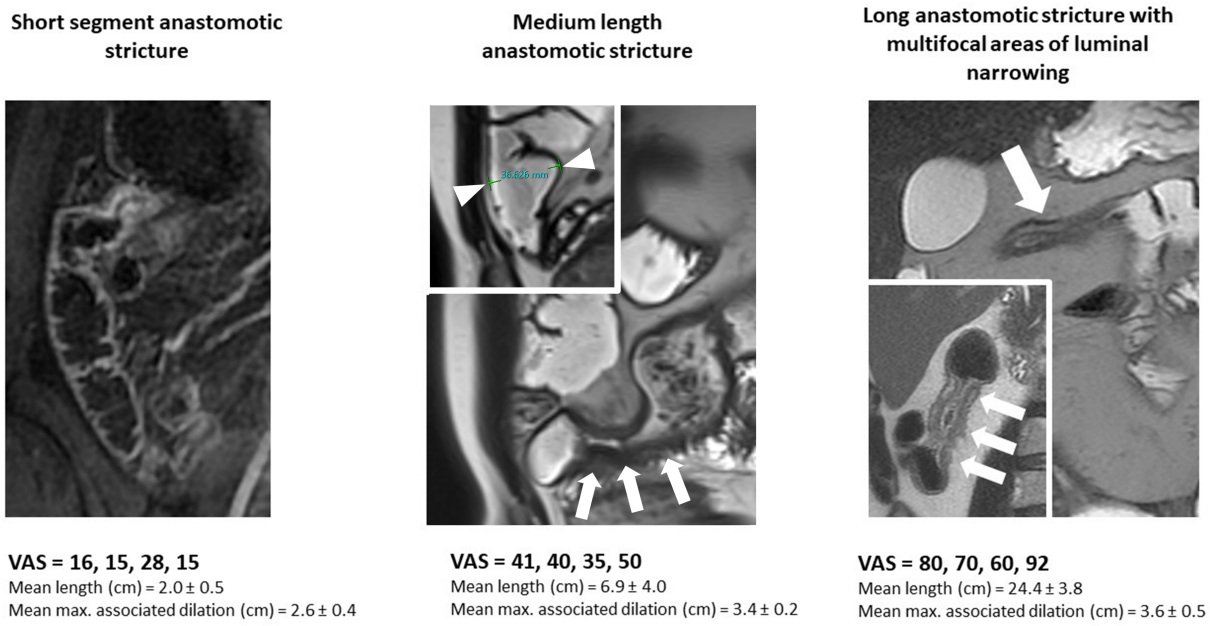
Figure 6. Coronal single-shot fast spin-echo and post-gadolinium images demonstrating the proximal, mid, and distal aspects of a naïve stricture on MR enterography exams taken from a 71-year-old female with a long inflammatory stricture with multi focal areas of luminal narrowing and intervening areas of active inflammation involving the terminal ileum. The central readers measured this stricture as being 70.1 cm, 69.8 cm, 71.7 cm and 71.6 cm in length, respectively. The visual analogue scale estimate of stricture severity was rated by the central readers as 65, 90, 80, and 75, respectively. The white arrows point to a single stricture that was assessed by all readers.

SUPPLEMENTAL MATERIAL

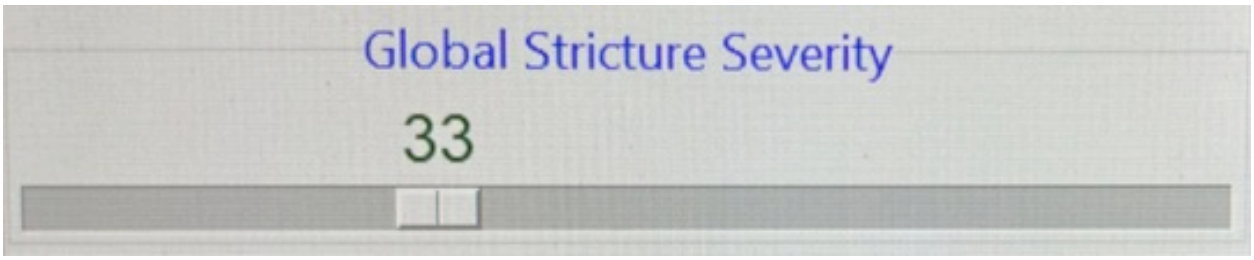
Supplemental Figures

Supplemental Figure 1A. Visual analog scale of stricture severity scores[1] for 3 representative cases. Corresponding mean (SD) values for stricture length and maximum small bowel dilation are also provided. Abbreviations: VAS, visual analog scale; max., maximum.

[1]Stricture severity was rated on a VAS where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen). Central readers provided a global assessment of stricture severity by placing a mark on the horizontal line consistent with their evaluation.

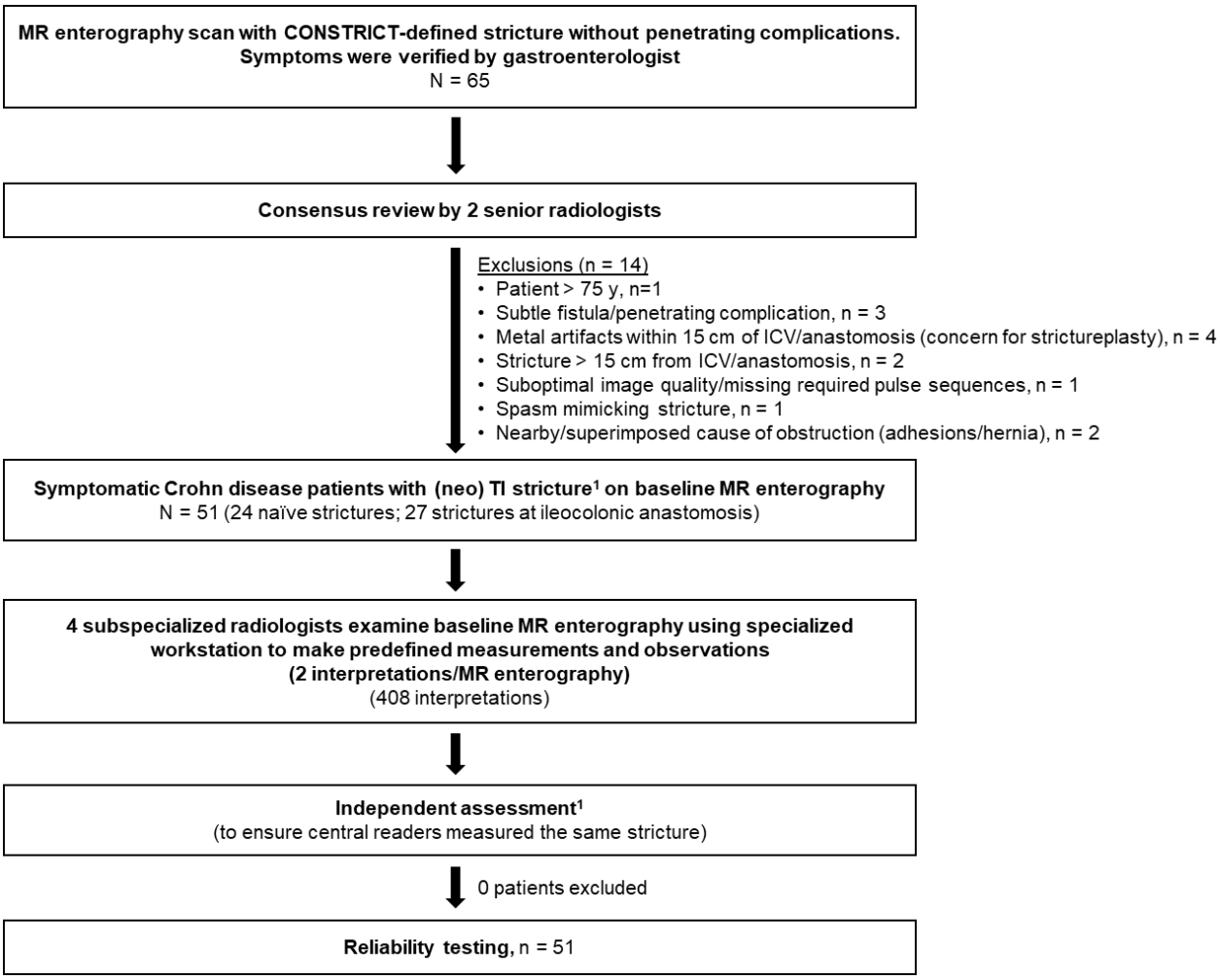


Supplemental Figure 1B. The slider bar used by central readers to rate global stricture severity on a visual analog scale (VAS) from 1 to 100. Instructions on how to score the VAS were provided in the Training Manual (**Supplemental Appendix 1**).



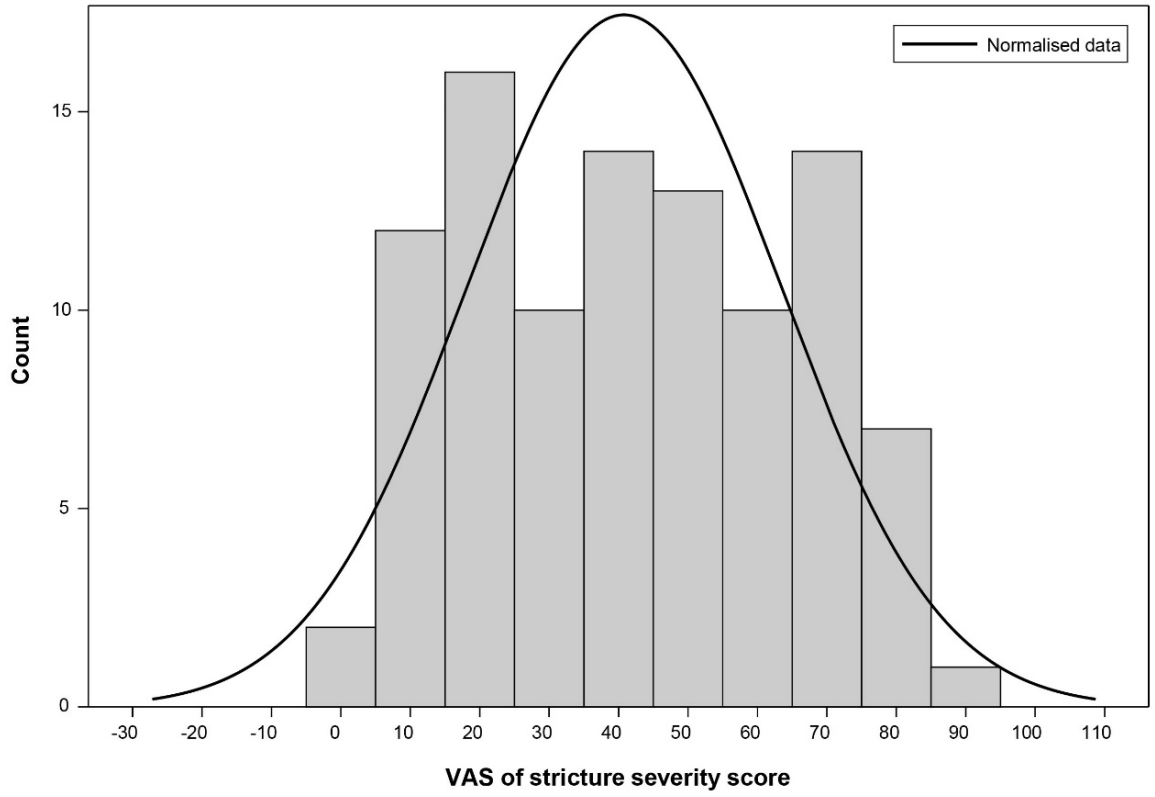
Supplemental Figure 2. Creation of the validation group. Part 2 (validation phase) included a single MR enterography exam from 51 patients with Crohn disease who had symptomatic terminal ileal strictures. Abbreviations: ICV, ileocecal valve; TI, terminal ileal.

[1]Independent assessment for the validation group was performed by the senior radiologist readers (JG and MEB) using database review to ensure there was spatial overlap in the strictures measured by the central readers. No MR enterography exams were excluded due to a lack of overlap.



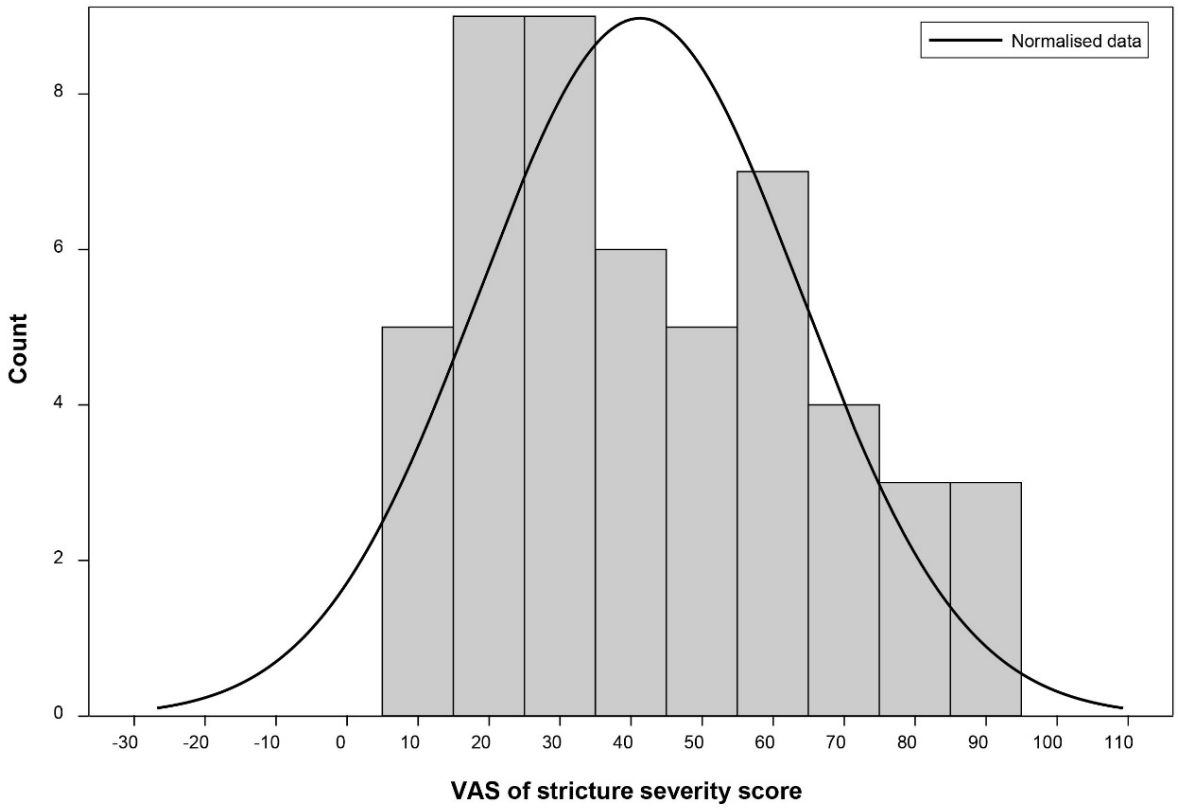
Supplemental Figure 3. Distribution of visual analog scale of stricture severity scores[1] in the development group. Abbreviation: VAS, visual analog scale.

[1]Stricture severity was rated on a VAS where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen). Central readers provided a global assessment of stricture severity by placing a mark on the horizontal line consistent with their evaluation.



Supplemental Figure 4. Distribution of visual analog scale of stricture severity scores[1] in the validation group. Abbreviation: VAS, visual analog scale.

[1]Stricture severity was rated on a VAS where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen). Central readers provided a global assessment of stricture severity by placing a mark on the horizontal line consistent with their evaluation.



Supplemental Tables

Supplemental Table 1. CONSTRICT stricture criteria (13).

Criterion	Definition
Localized luminal narrowing	Luminal diameter reduction of at least 50%, measured relative to the normal adjacent bowel loop
Bowel wall thickening	A 25% increase in wall thickness relative to the adjacent nonaffected bowel
Pre-stricture dilation	Luminal diameter greater than 3 cm

Supplemental Table 2. Stricture characteristics for the development and validation groups[a]

Characteristic	Development group (N=99)[b]	Validation group (N=51)
Number of strictures, n (%)		
1	96 (97.0)	50 (98.0)
2	3 (3.0)	1 (2.0)
Stricture type, n (%)		
Naïve	55 (55.6)	24 (47.1)
Anastomotic	44 (44.4)	27 (52.9)
Stricture length (cm)		
Mean (SD)	12.6 (11.6)	14.3 (13.3)
Median (IQR)	9.2 (4.3, 16.3)	12.0 (4.2, 19.3)
Stricture distance to ileocecal valve/anastomosis (mm)		
Mean (SD)	5.3 (21.8)	5.1 (13.5)
Median (IQR)	0 (0, 0)[c]	0 (0, 0)[c]
Minimum luminal diameter within stricture (mm)		
Mean (SD)	3.0 (1.9)	2.8 (1.3)
Median (IQR)	2.6 (1.8, 3.6)	2.6 (2.0, 3.4)
Maximal luminal diameter within stricture (mm)		
Mean (SD)	14.6 (10)	16.0 (10.6)
Median (IQR)	12.2 (7.1, 21.8)	12.1 (7.4, 24.0)
Luminal diameter of normal-appearing ileal loop (mm)		
Mean (SD)	18.7 (3.8)	NA
Median (IQR)	18.1 (16.1, 20.8)	NA
Percentage of stricture with luminal narrowing, n (%)		
<50%	11 (11.1)	6 (11.8)
50%-75%	82 (82.8)	45 (88.2)
>75%	6 (6.1)	0 (0)
Maximal stricture wall thickness (mm)[d]		
Mean (SD)	9.3 (2.3)	8.7 (2.2)
Median (IQR)	9.1 (8.0, 10.8)	8.5 (7.1, 10.0)
Wall thickness of normal-appearing adjacent ileal loop distended with enteric contrast (mm)		
Mean (SD)	2.3 (0.6)	NA
Median (IQR)	2.3 (1.9, 2.7)	NA
Maximal associated small bowel dilation (mm)		
Mean (SD)	32 (9.0)	38.0 (10.8)
Median (IQR)	30.3 (26.8, 35.9)	35.0 (30.5, 42.8)

Intramural T2 signal on fat-suppressed images[e], n (%)		
Absent	8 (8.1)	3 (5.9)
Mild	47 (47.5)	22 (43.1)
Moderate	37 (37.4)	24 (47.1)
Severe	7 (7.1)	2 (3.9)
Mural hyperenhancement, n (%)		
Absent	1 (1.0)	0 (0)
Mild	13 (13.1)	13 (25.5)
Moderate	38 (38.4)	27 (52.9)
Severe	47 (47.5)	11 (21.6)
Luminal ulcerations, n (%)		
Absent	53 (53.5)	23 (45.1)
Present	46 (46.5)	28 (54.9)
Pattern of enhancement: portal/enteric phase, n (%)		
Stratified	22 (22.2)	NA
Homogenous	25 (25.3)	NA
Luminal (inner wall only)	52 (52.5)	NA
Pattern of enhancement: delayed gadolinium phase[f] , n (%)		
Homogenous	47 (73.4)	34 (66.7)
Not homogenous	17 (26.6)	17 (33.3)
Intensity of delayed gadolinium phase compared to portal/enteric phase[f], n (%)		
Equal to portal	19 (29.7)	NA
Greater than portal	15 (23.4)	NA
Less than portal	30 (46.9)	NA
Intramural fat, n (%)		
Absent	12 (12.1)	NA
Present	87 (87.9)	NA
Sacculation, n (%)		
Absent	68 (68.7)	27 (52.9)
Present	31 (31.3)	24 (47.1)
Whiskering/serosal wall irregularity, n (%)		
Absent	68 (68.7)	NA
Present	31 (31.3)	NA
Fibrofatty proliferation, n (%)		
Absent	40 (40.4)	NA
Present	59 (59.6)	NA
Peri-enteric fat stranding, n (%)		
Absent	75 (75.8)	30 (58.8)
Present	24 (24.2)	21 (41.2)
Peri-enteric fat edema or free fluid, n (%)		

Absent	59 (59.6)	NA
Present	40 (40.4)	NA
Vasa recta distension, n (%)		
Absent	30 (30.3)	25 (49.0)
Present	69 (69.7)	26 (51.0)
Regional enlarged lymph nodes, n (%)		
Absent	71 (71.7)	NA
Present	28 (28.3)	NA
Restricted diffusion (DWI)[g], n (%)		
Absent	3 (4.9)	19 (37.3)
Present	58 (95.1)	32 (62.7)
Visual analog scale[h]		
Mean (SD)	40.8 (22.6)[h]	41.3 (20.0)
Median (IQR)	45.0 (30.0, 65.0)	37.5 (20.0, 60.0)

Abbreviations: DWI, diffusion-weighted imaging; IQR, interquartile range; NA, not applicable.

[a]Stricture characteristics for the development and validation based on the central reader assessments.

[b]Data missing for 1 patient in the development group; see **Figure 1**.

[c]There were 92 patients in the development group and 44 patients in the validation group who had a stricture abutting the ileocecal valve or anastomosis (ie, the distance between the stricture and the ileocecal valve or anastomosis was 0 cm).

[d]The average was taken from 3 measurements.

[e]A surrogate was used if intramural T2 signal on fat-suppressed images were unavailable or fat suppression was inadequate.

[f]Delayed gadolinium images were not available for 35 patients.

[g]DWI images were not available for 38 patients.

[h]Score at follow-up.

Supplemental Table 3. Intra-rater reliability in the development phase according to central reader.

Feature	Reader 1	Reader 2	Reader 3	Reader 4
MR enterography imaging measurements				
Stricture length	0.92	0.93	0.91	0.92
Maximal associated small bowel dilation	0.81	0.80	0.76	0.79
Maximal stricture wall thickness	0.70	0.63	0.66	0.75
Maximal luminal diameter within stricture	0.70	0.69	0.49	0.66
Luminal diameter of normal appearing ileal loop	0.53	0.28	0.36	0.49
Minimum luminal diameter within stricture	0.63	0.34	0.06	0.44
Stricture distance to ileocecal valve/anastomosis	0.60	0.55	0.84	0.85
Wall thickness of normal loop	0.47	0.01	0.12	0.19
MR enterography imaging observations				
Stricture type	0.80	0.71	0.63	0.84
Pattern of enhancement: delayed gadolinium phase	0.59	0.66	0.69	0.50
Restricted diffusion (DWI)	0.67	0.00	0.44	0.05
Sacculation	0.67	0.53	0.47	0.75
Mural hyperenhancement	0.67	0.61	0.46	0.56
Perienteric fat edema or free fluid	0.74	0.31	0.40	0.73
Percentage of stricture with luminal narrowing	0.44	0.29	0.48	0.61
Vasa recta distension	0.62	0.41	0.29	0.66
Regional enlarged lymph nodes	0.53	0.57	0.42	0.66
Fibrofatty proliferation	0.66	0.56	0.25	0.69
Whiskering/serosal wall irregularity	0.45	0.48	0.44	0.56
Intramural T2 signal on fat-suppressed images	0.58	0.47	0.36	0.73
Luminal ulcerations	0.57	0.43	0.38	0.61
Perienteric fat stranding	0.58	0.34	0.26	0.47
Intramural fat	0.70	0.17	0.23	0.46
Pattern of enhancement: portal/enteric phase	0.22	0.09	0.40	0.41
Intensity of delayed gadolinium compared to portal/enteric	0.29	0.31	0.20	0.70
Visual analog scale[a]				
Stricture severity	0.71	0.82	0.79	0.64

Abbreviations: DWI, diffusion weighted imaging.

[a]Stricture severity was rated on a visual analog scale where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen).

Supplemental Table 4. Intra-rater reliability in the validation phase according to central reader.

Feature	Reader 1	Reader 2	Reader 3	Reader 4
MR enterography imaging measurements				
Stricture length	0.94	0.95	0.96	0.97
Maximal associated small bowel dilation	0.75	0.97	0.91	0.80
Maximal stricture wall thickness	0.66	0.42	0.85	0.84
Maximal luminal diameter within stricture	0.50	0.86	0.58	0.72
MR enterography imaging observations				
Stricture type	0.75	0.96	0.88	0.88
Pattern of enhancement: delayed gadolinium phase	0.53	0.77	0.60	0.72
Restricted diffusion (DWI)	0.62	1.00	0.57	0.00
Sacculation	0.62	0.97	0.78	0.73
Mural hyperenhancement	0.31	0.73	0.68	0.73
Vasa recta distension	0.42	0.81	0.72	0.68
Intramural T2 signal on fat-suppressed images	0.37	0.75	0.75	0.76
Luminal ulcerations	0.69	0.71	0.59	0.30
Perienteric fat stranding	0.56	0.77	0.60	0.73
Visual analog scale[a]				
Stricture severity	0.70	0.92	0.85	0.86

Abbreviations: DWI, diffusion weighted imaging.

[a]Stricture severity was rated on a visual analog scale where scoring was performed using a 100 mm continuous horizontal line that ranged from 0 (completely normal) to 100 (worst disease ever seen).

Supplemental Table 5. Responsiveness of radiographic features with at least moderate inter-rater reliability in the development group (N=98[a]).

Variable	AUC (95% CI)	Odds Ratio Effect Size (95% CI)
Maximal associated small bowel dilation	0.59 (0.47, 0.70)	0.97 (0.92, 1.01)
Pattern of enhancement: delayed gadolinium phase	0.58 (0.47, 0.69)	2.1 (0.7, 6.1)
Sacculation	0.56 (0.48, 0.65)	0.55 (0.24, 1.26)
Stricture length	0.55 (0.44, 0.66)	1.00 (0.99, 1.00)
Stricture type	0.54 (0.47, 0.61)	1.7 (0.6, 4.9)
Restricted diffusion (DWI)	0.54 (0.47, 0.61)	3.0 (0.3, 29.5)
Mural hyperenhancement	0.54 (0.43, 0.64)	0.86 (0.52, 1.44)
Maximal luminal diameter within stricture	0.52 (0.41, 0.64)	0.99 (0.95, 1.02)
Maximal stricture wall thickness	0.48 (0.36, 0.59)	0.99 (0.85, 1.15)

Abbreviations: AUC, area under the receiver operating curve. DWI, diffusion weighted imaging. AUC values were interpreted using Cohen's effect sizes (0.56=small; 0.64=medium; 0.71=large).

[a] Data missing for 2 patients in the development group; see **Figure 1.**

Supplemental Appendix 1. MR enterography scanning technique.

All MR enterography scans were performed as part of routine clinical care according to existing standardized clinical protocols used by the Mayo Clinic and Cleveland Clinic Foundation at the time of exam. All included exams complied with technical parameters published by the Society of Abdominal Radiology (SAR) Crohn Disease- focused Panel (36). At least 900 mL of biphasic enteric contrast was orally administered in divided doses 40 to 60 minutes before MRE image acquisition. Motion-robust T2-weighted pulse sequences (single-shot fast spin- echo [SSFSE or HASTE] and/or true fast imaging with steady state precession [True-FISP or FIESTA]) were obtained with and without fat saturation in addition to post-gadolinium, fat-saturated, fast spoiled gradient echo imaging of the small bowel, which was performed in multiple planes to include the abdomen, pelvis, and perineum (including the anal canal) using either 1.5 or 3.0 T clinical MRI systems. Glucagon was administered subcutaneously or intravenously to inhibit small bowel motion during contrast enhancement. Contrast-enhanced imaging was initiated between the enteric and portal phases of enhancement (ie., 50-70 seconds after injection of intravenous contrast) with imaging in axial and coronal planes performed with slice thicknesses less than or equal to 5.0 mm.

Image quality review prior to study inclusion confirmed the use of required pulse sequences, 5.0 mm or less slice thicknesses, and acceptable image quality. Review of individual pulse sequence parameters was not performed. Overall image quality was rated as non-evaluable, poor, fair, and good. Specific image quality aspects, including peristaltic motion artifacts, poor contrast enhancement, noisy images, poor bowel distention, were also evaluated. Image quality review was performed by JGF and MEB, who determined whether or not an MR enterography exam was adequate for central reader review. Pre-contrast, delayed post- gadolinium 7-minute imaging and diffusion weighted images were not required, but included with subsequent image review, when available.

Supplemental Appendix 2. STAR Central Reader Training Manual.

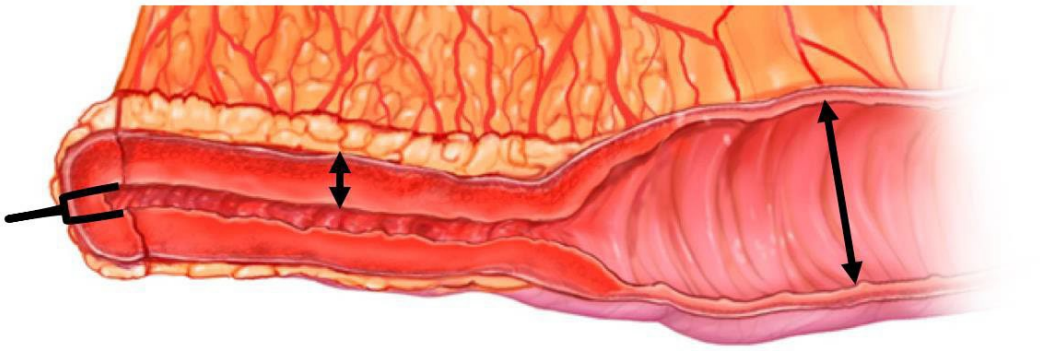
Revisions to the original Training Manual are delineated with an underscore.

This manual was provided to and reviewed by radiology central readers prior to in-person training and testing for both the derivation and validation cohorts. Underlined items represent updates made to the initial training manual after the validation study, based on feedback and questions from central readers/co-investigators after evaluating the derivation cohort results.

I. **Identify ileal stricture(s)** that extend within 15 cm of the ileocecal valve or ileocolic anastomosis (i.e., the distal margin of the stricture is within 15 cm of ICV/anastomosis)

   a. **Load coronal SSFSE/Gd images first as a global preview** (the rationale is that SSFSE is high quality across manufacturers, and short segment strictures are highlighted on post-Gd sequences)

      i. **Morphologic diagnosis:** wall thickening, luminal narrowing, and upstream dilation. CONSTRICT and STAR criteria specifically require:
         1. Wall thickening of at least 25% compared with nearby normal bowel (implies 4 mm thickness at minimum)
         2. Luminal narrowing of at least 50% compared to nearby, non- obstructed distal small bowel
         3. Unequivocal proximal small bowel dilation on MRE. If the stricture cannot be traversed with a colonoscope, associated maximum small bowel dilation should be unequivocal and 2.5 cm or greater (e.g., present on multiple sequences). If endoscopic criteria for stricture identification have not been met, associated maximum small bowel dilation on MRE should be 3 cm on any sequence.

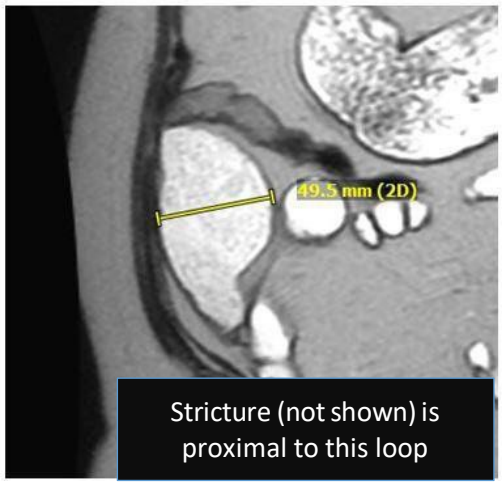


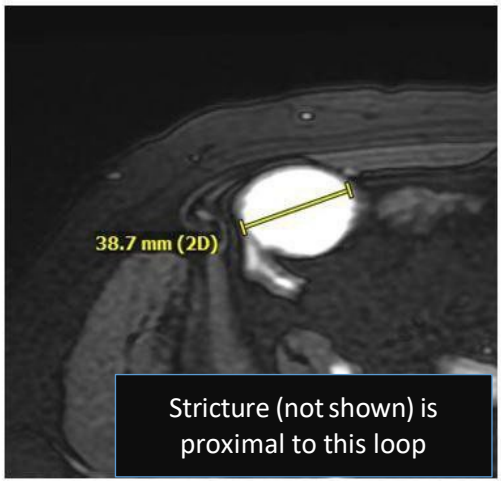
Narrowing with
proximal bowel dilation

         4. For this research study, some exams may be from follow-up, and some may be from baseline exams in patients with strictures that could not be traversed with an endoscope, so proximal small bowel dilation (of ≥ 3 cm or any degree) may or may not be present (e.g., a stricture with inflammation treated with a TNF- inhibitor may get shorter and still possess wall thickening and luminal narrowing, but without proximal small bowel dilation).

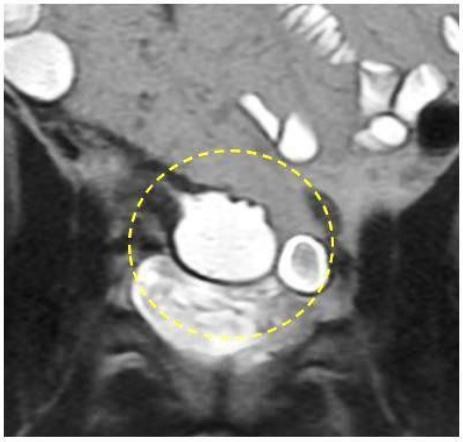
5. As this study seeks to track the evolution of such morphological and signal changes over time, mark at least 1 lesion with wall thickening and luminal narrowing within 15 cm of small bowel per case. Baseline exams may have up to 2 strictures and follow-up exams may have up to 3 strictures within the 15 cm of the distal small bowel that is to be analyzed. Please see **Section II** for what defines a single stricture.

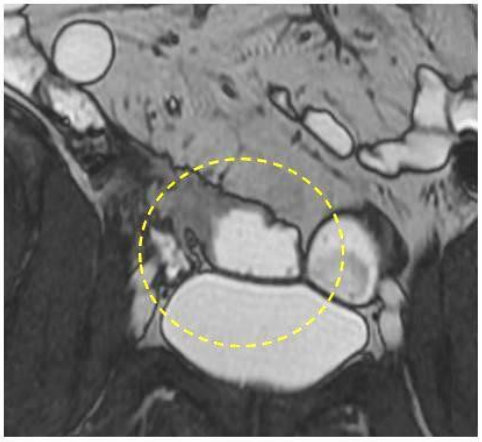
ii. For **max associated small bowel dilation (labeled proximal small bowel dilation above)**, the coronal SSFSE and post-Gd sequences are generally best, but also evaluate axial true-FISP or HASTE images.





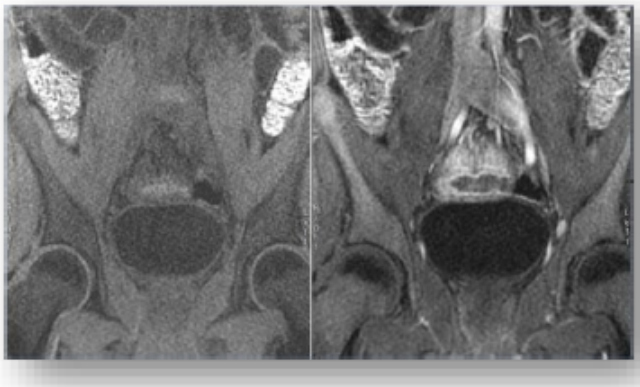
1. Measurement of **max associated small bowel dilation** is from mid-bowel wall to mid-bowel wall, measured orthogonally to luminal centerline or axis of the small bowel (see above).
2. Choose the maximum prestenotic dilation in any plane or sequence, in which the diameter is reasonably orthogonal to the lumen. Practically, the greatest degree of prestenotic distension is often on the HASTE sequence.
    a. Dilated prestenotic bowel is often oval-shaped and not round. Measure the longest linear dimension.
    b. It is normal for **max associated small bowel dilation** to change over the course of an exam as enteric contrast travels to the colon. Make sure to measure the greatest degree of prestenotic dilation.



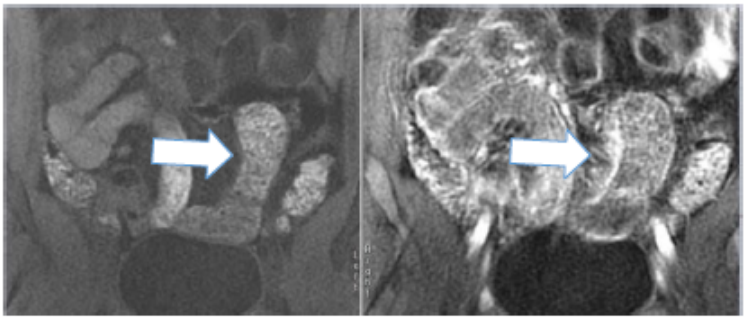


- Some strictures will have 2 or more areas of severe narrowing within the stricture, and the **max associated small bowel dilation** may be within the stricture, rather than proximal to the stricture, if the dominant area of narrowing is located distally within the stricture (see Figure below: "What constitutes a single stricture?").
- **Max associated small bowel dilation *CANNOT*** be measured in an enteric anastomosis (e.g., side-to-side ileoileostomy), as this will affect the diameter and not reflect the obstructive component.
- Remember that the lumen may be filled with feces or unopacified succus rather than high signal fluid and thus can appear dark in T2-weighted imaging.

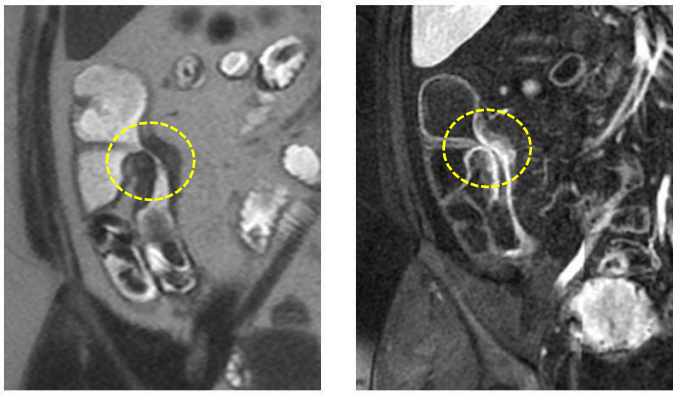
Images show a distal ileal stricture with inflammation without and with contrast (note comb sign) in the top row. The bottom row shows the distal ileum proximal to stricture. Note the high signal succus filling lumen. For lumen measurement, measure the lumen filled with high signal succus.

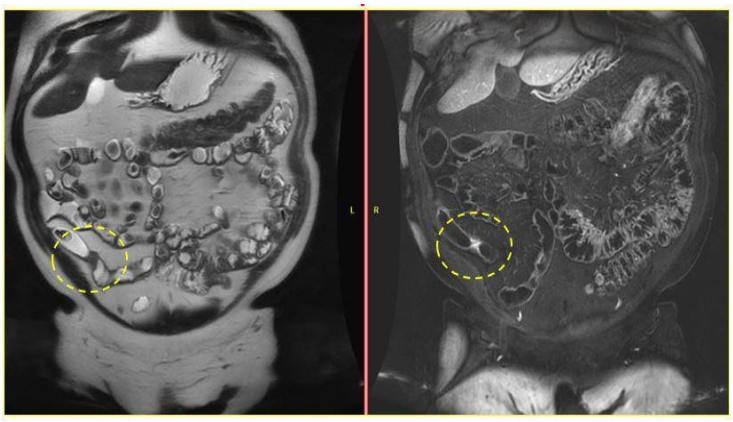


High signal enteric contents can also make interpretation of Gd enhancement challenging. Make sure to measure using bowel wall on contrast-enhanced images, especially for maximum associated small bowel dilation.

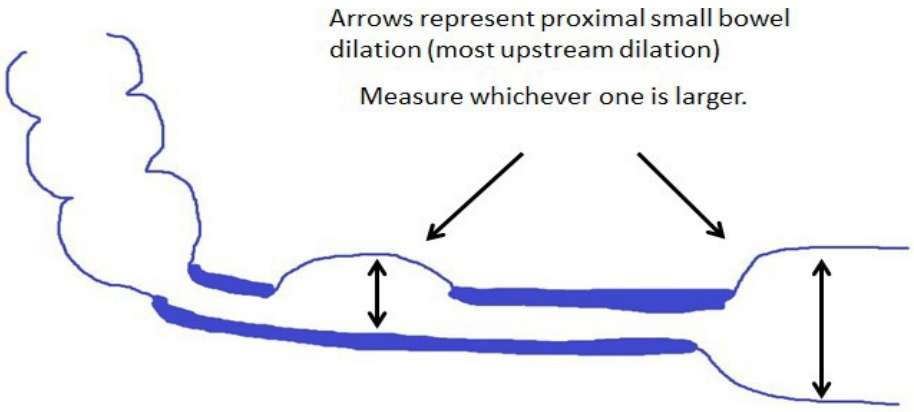


Some short segment strictures are difficult to see when the lumen is obscured and may appear as 2 adjacent bowel loops. **Gd sequences** are often helpful in this scenario when a connection between the (neo)terminal ileum and colon cannot be appreciated on T2- weighted imaging, as strictures with an inflammatory component will result in avid enhancement.





If there is a long stricture with continuous disease along the mesenteric border only, or single inflammatory lesion with multiple regions of luminal narrowing (***please see definition of single vs. multiple strictures below***), select the location to measure proximal small bowel dilation in the location with the most upstream dilation (i.e., the most dilated bowel, potentially in the middle of the lesion). For this study, this may mean that the most dilated proximal bowel is in an area of sacculation.



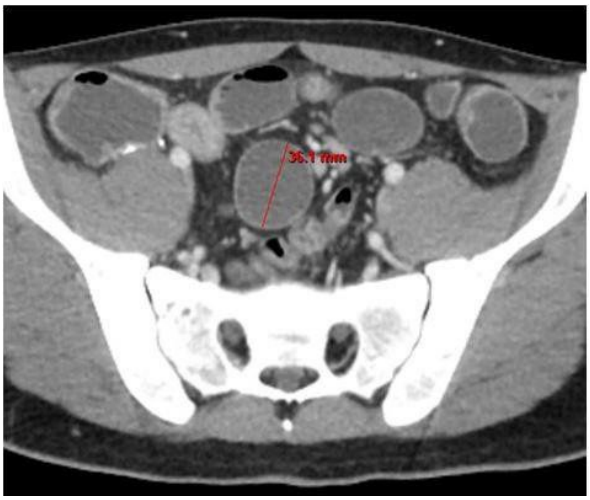
Continuous inflammation along the mesenteric border and patchy inflammation along the antimesenteric border.

II. **What constitutes single vs. multiple strictures?** Occasionally, it will be difficult to ascertain if single or multiple strictures are present based on areas of luminal narrowing and dilation within the terminal ileum. Several rules have been developed as a guide:

a. ***Continuous inflammation defines a single stricture***, although it may have several areas of discrete luminal narrowing within it. Such strictures will have contiguous but often asymmetric inflammation, but without continuous luminal narrowing. ***The continuous inflammation may be along the mesenteric border only***. Bowel in between segments with narrowed lumen will have wall thickening (often asymmetrically and along the mesenteric border, with anti-mesenteric sacculation).

Please open and view the accompanying PPT. See GIF # 1A – continuous inflammation defines a single stricture

Please open and view the accompanying PPT. See GIF # 1B – continuous inflammation defines a single stricture



b. ***Nearby (≤ 3 cm) adjacency defines a single stricture***. If there are 2 strictures (each with wall thickening and luminal narrowing) separated by ≤ 3 cm of normal-appearing bowel, the pair of strictures should be marked as a single stricture. The 3 cm threshold was chosen because it was felt that (1) this length is reproducible and will allow for clear and separate identification of distinct strictures; (2) adjacent strictures separated by 3 cm or less behave functionally like a single stricture; and (3) nearby adjacent strictures often evolve from or toward a single stricture when more inflammation is present. If there is a long distal ileal segment with discontinuous disease (separated by 3 cm or less), consider these discontinuous strictures as 1 stricture.

Please open and view the accompanying PPT. See GIF #2A – nearby adjacency ≤ 3 cm defines a single stricture
Please open and view the accompanying PPT. See GIF #2B – nearby adjacency ≤ 3 cm defines a single stricture
Please open and view the accompanying PPT. See GIF #2C – nearby adjacency ≤ 3 cm defines a single stricture

III. **Stricture Morphologies** are limited by the inclusion criteria of this study; patients with fistulas on baseline or follow-up exams are excluded, and stable sinus tracts are permitted. Patients with strictureplasty within the distal ileum are also excluded, as are J pouch patients, and patients with ostomies with strictures located in the abdominal wall. Patients with ostomies and intra-abdominal terminal ileal strictures are permitted.

Inclusion criteria also include ≤ 2 strictures at baseline exam that extend to within 15 cm of the ileocecal valve or anastomosis (please see definition of single vs. multiple strictures above). You will only be measuring and examining strictures that extend to within 15 cm of the ileocecal valve or ileocolic anastomosis. (For patients with 2 strictures, the entirety of the proximal stricture does not need to be within the 15 cm boundary—just a portion. If there is a question about 3 strictures within the boundary, carefully consider if you are following the rules for what constitutes a stricture).

For patients without a correlative endoscopy, maximal proximal small bowel dilation was measured at 3 cm by a central reader. In patients in which a colonoscope could not traverse the TI or anastomosis (i.e., stricture defined by CONSTRICT endoscopic criteria), unequivocal proximal SB dilation of any degree was required (e.g., multiple pulse sequences, > 2.5 cm). These inclusion criteria imply that only the following stricture morphologies will be included:

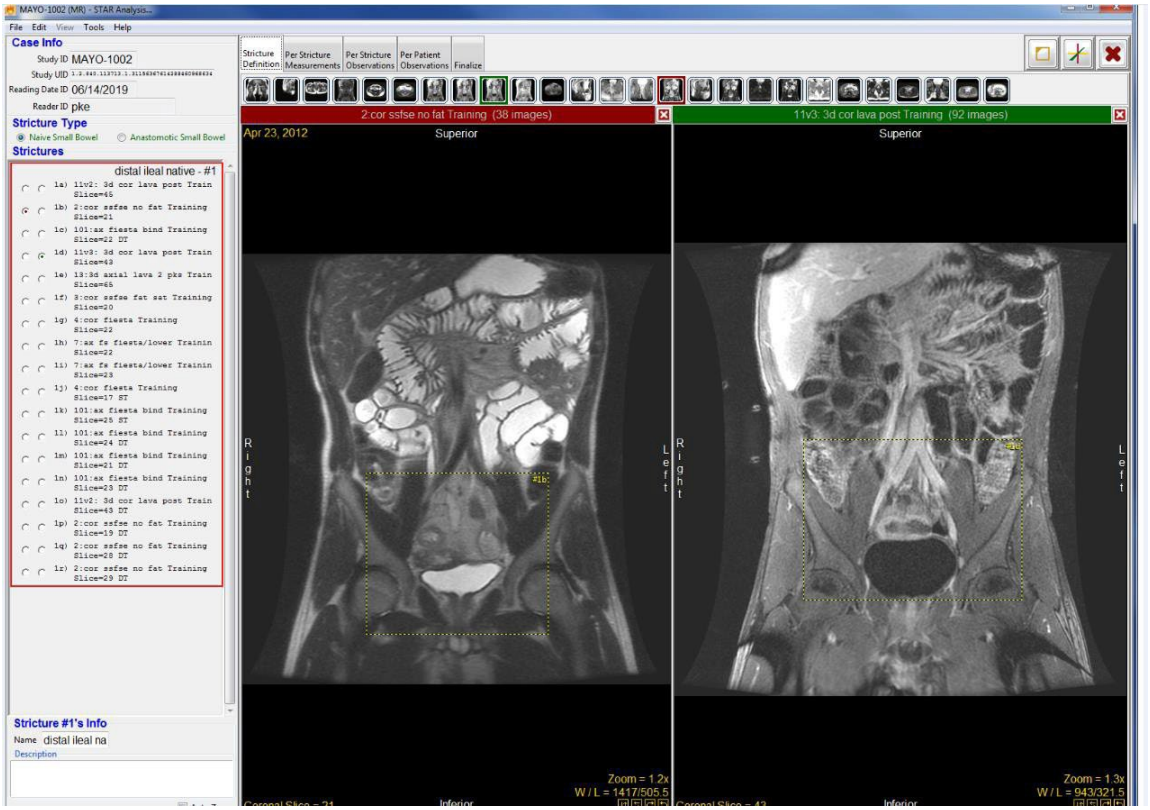
i. Short or long segment inflammatory strictures (but no strictures with penetrating complications other than sinus tracts)
ii. Predominantly fibrotic strictures (i.e., low T2 and DWI signal, enhancement which increases over time, and absence of imaging findings of inflammation)

iii. **State if the stricture is Naïve or Anastomotic.** This definition will be based on whether the ileocecal valve is present or not. Neoterminal ileal strictures that do not abut an anastomosis will be classified as “anastomotic” strictures in this study (i.e., neoterminal ileal strictures are defined by presence of a nearby anastomosis, and not by abutment to the anastomosis). This question is for research purposes only; stricture types will be classified by surgical history in the study.

STRICTURE MEASUREMENTS – The workstation will permit drag-and-drop comparison of pairs of pulse sequences on a diagnostic monitor. Your first task will be to define strictures on multiple pre- and post-contrast pulse sequences for measurement. This will be done by circumscribing a rectangle around the stricture and grouping images circumscribing the same stricture together. Rectangular ROIs circumscribing the same stricture will be grouped. You will be shown how to do this during reader training. At least 1 stricture or stenosing lesion must be defined on every exam. To speed your measurement/evaluation, you will need to circumscribe each stricture within a square. Measurements you will need to take include:

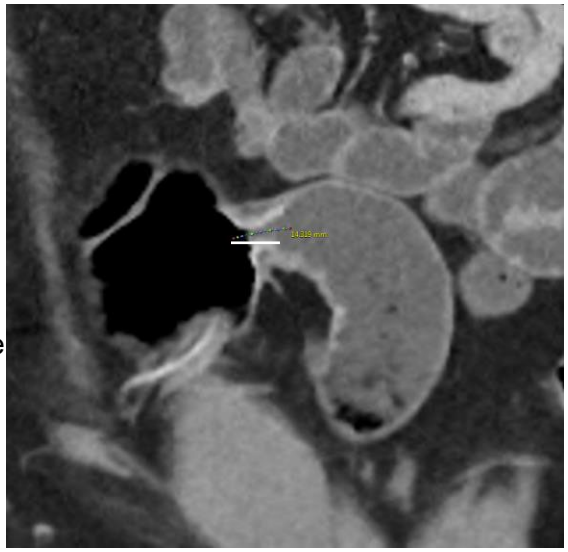
a. Stricture length
b. Stricture distance to the ileocecal valve or anastomosis
c. Narrowest luminal diameter within stricture
d. Maximum luminal diameter within stricture
e. Maximum wall thickness within stricture (#1, #2, #3)
f. Maximum associated small bowel dilation

Begin your evaluation by looking at coronal SSFSE and post-Gd images.



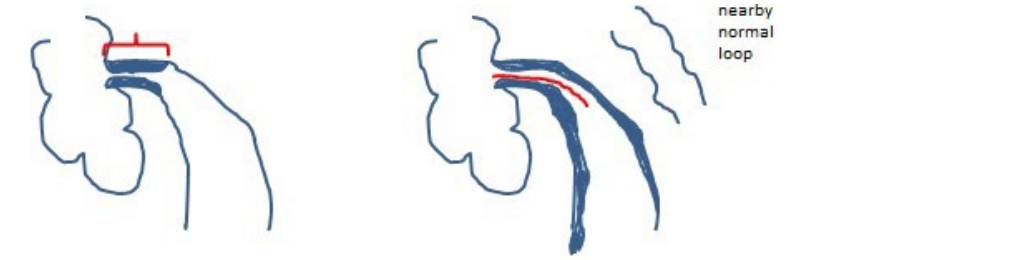
Measurements will be performed in the measurements tab of the workstation.

1. **Measure stricture length** –Strictures start and end where luminal narrowing starts and stops. Consequently, there may be a long segment with active inflammation, but the narrowed regions within it that constitute a single stricture may be shorter (i.e., some of the actively inflamed loops will not be narrowed). Measurements will be performed with a spline tool by paging through images and dropping points in the lumen. Long inflammatory and discontinuous strictures will have intervening areas where the lumen may not be narrowed by 50%. In these strictures, the length will include these non-narrowed segments, but the proximal and distal ends of the stricture will possess luminal narrowing. Based on this definition, it should be realized that the length of enteric inflammation and small bowel thickening may be longer than the stricture length. Alternatively, if very short, do your best (the tool records to 0.1 mm). Do not perseverate/remeasure a stricture that is 2.0 or 3.0 mm in length. Look for focal field artifact from staple lines and focal Gd enhancement to help you identify the difficult-to-find short segment anastomotic stricture.



Please open and view the accompanying PPT. See GIF #3 – Stricture length measurement.

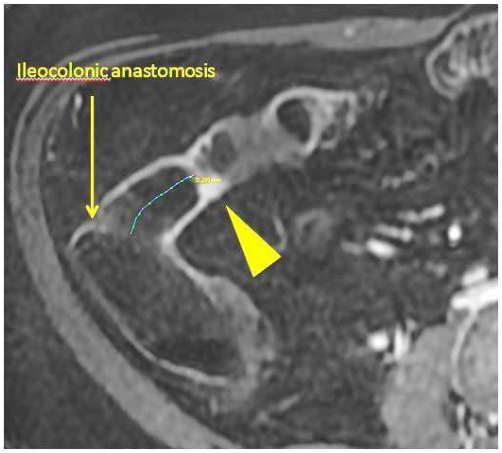
Occasionally, the proximal boundary of a stricture is poorly defined because there is a smooth tapering to the narrowed lumen within the stricture. In this case, try to assign the proximal boundary to an area where the lumen is about 50% of a nearby ileal loop.



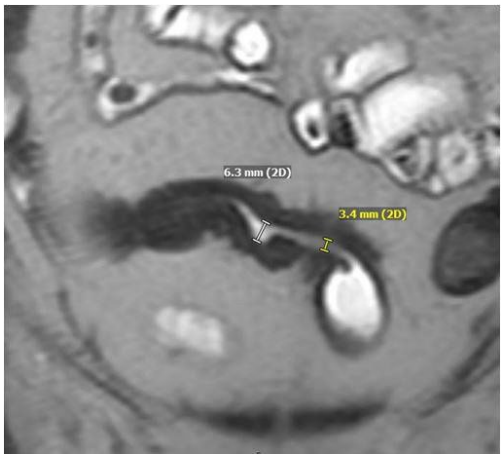
Normal Sharp Boundary

Less Clear Boundary – Smooth Tapering
Define boundary by 50% of width of normal ileum

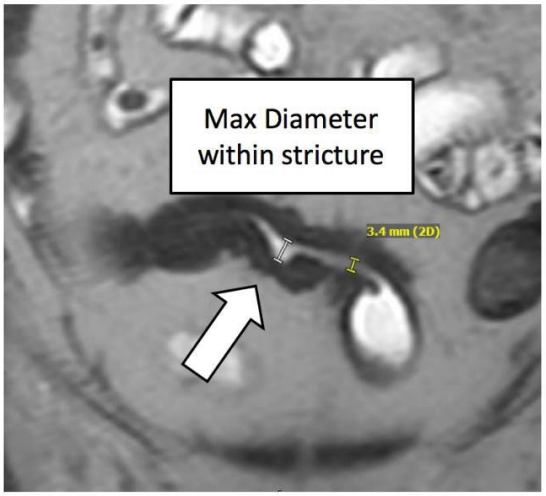
2. **Distance to valve or anastomosis**. You will be asked if the stricture abuts the ileocecal valve or anastomosis. If it does not, measure the distance from the ileocecal valve or anastomosis to the distal end of the stricture.

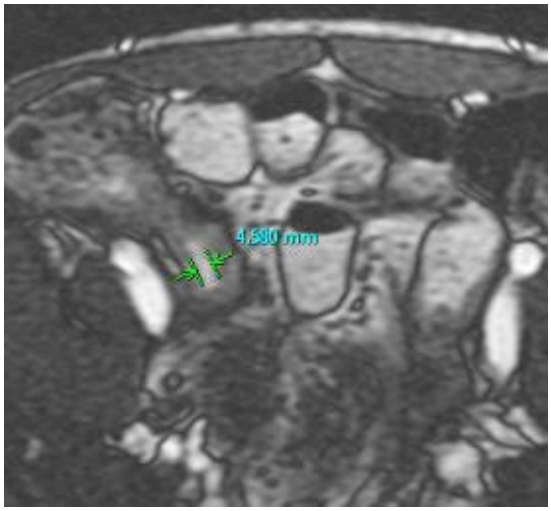


3. **Minimum Stricture luminal diameter** - for very narrow segments, do not perseverate on accuracy if measurement is < 1 mm (just ensure that the measurement is < 1 mm). Do your best. Using the caliper tool, measure luminal diameter from inner wall to inner wall in the most narrowed region. Often the SSFSE or HASTE images are best. Look at the stricture in 2 planes.



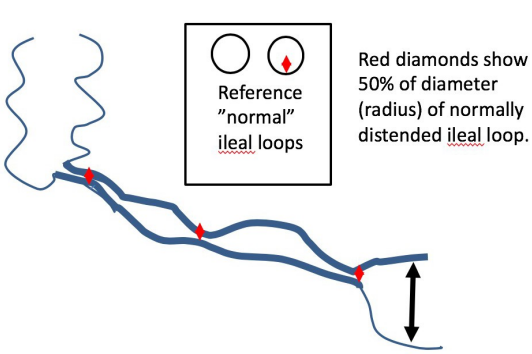
4. **Maximum Stricture luminal diameter** – look at 2 different planes and measure the greatest luminal diameter within the stricture using the caliper tool. If there are 2 nearby strictures, the maximum luminal diameter may be in between the 2 strictures.



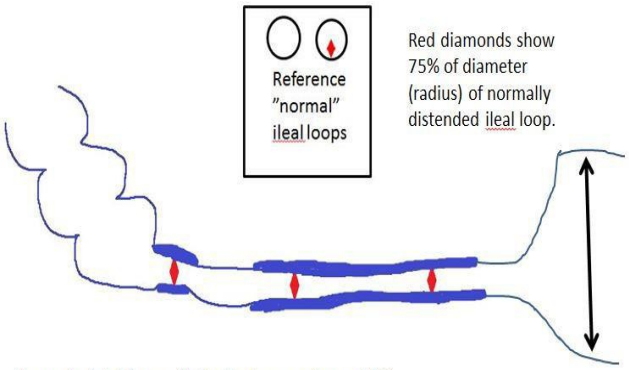


5. **Percent of stricture with luminal narrowing** - What is your best estimate of the percent of the stricture in which the luminal area is < 50% of a normally distended ileal loop? Many strictures will have luminal narrowing throughout the stricture, but discontinuous strictures or those with asymmetric inflammation may contain regions without luminal narrowing.

    a. **Choices: < 50% or ≥ 50%**

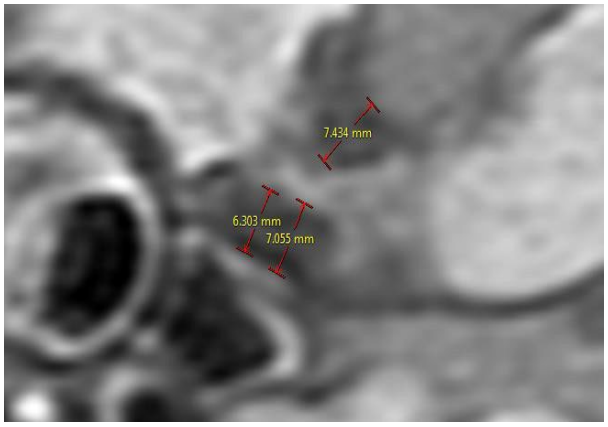


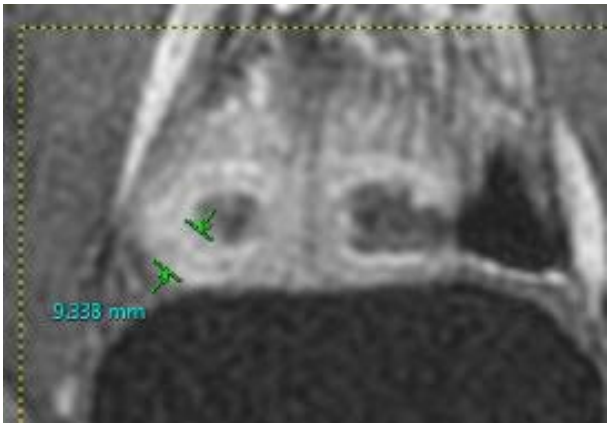
Percent of stricture with luminal narrowing : < 50%



Percent of stricture with luminal narrowing : >75%

6. **Maximum wall thickness of stricture -** look at two different planes and measure the maximal wall thickness within the stricture. Take 3 measurements using the caliper tool. Consider taking measurements on SSFSE or FIESTA images first as these pulse sequences generally have fewer motion artifacts.

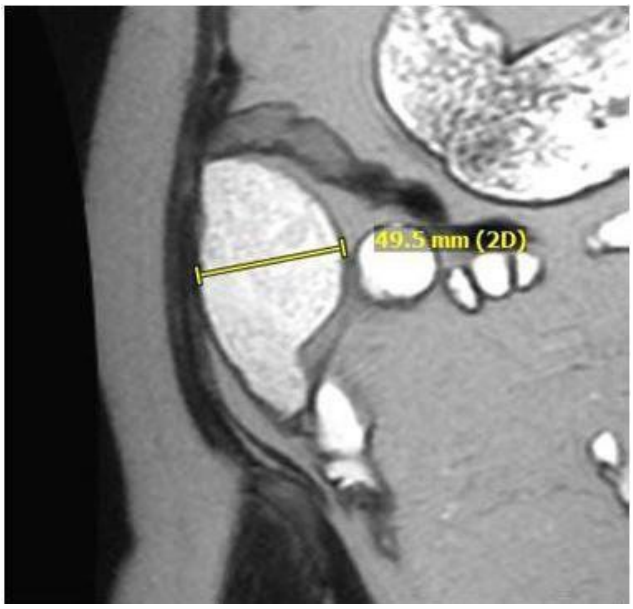




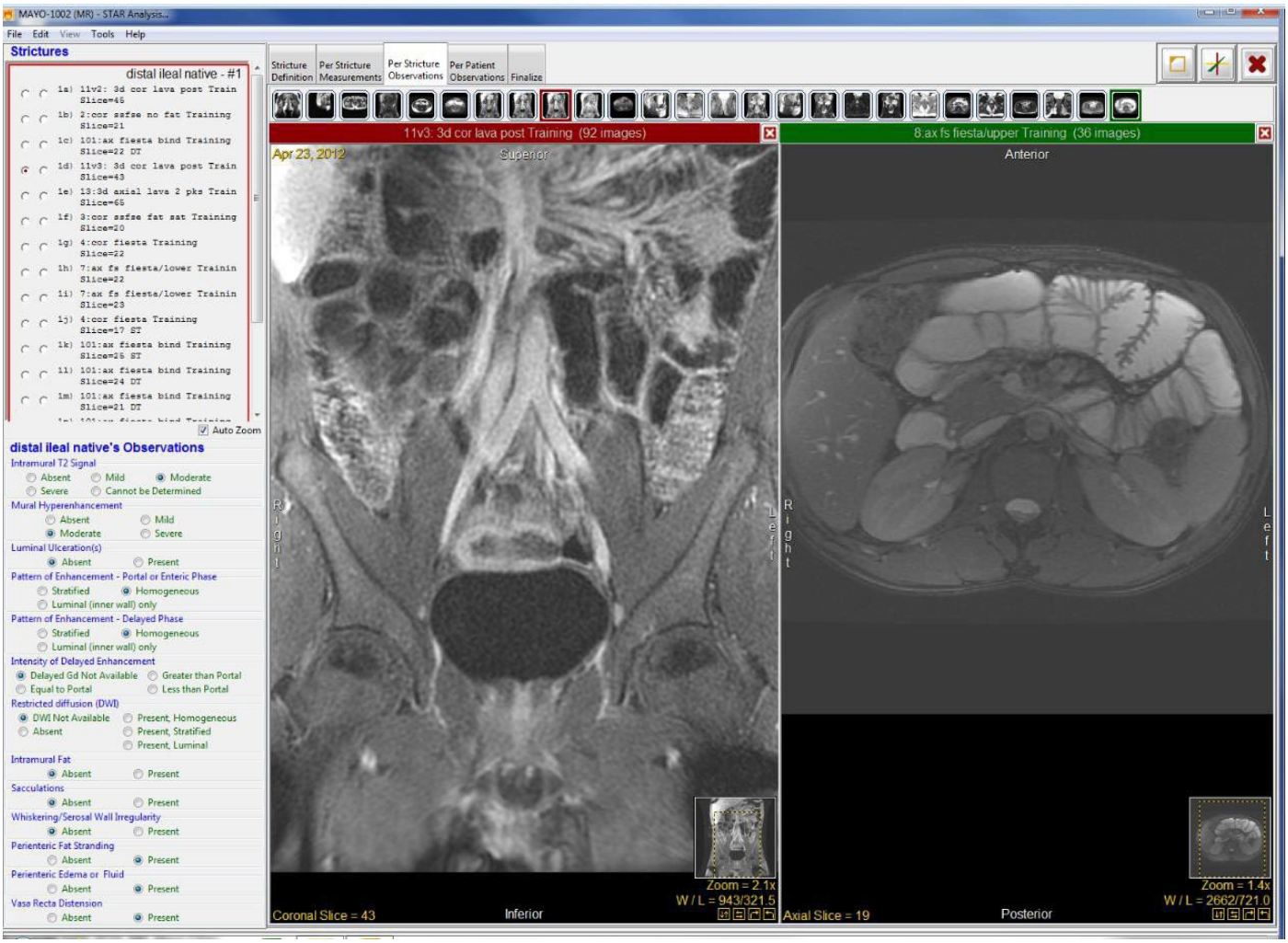
7. **<u>Max associated</u> small bowel dilation** – <u>as described earlier, this is usually proximal to a stricture, but may occur within a stricture</u>. Using the caliper tool, measure orthogonally to the lumen, from mid wall to mid wall. Remember that not all dilated segments are filled with fluid- it may be gas with low signal or succus with high signal. Remember:

iv. **<u>Max associated</u> small bowel dilation**, the coronal HASTE and post- Gd sequences are generally best, but also quickly examine axial true- FISP or HASTE

1. Choose the maximum prestenotic dilation in any plane or sequence. Dilated prestenotic bowel is often oval-shaped. Measure the longest linear dimension.



STRICTURE OBSERVATIONS – These will be made in a separate tab that permit drag-and- drop comparison of pairs of pulse sequences. On the left there will be a menu of observations to be evaluated. *Select the answer that is most representative of the stricture as a whole (i.e., not the worst part, but the most representative answer).* For certain observations that grade intramural T2-signal and enhancement, compare the appropriate image to the laminated images in front of the workstation.



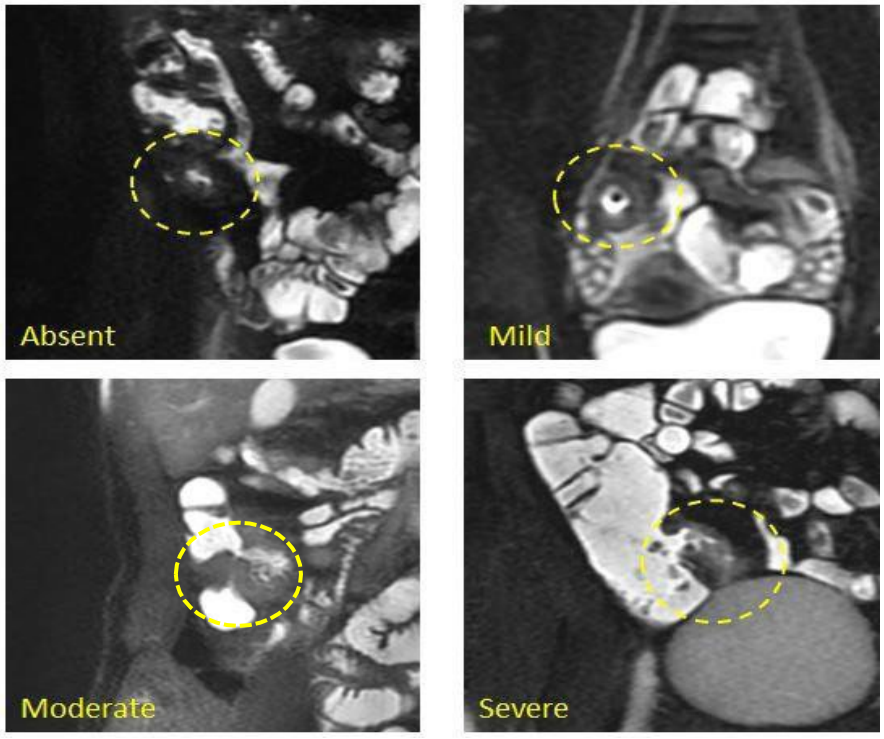
1. **Intramural T2 signal**-edema-fat sat SSFSE fat sat HASTE, HASTE SPAIR present or absent (**DO NOT RATE T2 INTENSITY ON FAT-SAT TRUE-FISP!**)

    a. Score the predominant intramural signal T2 intensity, not the pixel with greatest pixel signal. This selection should be carried out irrespective of the pattern of intramural signal.

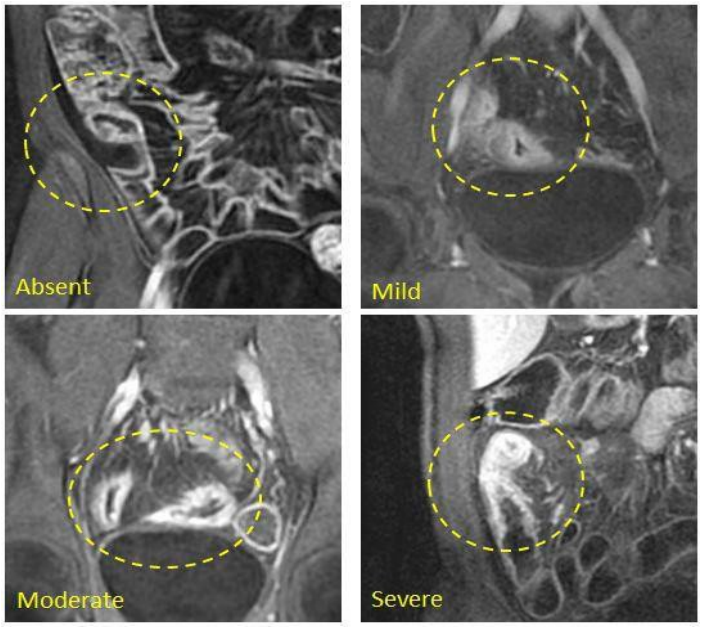
T2 intramural signal intensity varies slightly by the pulse sequence used. Consequently, we want readers to compare the predominant T2 signal to a standard set of images. Pick what you think is the best answer, if both fat-saturated true-FISP and fat-saturated SSFSE are present. **For T2 signal, rate the predominate signal throughout the stricture.**

		Skeletal muscle	Liver/spleen	Vessel
Score	Absent	Mild	Moderate	Severe
Mural T2 Signal	Equivalent to normal bowel wall	Minor increase in signal: bowel wall appears dark grey on fat-saturated images	Moderate increase in signal: bowel wall appears light grey on fat-saturated images	Marked increase in signal: bowel wall contains areas of high signal approaching that of luminal content

Reference images for fat-saturated SSFSE:



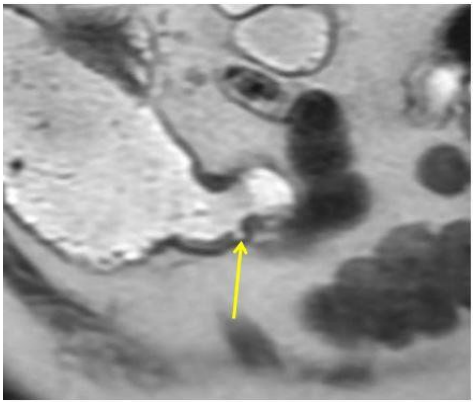
2. **Mural hyperenhancement –** Grade mural hyperenhancement on enteric phase or portal venous phase images. Grade enhancement qualitatively based upon the images supplied by the workstation. Do not consider the pattern of enhancement (e.g., stratified, homogeneous, layered). If there is layered enhancement, hyperenhancement should be graded using the brightest layer of the bowel wall. For interpretation, select the first phase in which the SMV and portal vein are opacified. DO NOT grade delayed sequences, as these will be assessed separately. **Select the most enhancing part of the stricture to grade.**



Mural Enhancement Grades

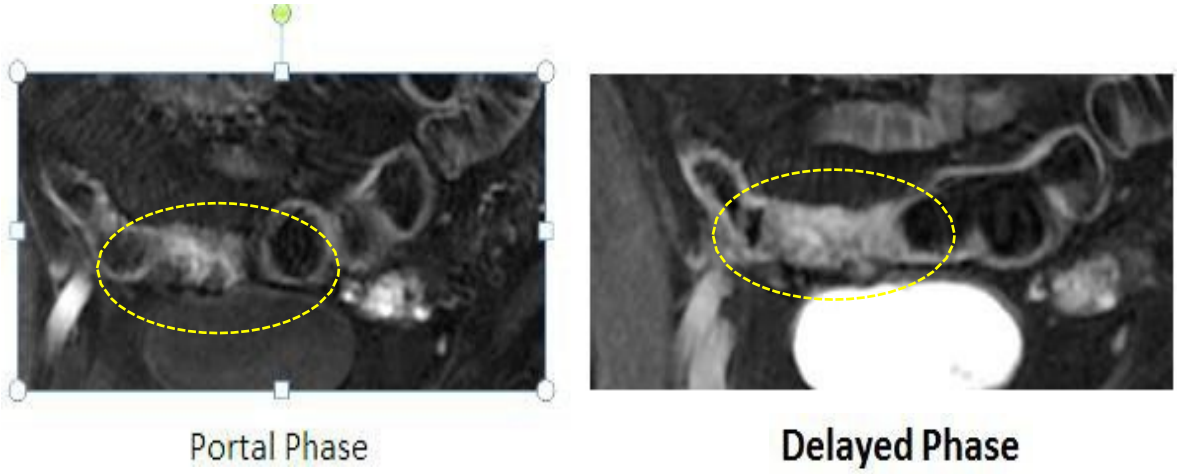
(Enteric Phase)

Score	Absent	Mild	Moderate	Severe
T1 Enhancement	Equivalent to normal bowel wall	Minor enhancement: bowel wall signal greater than normal small bowel but significantly less than nearby vascular structures	Moderate enhancement: bowel wall signal increased but somewhat less than nearby vascular structures	Marked enhancement: bowel wall signal approaches that of nearby vascular structures



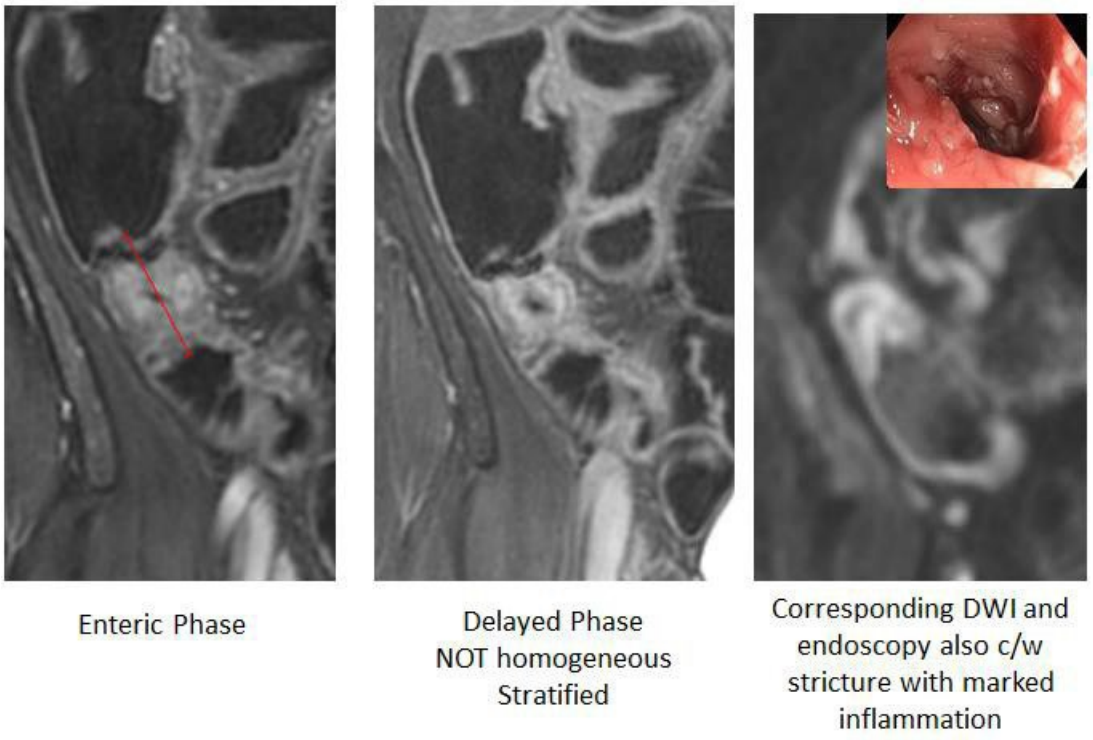
Ulceration – Grade as present or absent. Ulcers are breaks in the luminal wall that extend into the bowel wall and can be filled with air or enteric contrast, and are an important marker of severe inflammation.
Ulcers are most often seen on HASTE or true-FISP images, but are sometimes seen best on post-Gd images. There may be a few sinus tracts, which should not be confused with ulcers.

3. **Pattern of Enhancement - Delayed**: State if delayed Gd images are present or not provided. Otherwise, rate the pattern of enhancement on delayed images as homogeneous vs. not homogeneous (stratified or inner wall only). The example below shows delayed homogeneous enhancement.



Homogeneous delayed enhancement

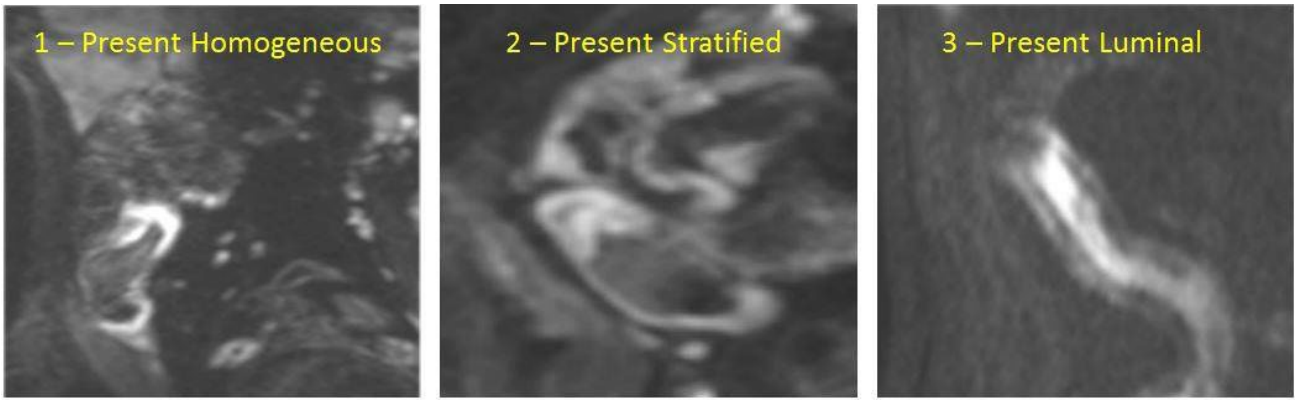
Below is an example of delayed stratified (i.e., not homogeneous) enhancement on delayed images.



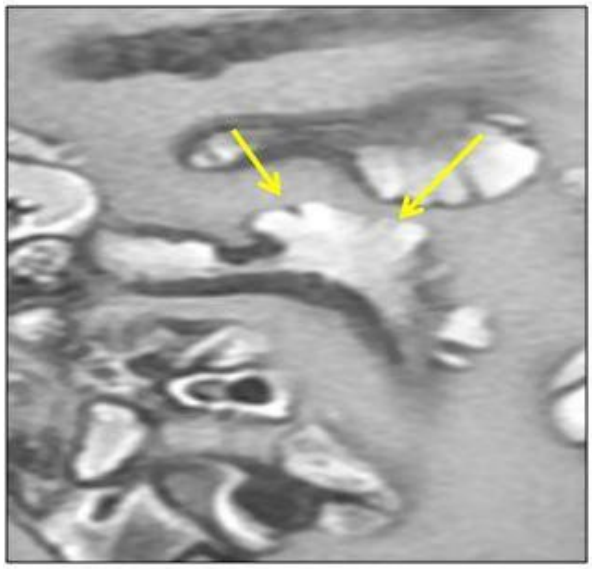
4. **Restricted diffusion -** Select the appropriate button if DWI images are not provided. Otherwise rate DWI signal on the highest B-value images per below. Please check to see if the ADC maps are available. If so review them to ensure that DWI signal does not result from T2 shine-through.

Restricted Diffusion

0 – absent, 1 – present (can be homogeneous, stratified, or luminal)



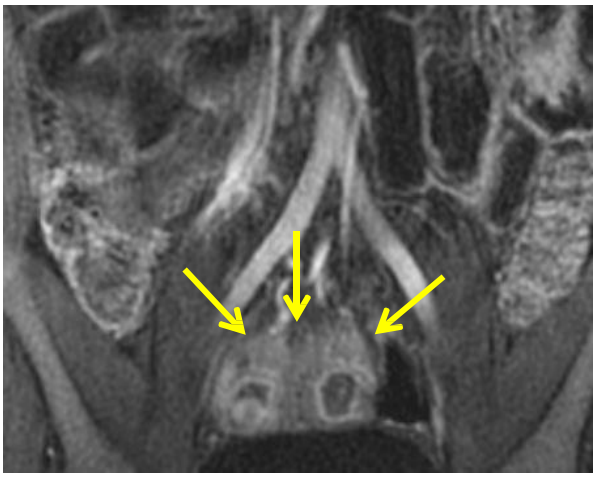
5. Sacculations along antimesenteric border – Rate as present or absent within the length of the stricture.

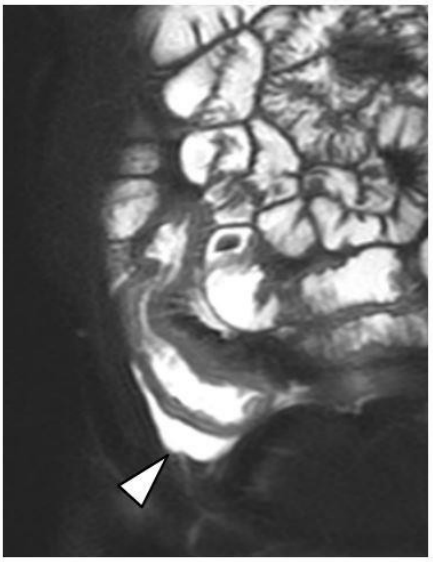


Sacculation

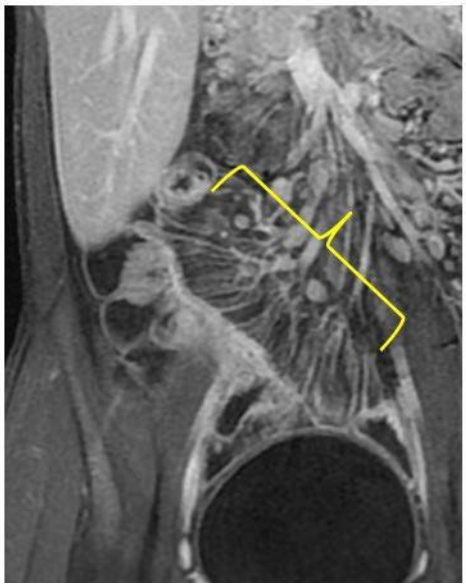
Perienteric/Mesenteric Observations – These observations will be made in the Observations tab on a per stricture basis.

1. **Perienteric edema, fluid or stranding** – Stranding is visible on post-contrast sequences. Evaluate fluid and edema on T2-weighted fat-suppressed images. Rate as present or absent.





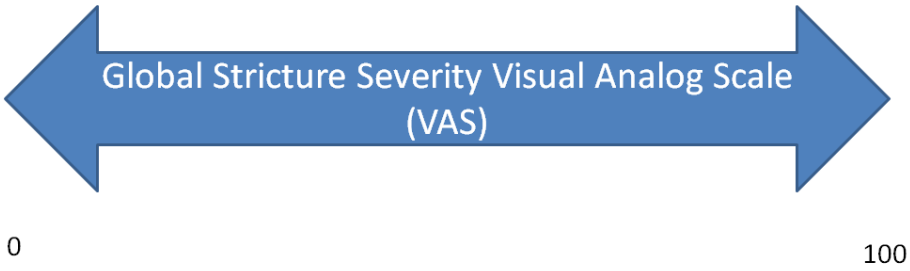
2. **Vasa recta distension –** This should be rated as present when there is an **INCREASED NUMBER AND SIZE** of vasa recta. Rate unequivocal cases only. The normal vasa recta are often not seen. Rate as present or absent.



Patient Level Observations – These are also made in the measurements tab.

Visual Analog Scale (VAS):

Give your overall impression of stricture severity from 1 to 100, with 1 representing the mildest stricture you have ever seen to 100 representing the worst you have ever seen without penetrating complications.



A score of 100 represents the worst stricture you have ever seen without a fistula or abscess, while a score of 1 represents the mildest stricture you have ever seen. Take into account all strictures that extend to within 15 cm of the ileocecal valve or anastomosis. The Global stricture score includes both inflammatory and non-inflammatory components, which are scored separately. Factors such as degree of luminal narrowing and proximal obstruction, mesenteric edema, and sinus tracts should increase the score. Short and long strictures can have high scores.

Findings Increasing Inflammation

- Wall Thickness with T2 hyperintensity on FATSAT (edema)
- Ulcers, ↑ DWI
- Early and intense hyperenhancement

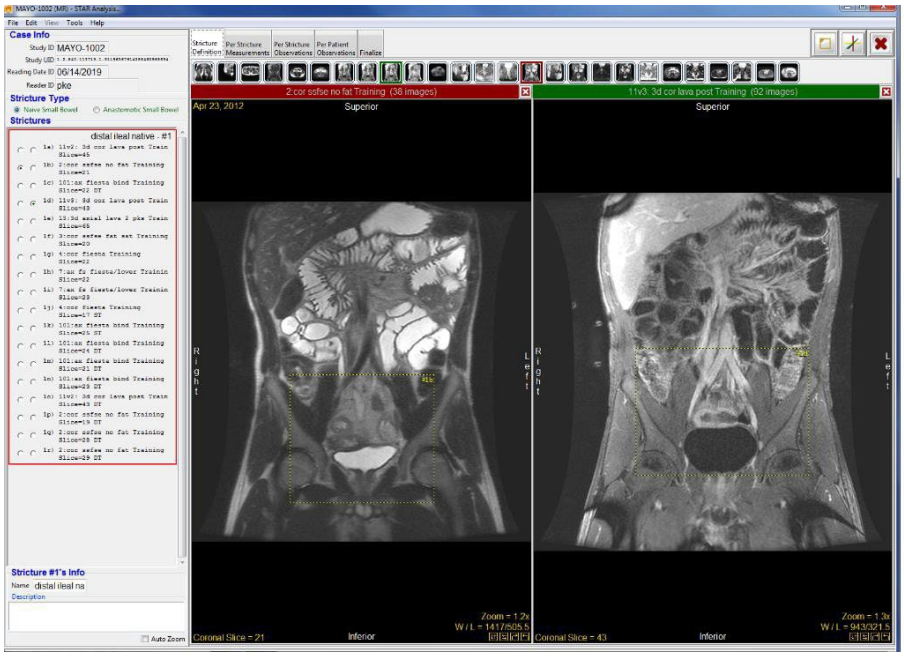
Findings Increasing Non-inflammation Damage

- Wall thickness with T2 hypointensity on FAT SAT (fat or fibrosis)
- Sacculation
- Whiskering
- Delayed enhancement
- Fibrofatty proliferation

Supplemental Appendix 3. Stenosis Therapy and Anti-Fibrosis Research (STAR) inflammatory bowel disease computer workstation and processes.

A computer workstation was built to facilitate image review and annotation by the central readers and ensure that the same stricture was annotated across central readers. A secure database maintained all annotations, measurements, and observations.

Image annotation was initiated by drawing bounding boxes around the stricture on required pulse sequences. A series of tabs then guided the central reader to make quantitative measurements using customised tools and record specific imaging features.



Quantitative features were recorded in a single tab. Image series were interpolated to volumes using information in the DICOM header. All measurements except for stricture length could be made in single images. For stricture length, radiologists used a 3D spline tool and clicked within the bowel lumen from the proximal to the distal end of the stricture. Tools permitted different pulse sequences to be quickly exchanged so that measurements could be compared between image series.

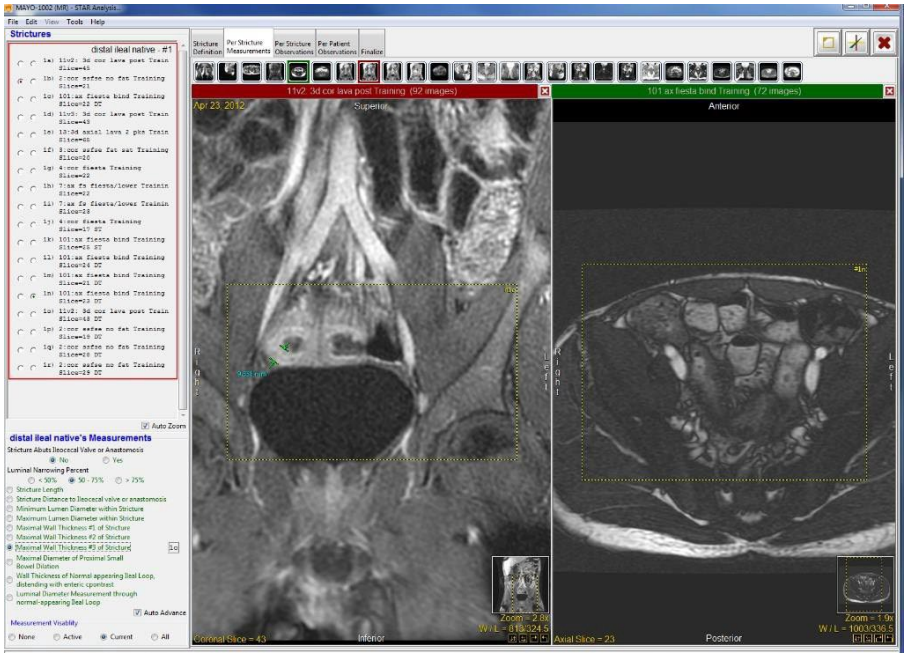
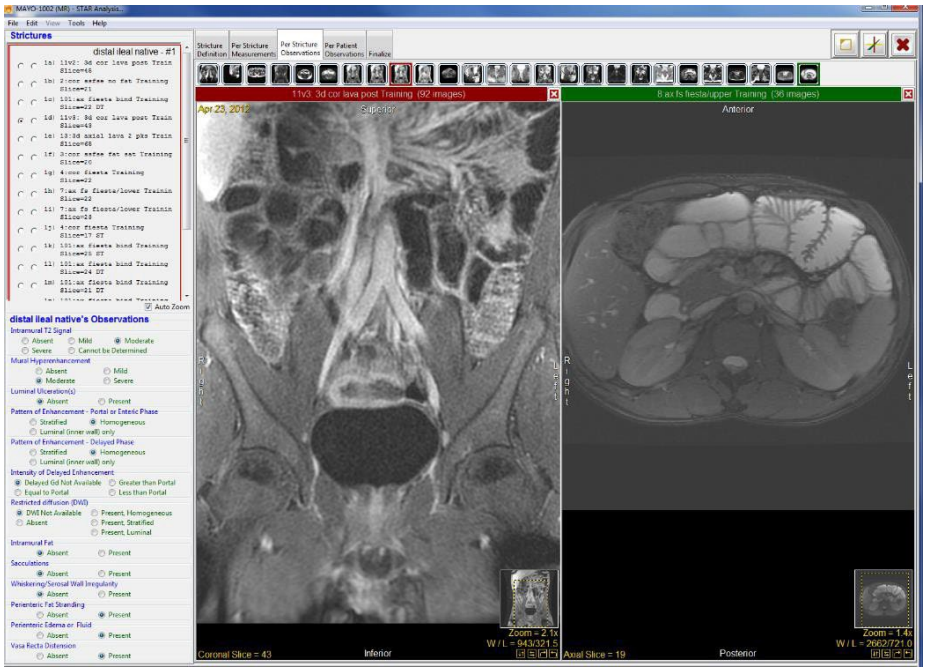
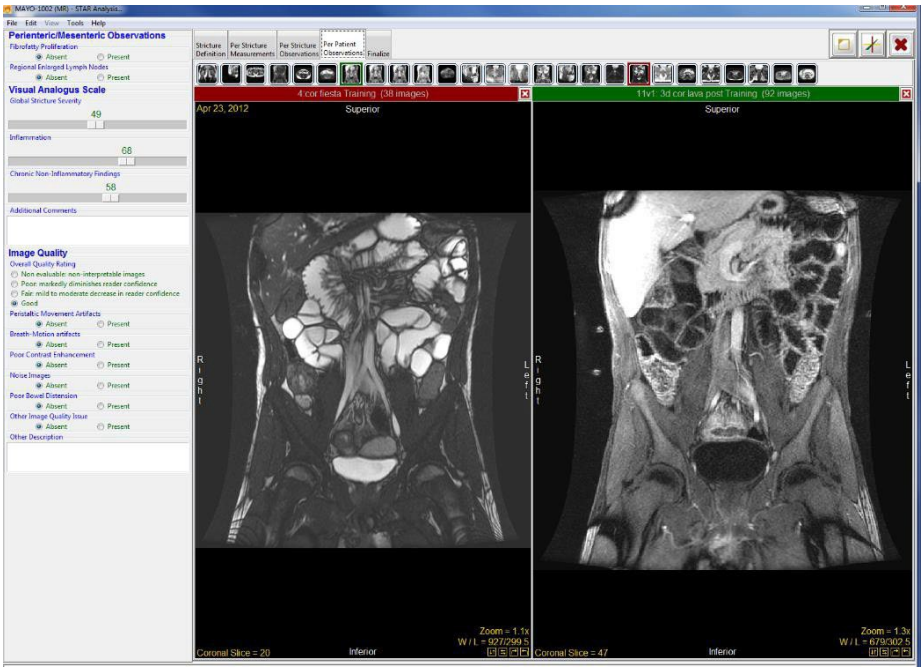


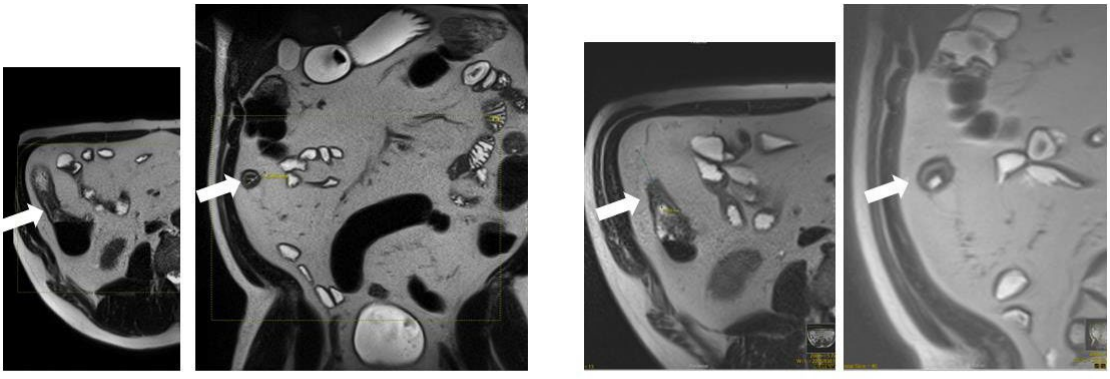
Image features were evaluated by having the radiologists drag the relevant pulse sequences into the 2-panel viewer. A menu to the left of the 2-panel viewer presented the radiologist with features to be graded and rating options.



The visual analog scale (VAS) and image quality metrics were graded on the final tab after all measurements and observations were made. Each case was locked upon submission, meaning that the radiologist was unable to review the images and feature ratings after the case was submitted. The study database was transmitted to Alimentiv, Inc. at regular intervals.



Once all of the baseline exams (read twice by each central reader) and follow-up exams (read once by 1 central reader) were evaluated, an independent assessment was performed by a subspecialized gastrointestinal radiologist (JGF, with 25 years of experience as a staff abdominal radiologist) who had not performed the initial reading. The purpose of this was to compare the annotation of terminal ileal strictures between the baseline and follow-up MRE exams. This was done to ensure that the central readers had measured the same stricture on both exams, as it was possible to measure different terminal ileal strictures between exams. Below are images selected from a central reader's annotations from a baseline and follow-up MRE, demonstrating that the central reader was evaluating the same stricture. The non-central reader radiologist determined if the comparison was valid; he was not permitted to alter any central reader markings.



CENT0013
Baseline Exam
Reader 3

CENT0080
Follow-up Exam
Reader 3

VALID

Supplemental Appendix 4: Additional Statistical Methods

Within-patient standard deviation for reliable features

In exploratory analyses the within-patient SD for the most reliable features was calculated. Random effects models estimated within-patient variance. Continuous measurements were the outcome variables, and the patient was the random effect. The SD was calculated as the square root of the within-patient variance.

Model building

A model of stricture severity was developed using linear regression modelling. Features with moderate inter-rater reliability were the independent variables and the dependent variable was the stricture severity VAS. Features with $P \leq 0.10$ were selected for model construction. The final model was built according to a backward step-down approach with $P \leq 0.05$ (as calculated by Student's t-test) used for feature retention. Final model stability was assessed and calibrated using the bootstrap method with 2000 replicates.

Sample size calculation

Sample size calculation was estimated a priori based on the 1-way random effects model (20). Assuming an ICC of 0.75, evaluation of 50 or 100 exams by 4 central readers ensured the 1-sided 95% CI would exceed 0.70 with 80% probability.