Observer performance for adaptive, image-based denoising and filtered back projection compared to scanner-based iterative reconstruction for lower dose CT enterography

Joel G. Fletcher[1], **Amy K. Hara**[2], **Jeff L. Fidler**[1], **Alvin C. Silva**[2], **John M. Barlow**[1], **Rickey E. Carter**[3], **Adam Bartley**[3], **Maria Shiung**[1], **David R. Holmes III**[4], **Nicolas K. Weber**[5], **David H. Bruining**[5], **Lifeng Yu**[1], and **Cynthia H. McCollough**[1]

[1]Department of Radiology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

[2]Department of Radiology, Mayo Clinic, Scottsdale, AZ, USA

[3]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

[4]Department of Biomedical Imaging Resources, Mayo Clinic, Rochester, MN, USA

[5]Department of Gastroenterology, Mayo Clinic, Rochester, MN, USA

Abstract

Purpose—The purpose of this study was to compare observer performance for detection of intestinal inflammation for low-dose CT enterography (LD-CTE) using scanner-based iterative reconstruction (IR) vs. vendor-independent, adaptive image-based noise reduction (ANLM) or filtered back projection (FBP).

Methods—Sixty-two LD-CTE exams were performed. LD-CTE images were reconstructed using IR, ANLM, and FBP. Three readers, blinded to image type, marked intestinal inflammation directly on patient images using a specialized workstation over three sessions, interpreting one image type/patient/session. Reference standard was created by a gastroenterologist and radiologist, who reviewed all available data including dismissal Gastroenterology records, and who marked all inflamed bowel segments on the same workstation. Reader and reference localizations were then compared. Non-inferiority was tested using Jackknife free-response ROC (JAFROC) figures of merit (FOM) for ANLM and FBP compared to IR. Patient-level analyses for the presence or absence of inflammation were also conducted.

Results—There were 46 inflamed bowel segments in 24/62 patients (CTDI$_{vol}$ interquartile range 6.9–10.1 mGy). JAFROC FOM for ANLM and FBP were 0.84 (95% CI 0.75–0.92) and 0.84 (95% CI 0.75–0.92), and were statistically non-inferior to IR (FOM 0.84; 95% CI 0.76–0.93). Patient-level pooled confidence intervals for sensitivity widely overlapped, as did specificities.

Image quality was rated as better with IR and AMLM compared to FBP ($p < 0.0001$), with no difference in reading times ($p = 0.89$).

Conclusions—Vendor-independent adaptive image-based noise reduction and FBP provided observer performance that was non-inferior to scanner-based IR methods. Adaptive image-based noise reduction maintained or improved upon image quality ratings compared to FBP when performing CTE at lower dose levels.

Keywords

Crohn's disease; Radiation dose; CT; Iterative reconstruction

CT enterography is performed with a large volume of enteric contrast to distend the small bowel, and intravenous contrast to enhance segments with intestinal inflammation or mass lesions [1, 2]. CT enterography has been shown to change patient management in approximately half of patients with known or suspected Crohn's Disease in the outpatient setting [3, 4]. Similarly, CT has been shown to change patient management in symptomatic patients in the emergency room [5, 6]. While MR is the recommended test when assessing for disease response to therapy or when examining younger patients [7], CT enterography may be preferred in many situations because of planned intervention, lack of timely MR availability, or other factors [8].

Due to the repetitive use of CT imaging to examine for complications of Crohn's Disease, many institutions perform CT enterography using lower radiation doses than a routine abdominopelvic CT examination. In general, lower dose CT images have more noise and inferior image quality than higher radiation dose exams if noise reduction techniques are not applied. Despite these limitations, multiple studies have shown that radiation doses can be substantially lowered for CT enterography without sacrifices in observer performance [9–12]. Noise reduction techniques such as iterative reconstruction improve image quality [13, 14], but not necessarily observer performance [15–17]. Scanner-based iterative reconstruction methods such as Adaptive Statistical Iterative Reconstruction (ASIR, GE Healthcare) and Sinogram-affirmed Iterative Reconstruction (SAFIRE, Siemens Healthcare) are performed within the scanner image construction system, and while expensive, minimally disrupt workflow and reconstruct images quickly. However, it may be cost prohibitive or not feasible to equip older CT systems with iterative reconstruction, as replacement of the image construction system may be required [18]. Another option is vendor-independent image-based noise reduction strategies [19–22]. In our experience with several vendor-independent CT noise reduction methods, these image-based methods are relatively inexpensive to install and can be useful in CT practices that employmultiple different scannermakes and models, as a single server can service multiple CT scanners. These typically require some optimization based on dose level, slice thickness, and diagnostic task.

While most institutions now have at least one state-of-the-art CT system with iterative reconstruction, it can be daunting or impossible to triage all Crohn's patients to state-of-the-art scanners with such technology, particularly with other patient groups and diagnostic tasks competing for the same access. So it is highly desirable to develop low-dose CT

enterography methods that are vendor neutral. One such vendor neutral solution has recently been described as adaptive non-local means (ANLM). ANLM adapts denoising strength according to the local noise level [23] and is independent of scanner model, software or hardware.

Our purpose was to compare observer performance for intestinal inflammation between reconstruction methods for low-dose CT enterography. Specifically, we compared reader identification of intestinal inflammation between scanner-based iterative reconstruction methods (ASIR and SAFIRE) and an image-based noise reduction method (ANLM) and filtered back projection.

Methods

After IRB approval and patient consent to the retrospective use of medical records for research purposes, we collected CT enterography exams performed using standard low-dose enterography techniques and reconstructed with commercially available iterative reconstruction methods in patients with known or suspected Crohn's Disease. Other inclusion criteria required: (1) a Gastroenterology dismissal note incorporating CTE findings and all other endoscopic, clinical, and laboratory information, (2) availability of filtered back projection images with slice thicknesses matching IR images, and (3) availability of thin FBP images (for ANLM processing). Patients were excluded if any of the inclusion criteria were not met. Due to the study design, we attempted to find consecutive patients using vendor-based iterative reconstruction, with a target population of 50% of patients being positive for enteric inflammation and 50% being negative. A final sample size of 62 was administratively chosen based on the amount of cases radiologists could evaluate in a single day. A summary of the overall study is given in (Fig. 1).

CT acquisition and image reconstruction

Low-dose CT enterography exams were performed according to standard clinical protocols at two sites within the Mayo Clinic Enterprise. Low-dose CT enterography exams were performed as previously described at Mayo Clinic, Arizona using a GE Discovery 750 CT scanner (GE Healthcare, Milwaukee, WI), and at Mayo Clinic Rochester using a Siemens Definition FLASH (Siemens Healthcare, Malvern, PA) [12]. These exams are performed in patients with a suspicion of small bowel inflammation due to Crohn's disease or other causes. Routine dose CT enterography exams are utilized in patients with obscure GI bleeding or when there is concern for a small bowel or pancreatic neoplasm. For low-dose CTE on the GE system, 100 kV was employed as a fixed KV with a noise index of 31 for patients within this retrospective study. ASIR blending was set to 70%. On the Siemens system, a vendor-supplied kV selection tool was employed to select the most dose-efficient tube potential [24], with an automatic exposure control setting of 160 quality reference mAs and a SAFIRE reconstruction kernel (kernel I-40 with a strength of 2). All images were reconstructed at 3.75 and 3 mm for the GE and Siemens systems, respectively, using an increment of 2 mm. In order to match the vendor-supplied IR images, filtered back projection images using the same slice thickness and reconstruction increment were performed. Thin 0.625 or 1 mm FBP images were sent to a computer server, which

performed the ANLM denoising, and which reconstructed axial slice thicknesses to match the corresponding FBP and IR images for each patient. ANLM is an image-based, vendor-independent noise reduction method which adapts the denoising strength to the local noise level, and which has been analytically estimated and validated with acceptable high and low contrast resolution [23].

Reference standard

The reference standard was created by a gastroenterologist and radiologist not participating in CTE interpretation. Prior to reviewing the CTE images, the physicians reviewed the dismissal gastroenterology note, all prior gastroenterology notes including outside records, and all available endoscopy reports, histopathology reports, surgical reports, relevant serum laboratory values, as well as all other clinical notes, to determine the presence or absence of intestinal inflammation. Based on this retrospective chart review, the gastroenterologist and radiologist then decided if the patient had Crohn's disease. When intestinal inflammation was due to causes other than Crohn's disease, this was also recorded. Locations of intestinal inflammation within the gut on prior imaging, endoscopy or capsule endoscopy, and surgery were noted.

Subsequently, the clinical vendor-based IR CTE images were viewed by these physicians on a specially configured computer workstation. All intestinal inflammation was marked directly on the workstation images using a circumscription tool, with these results archived. All markings had to be congruent with the previously determined reference diagnosis and all known location information. For this anatomic marking and circumscription of intestinal inflammation, which served as our reference standard, a single inflammatory lesion was defined as a segment of bowel involved by continuous inflammation and separated from other inflammatory lesions by normal-appearing bowel. For example, in pancolitis, the entire colon is counted as one inflammatory lesion; if a patient had Crohn's proctosigmoiditis and ileocecal inflammation, two inflammatory lesions were recorded. When reference standard diagnosis for inflammation in the GI tract was positive, but there were no abnormalities visible on the CTE images, the bowel segment corresponding to the reference standard was still marked on CTE images (so failure to identify this inflammation would be considered a false negative). Penetrating complications (fistulae, abscess) were also marked. Apart from the workstation, physicians recorded inflammation by segment for demographic purposes (anorectum, left colon, right colon, the distal 20 cm of neo-terminal ileum, jejunum to terminal ileum, and stomach/duodenum). Inflammatory severity in each of these regions was graded as follows: mild (segmental hyperenhancement with wall thickness of 3–5 mm), moderate (segmental hyperenhancement with wall thickening greater than 5 mm), or severe (segmental hyperenhancement with wall thickening greater than 10 mm or greater than 5 mm with perienteric stranding or penetrating ulcers) [25].

Reader evaluation

Three gastrointestinal radiologists reviewed the LD-CTE cases at three sessions using the same customized computer workstation used to create the reference standard. At each session, each radiologist scrolled up and down through the volume data and interactively marked all inflammatory lesions, circling them with a tool, and using a drop-down list to

state bowel segment and the presence of inflammation or penetrating complications. Cases were randomized for each interpretation session such that each patient exam was shown only once during each session, with the image reconstruction method randomly prescribed by statistical collaborators. At each interpretation session, radiologists were instructed to only evaluate the gastrointestinal tract from the stomach to anus. Confidence scores on a 0–100 scale were used to determine the confidence of the localization. Interpretation sessions were separated by a minimum of 7 days, and reader interpretation time per case was automatically recorded by the computer workstation.

Following visual examination of the GI tract, each reader completed an image quality survey based on European Quality CT criteria [26], and using drop-down menus, with readers rating overall image quality, sharpness, noise, and noise texture ("Appendix").

Matching of reference and reader markings

For matching of reference and reader markings, the computer workstation was re-configured so that reference and reader detections would be simultaneously displayed for each interpretation (Fig. 2). Two experienced gastrointestinal radiologists not participating in reader interpretation matched reference and reader markings using a grid which displayed all the reader and reference detections. Unmatched reference at detections was classified as false negatives. Unmatched reader detections were characterized as either false positive or non-relevant (e.g., findings in solid organs and outside of bowel).

Statistical analysis

The primary purpose of this study was to determine if ANLM or FBP resulted in inferior performance to vendor-supplied IR methods for low-dose CTE. Consequently, the primary analysis was conducted using non-inferiority principles [27] on the estimated difference between observer performance for the vendor-based iterative reconstruction methods verses ANLM and FBP. Observer performance for each reconstruction type was estimated by calculating a figure of merit (FOM) from the area under the jackknife after free-response ROC (JAFROC) curve [28, 29] using the Dorfman-Berbaum-Metz model with fixed-reader, random cases using JAFROC Version 4.2 [30]. Differences between the FOM for vendor-supplied iterative reconstruction and ANLM and FBP images, as well as their corresponding 95% confidence intervals, were calculated across readers. Using a non-inferiority study design, the limit of non-inferiority was set at 0.1 prior to initiation of the study, meaning that if the lower limit of the 95% confidence interval for the difference between reconstruction methods was greater than −0.1, then ANLM or FBP reconstructions were non-inferior. If the lower limit of the 95% confidence interval for the difference between reconstruction methods was less than −0.1, then the ANLM or FBP reconstructions would be interpreted as non-inferiority not shown (i.e., statistically inconclusive). Patient-level analyses for the detection of inflammation were also computed. Pooled estimates and 95% confidence intervals for sensitivity and specificity were estimated using generalized estimating equations using a generalized linear model consisting of a binomial distribution and identity link. These analyses were conducted using the SAS System version 9.3 and the GENMOD procedure (Cary, NC). Finally, image quality ratings were used to create a composite rating

(mean score) within each patient—reconstruction pairing. Differences in these composite scores across reconstruction types were assessed using a random intercept mixed model to account for the clustering of ratings patient case. SAS PROC MIXED was used for these analyses.

Results

Patients and exams

A total of 62 patients who underwent LD-CTE using vendor-supplied iterative reconstruction methods met inclusion but not exclusion criteria for our retrospective study. Of the selected cases, 38 were negative by the reference standard. The remaining twenty-four patients (39%; 24/62) had 46 inflammatory lesions within the gastrointestinal tract (mean of 1.9 lesions/patient; range 1–6). The majority of the lesions were located in the small bowel ($n = 30$, 65%), with the remainder located in the colon ($n = 9$, 20%), rectum ($n = 6$, 13%), or stomach/duodenum ($n = 1$, 2%). Inflammatory severity of these segments was graded as mild (16/46; 35%), moderate (24%), severe (37%) by the GI radiologist and gastroenterologist creating the reference standard. There were two patients with colonic inflammation by the reference standard without identifiable abnormalities on LD-CTE images (two ulcerative colitis patients—one with mild inflammation in the right colon; and the other with pouchitis). There were five patients with penetrating complications.

All LD-CTE exams were acquired using automatic exposure control. Thirty-two of the LD-CTE exams used Adaptive Statistical Iterative Reconstruction (ASIR, GE Healthcare) as the vendor-supplied IR method while 30 used Sinogram-affirmed Iterative Reconstruction (SAFIRE, Siemens Healthcare). The median $CTDI_{vol}$ was 9.3 ± 4.1 mGy (range 4.9–29.4 mGy; 25th–75th percentile 6.9–10.1). The median BMI was 24.5 (range 17.0–42.9).

Observer performance for detecting intestinal inflammation

Table 1 shows observer performance for vendor-supplied IR as well as ANLM and FBP, in addition to estimated differences in performance between reconstruction methods, along with 95% confidence intervals. The estimated differences in the JAFROC FOM between scanner-based IR vs. ANLM or FBP were small (–0.002 and –0.005, respectively), and the lower limits of the 95% confidence intervals were above the –0.1 threshold for non-inferiority. Consequently, the null hypotheses for ANLM and FBP were rejected in favor of the alternative hypothesis that the reconstruction techniques were statistically non-inferior to the vendor-supplied IR. Table 1 also presents the reader-specific performance for identifying inflammatory lesions within the GI tract on a perlesion basis as represented by the JAFROC FOM. As expected from the overall estimates for the FOM, the reader-specific estimates were very similar among readers and reconstruction methods, but there was some suggestion that reader performance may vary according to reconstruction technique. Given the limited number of readers, this variation in performance was not estimable.

Table 2 shows reader-specific and pooled (i.e., common) estimates for a patient-level sensitivity and specificity on a per-reader basis, with wide overlap of sensitivity and specificity confidence intervals for each reconstruction method (Figs. 3, 4). Penetrating

complications occurred in five patients, including enteroenteric fistulae (two patients), perianal fistulae (two patients), enterocutaneous fistula (one patient), and mesenteric abscess (one patient). For penetrating complications, the pooled sensitivities across readers was 80% (95% CI 57–100%) for all techniques, with similar specificities: 98% (95% CI 95–100%) for FBP, 95% (95% CI 92–98%) for ANLM, and 95% (95% CI 92–98%) for IR (Fig. 5).

Image quality comparison

Table 3 shows image quality ratings and reader times for each reconstruction method. Reader rankings for overall image quality for IR and ANLM images were significantly higher than for FBP ($p < 0.0001$ for both comparisons). Sharpness was slightly improved (lower score better) for ANLM (mean rating 1.80) compared to IR and FBP (1.99 and 2.0, respectively; $p < 0.0004$). Noise was significantly worse (higher numeric score) for FBP (mean rank 2.84) compared to IR and ANLM (mean ranks 2.54 and 2.44, respectively, $p < 0.0001$). Noise texture was not significantly different between reconstruction methods. There was no difference in interpretation times between reconstruction methods for any reader ($p = 0.89$).

Discussion

We have demonstrated that for low-dose CTE, an image-based noise reduction method (ANLM) and filtered back projection (FBP) result in observer performance that is non-inferior to the observer performance obtained with scanner-based iterative reconstruction methods at the dose levels employed (interquartile range $CTDI_{vol}$ 6.9–10.1 mGy). Similar performance was observed despite the substantial improvements in image quality using both noise reduction methods. Furthermore, there was no difference in reader interpretation times between reconstruction methods.

These results have important implications for radiology practices with Crohn's patients and different CT scanner technologies. First, lower dose CTE protocols for Crohn's patients can be performed across a wide variety of CT platforms with differing technology. This realization may be especially important in the Emergency Room, where symptomatic presentation might preclude triage to a state-of-the-art CT system with iterative reconstruction. Multiple studies have highlighted the role of CT in altering patient management in this population [5, 6]. Second, observer performance is not linked to image quality: high contrast diagnostic tasks, such as CTE permit greater levels of dose reduction [18], and we observed no difference in performance despite observed differences in image quality. Consequently, the benefit of noise reduction may be personal preference and acceptance of lower dose images that are less noisy. For these purposes, an image-based noise reduction system is an acceptable method that can be employed in clinical practice for older CT systems. Image-based noise reduction is vendor independent, and such algorithms can be placed on computer servers, which can denoise thin images from multiple vendors simultaneously. The downside of image-based denoising is slight loss of workflow efficiency as thin FBP images are transferred to a server, denoised with new images reconstructed to the desired slice thickness and then transferred to PACS. This transfer and processing time takes approximately 5–10 min. In comparison, the commercial vendor-

based IR systems we used for study purposes are reconstructed immediately after patient scanning and sent directly to PACS. Nevertheless, image-based denoising provides an attractive alternative for noise reduction in large practices with heterogeneous scanner fleets or rural practices, where noise reduction will not be used on all cases. Subsequent to this study, we have successfully used the image-based ANLM noise reduction method for two older CT systems in our outpatient practice for CT enterography and CT urography.

Several studies have evaluated the performance of lower dose CTE in combination with a noise reduction system. Kaza et al. compared 120 kV CT enterography in patients weighing less than 160 pounds to 80 kV lower dose CT enterography with images reconstructed using ASIR [11]. Lower dose CTE images with FBP alone were not evaluated. They found lower dose CT enterography had slightly lower image quality scores, but no change in accuracy for inflammatory bowel disease in a subset of their population. In comparison, Lee et al. compared lower dose CTE images with iterative reconstruction in image space vs. with FBP alone at the level of the terminal ileum, and found that FBP images alone did not result in inferior performance [15]. Their results are similar to our study, but the iterative reconstruction they tested was not vendor independent and inflammation in the neo-terminal ileum alone was examined. In the current work, we evaluated for the necessity of noise reduction at all (with comparison to FBP images), and compared different denoising approaches for identification of inflammation throughout the length of the GI tract. It is difficult to compare the dose levels used in these prior studies to ours as the Michigan study evaluated only small patients, and we employed variable kV selection on a subset of our patients. However, Lee et al. used an automatic exposure control setting of 100 Quality Reference mAs for each of two X-ray tubes in their study, potentially similar to the 160 Quality Reference mAs we used in a single X-ray tube.

Our study has several limitations. We evaluated only 62 patients. Ideally we would have chosen 82 patients based on paired comparisons, a 90% expected proportion of agreement and alpha of 0.05, and null hypothesis agreement rate of 80%. Given practical limitations of reader time, our methodology was altered to include a more generalized analysis strategy that accounted for the multiple readers and permitted detection of intestinal inflammation anywhere along the GI tract. Additionally, our study cohort included several patients with non-IBD-related intestinal inflammation (e.g., radiation enteritis and sprue). However, we believe this heterogeneity reflects clinical practice where etiology and the presence of inflammation are often not known beforehand. Additionally, we did not require histopathology for identification of intestinal inflammation in the reference standard. Such a requirement would not have been possible given our aim to identify multiple locations of intestinal inflammation throughout the GI tract, as many areas would not be endoscopically accessible [31]. Rather, our study design was based on a non-inferiority methodology that examines the fidelity in reader performance between the state-of-the-art images (vendor-supplied IR) vs. cheaper and potentially more widely available alternatives (FBP and ANLM). Our minimum requirement for at least 7 days between reading sessions may have not permitted complete memory extinction for the readers. Finally, we only evaluated for non-inferiority at one dose level using similar protocols on two CT systems from two different vendors. Studies using lower radiation dose techniques may have different results.

In conclusion, we found that identification of intestinal inflammation using a vendor-independent, adaptive image-based noise reduction method (ANLM) and FBP were non-inferior compared to scanner-based iterative reconstruction for lower dose CT enterography at the dose levels employed. This finding should permit radiology departments and emergency rooms great flexibility in providing lower dose CT options for patients with Crohn's disease. Image-based denoising methods provide a cost-effective alternative to CT noise reduction that may be particularly useful in heterogenous CT fleets or for older CT systems, where iterative reconstruction is not available.

Acknowledgments

References

1. American College of Radiology. [Accessed 25 Mar 2013, 2014] ACR Appropriateness Criteria. (http://www.acrorg/quality-safety/appropriateness-criteria)

2. Al-Hawary MM, Kaza RK, Platt JF. CT enterography: concepts and advances in Crohn's disease imaging. Radiol Clin North Am. 2013; 51:1–16. [PubMed: 23182504]

3. 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. [PubMed: 21337477]

4. Higgins PD, Caoili E, Zimmermann M, et al. Computed tomographic enterography adds information to clinical management in small bowel Crohn's disease. Inflamm Bowel Dis. 2007; 13:262–268. [PubMed: 17206719]

5. Kerner C, Carey K, Mills AM, et al. Use of abdominopelvic computed tomography in emergency departments and rates of urgent diagnoses in Crohn's disease. Clin Gastroenterol Hepatol. 2012; 10:52–57. [PubMed: 21946122]

6. Israeli E, Ying S, Henderson B, et al. The impact of abdominal computed tomography in a tertiary referral centre emergency department on the management of patients with inflammatory bowel disease. Aliment Pharmacol Ther. 2013; 38:513–521. [PubMed: 23837741]

7. Guimaraes LS, Fidler JL, Fletcher JG, et al. Assessment of appropriateness of indications for CT enterography in younger patients. Inflamm Bowel Dis. 2010; 16:226–232. [PubMed: 19637359]

8. Fletcher JG. CT enterography technique: theme and variations. Abdom Imaging. 2008; 34:283–288. [PubMed: 18551337]

9. Allen BC, Baker ME, Einstein DM, et al. Effect of altering automatic exposure control settings and quality reference mAs on radiation dose, image quality, and diagnostic efficacy in MDCT enterography of active inflammatory Crohn's disease. Am J Roentgenol. 2010; 195:89–100. [PubMed: 20566801]

10. Kambadakone AR, Prakash P, Hahn PF, Sahani DV. Lowdose CT examinations in Crohn's disease: impact on image quality, diagnostic performance, and radiation dose. Am J Roentgenol. 2010; 195:78–88. [PubMed: 20566800]

11. Kaza RK, Platt JF, Al-Hawary MM, et al. CT enterography at 80 kVp with adaptive statistical iterative reconstruction versus at 120 kVp with standard reconstruction: image quality, diagnostic adequacy, and dose reduction. Am J Roentgenol. 2012; 198:1084–1092. [PubMed: 22528897]

12. Del Gaizo AJ, Fletcher JG, Yu L, et al. Reducing radiation dose in CT enterography. Radiographics. 2013; 33:1109–1124. [PubMed: 23842974]

13. Sagara Y, Hara AK, Pavlicek W, et al. Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. Am J Roentgenol. 2010; 195:713–719. [PubMed: 20729451]

14. Prakash P, Kalra MK, Kambadakone AK, et al. Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique. Invest Radiol. 2010; 45:202–210. [PubMed: 20177389]

15. Lee SJ, Park SH, Kim AY, et al. A prospective comparison of standard-dose CT enterography and 50% reduced-dose CT enterography with and without noise reduction for evaluating Crohn disease. Am J Roentgenol. 2011; 197:50–57. [PubMed: 21701010]

16. Baker ME, Dong F, Primak A, et al. Contrast-to-noise ratio and low-contrast object resolution on full- and low-dose MDCT: SAFIRE versus filtered back projection in a low-contrast object phantom and in the liver. Am J Roentgenol. 2012; 199:8–18. [PubMed: 22733888]

17. Goenka AH, Herts BR, Obuchowski NA, et al. Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: an 18-reader study. Radiology. 2014; 272:154–163. [PubMed: 24620913]

18. Ehman EC, Yu L, Manduca A, et al. Methods for clinical evaluation of noise reduction techniques in abdominopelvic CT. Radiographics. 2014; 34:849–862. [PubMed: 25019428]

19. Borgen L, Kalra MK, Laerum F, et al. Application of adaptive non-linear 2D and 3D postprocessing filters for reduced dose abdominal CT. Acta Radiol. 2012; 53:335–342. [PubMed: 22362136]

20. De Geer J, Sandborg M, Smedby O, Persson A. The efficacy of 2D, non-linear noise reduction filtering in cardiac imaging: a pilot study. Acta Radiol. 2011; 52:716–722. [PubMed: 21852437]

21. Lubner MG, Pickhardt PJ, Kim DH, et al. Prospective evaluation of prior image constrained compressed sensing (PICCS) algorithm in abdominal CT: a comparison of reduced dose with standard dose imaging. Abdom Imaging. 2015; 40:207–221. [PubMed: 24943136]

22. Nishimaru E, Ichikawa K, Okita I, et al. Development of a noise reduction filter algorithm for pediatric body images in multidetector CT. J Digit Imaging. 2010; 23:806–818. [PubMed: 19536600]

23. Li Z, Yu L, Trzasko JD, et al. Adaptive nonlocal means filtering based on local noise level for CT denoising. Med Phys. 2014; 41:011908. [PubMed: 24387516]

24. Yu L, Fletcher JG, Grant KL, et al. Automatic selection of tube potential for radiation dose reduction in vascular and contrast- enhanced abdominopelvic CT. Am J Roentgenol. 2013; 201:W297–W306. [PubMed: 23883244]

25. Faubion WA Jr, Fletcher JG, O'Byrne S, et al. EMerging BiomARKers in Inflammatory Bowel Disease (EMBARK) study identifies fecal calprotectin, serum MMP9, and serum IL-22 as a novel combination of biomarkers for Crohn's disease activity: role of cross-sectional imaging. Am J Gastroentero. 2013; 108:1891–1900.

26. European Commission. European guidelines on quality criteria for computed tomography (EUR 16262 EN). Luxembourg: European Commission & The Office For Official Publications of the European Communities; 2000.

27. Murray G. Points to consider on switching between superiority and non-inferiority. Clin Pharmacol. 2001; 52:223–228.

28. Chakraborty DP. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med Phys. 1989; 16:561–568. [PubMed: 2770630]

29. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys. 2004; 31:2313–2330. [PubMed: 15377098]

30. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol. 1992; 27:723–731. [PubMed: 1399456]

31. Samuel S, Bruining DH, Loftus EV, et al. Endoscopic skipping of the distal terminal ileum in Crohn's disease can lead to negative results from ileocolonoscopy. Clin Gastroenterol Hepatol. 2012; 10:1253–1259. [PubMed: 22503995]

Appendix: image quality assessment performed by radiologist readers

Image quality criteria were performed after observer interrogation of each dataset. Readers could scroll through the data while evaluating image quality. Criteria and rankings are described below:

Overall image quality (5-point scale)

1 = non-diagnostic due to excessive noise artifacts

2 = diagnosis questionable due to excessive noise/artifacts; moderate decrease in diagnostic confidence

3 = diagnostic with moderate but acceptable noise/artifacts

4 = mild noise, no change in diagnostic confidence

5 = routine diagnostic image quality

Image sharpness (5-point scale)

1 = very sharp

2 = mildly unsharp edges, no diagnostic difference

3 = moderately unsharp, questionable diagnostic difference

4 = noticeable blur with poorly defined edges

5 = non-diagnostic

Image noise (4-point scale)

1 = less than usual

2 = optimal 'routine' noise

3 = increased noise, does not affect interpretation

4 = increased noise affecting interpretation

Noise texture (4-point scale)

0 = no noticeable change

1 = no noticeable change after window settings changed

2 = perceptible change

3 = blotchiness or change affecting confidence
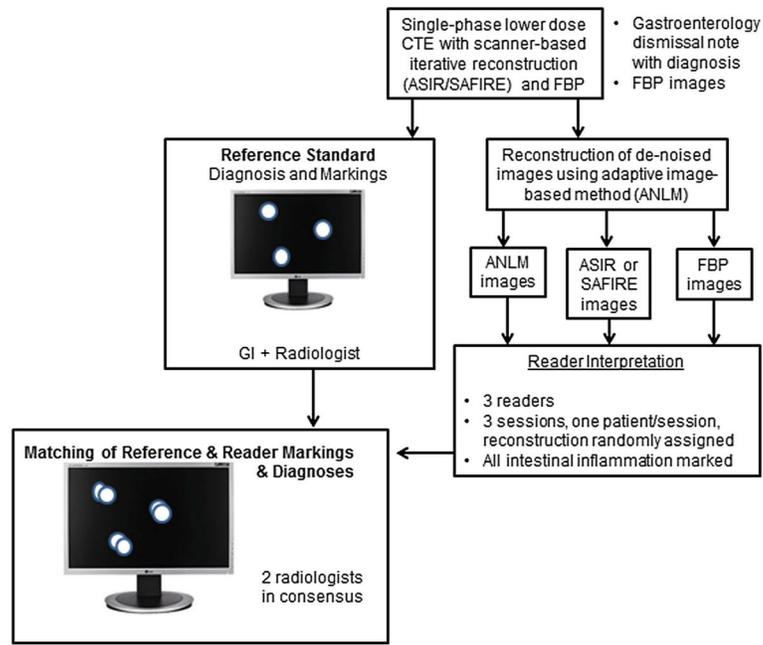
Fig. 1.
Overall study schema.
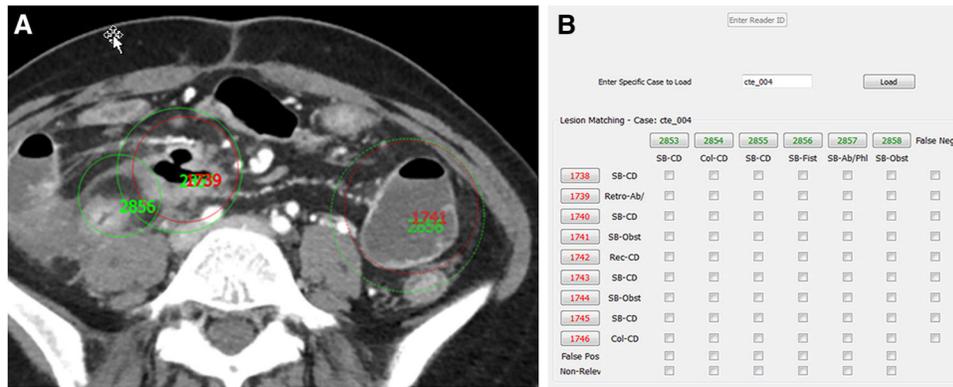
Fig. 2.
Image showing matching of reader (*green circles* and numbers) and reference (*red circles* and numbers) marking viewed by two GI radiologists not participating in CTE interpretation. Radiologists could scroll up and down through the dataset to determine if inflamed segments could were marked by reference physicians and radiologists readers (**A**). Each reference and reader marking was given a unique reference identification number. A table (**B**) then permitted matching of reader and reference markings. Unmatched reference identification numbers were labeled as false negatives, and unmatched reader markings were labeled as false positive or non-relevant (e.g., if marking a renal mass rather than a bowel finding).
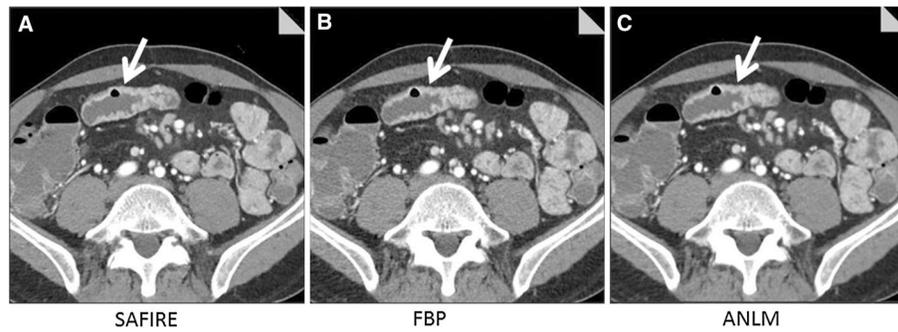
Fig. 3.
50-year-old male with Crohn's disease, night sweats, and weight loss underwent lower dose CT enterography ($CTDI_{vol}$ = 6.6 mGy). Images show inflamed distal jejunal segment with mural enhancement and wall thickening (*arrow*) reconstructed with SAFIRE (**A**), filtered back projection (**B**), and adaptive image-based noise reduction (**C**, ANLM). CT findings in conjunction with elevated sedimentation rate and C-reactive protein led treating gastroenterologist to diagnose recurrent jejunal Crohn's disease and add Purinethol in addition to Cimzia. All readers identified this inflamed bowel segment using all reconstruction types (except for one reader evaluating SAFIRE images).
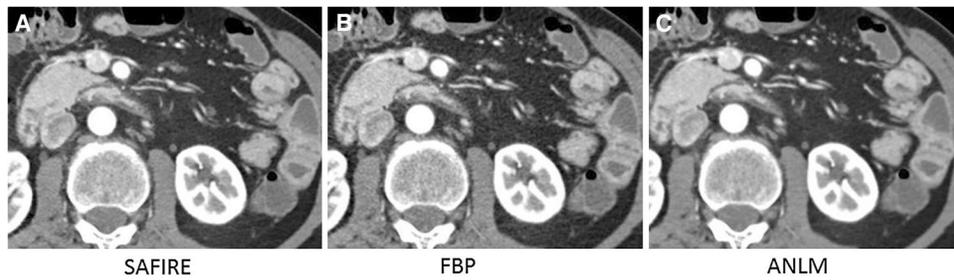
Fig. 4.
61-year-old male with urgency and loose stools underwent lower dose CT enterography (9.4 mGy) demonstrating symmetrical wall thickening, intramural fat, and mural hyperenhancement in the duodenum and proximal jejunum suggesting ulcerative duodenitis and jejunitis in the setting of celiac sprue. Transverse image reconstructed with SAFIRE (**A**), filtered back projection (**B**), and adaptive image-based noise reduction (**C**, ANLM) are shown. Reference standard is positive for duodenal and jejunal inflammation based on villous atrophy at endoscopy, and normalization of stools and weight gain on gluten-free diet. All three readers detected this jejunal inflammation.
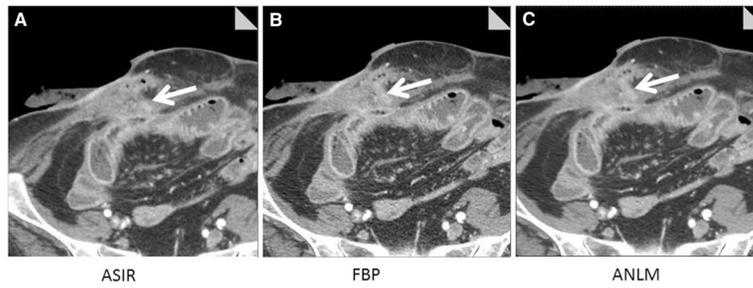
Fig. 5.
64-year-old female underwent lower dose CT enterography showing inflamed terminal ileum and enterocutaneous fistula (arrow), which was subsequently taken down at surgery. Transverse image reconstructed with ASIR (**A**), filtered back projection (**B**), and adaptive image-based noise reduction (**C**, ANLM) are shown. All three readers located the inflamed ileal loop and enterocutaneous fistula.

Table 1

Observer performance for intestinal inflammation using a figure of merit (FOM) and 95% confidence interval for a jackknife after free-response ROC (JACROC) analysis

Configuration	Reader	FOM	95% CI
Commercial IR (ASIR/SAFIRE)			
	1	0.881	–
	2	0.825	–
	3	0.814	–
	Pooled[a]	0.840	(0.755, 0.925)
ANLM			
	1	0.817	–
	2	0.891	–
	3	0.806	–
	Pooled[a]	0.838	(0.754, 0.922)
	Estimated difference[b]	−0.002	(−0.046, 0.042)
FBP			
	1	0.869	–
	2	0.845	–
	3	0.793	–
	Pooled[a]	0.836	(0.751, 0.920)
	Estimated difference[b]	−0.005	(−0.049, 0.039)

Estimated differences between ALM and FBP vs. vendor-supplied IR are also shown. Assessment for non-inferiority was performed using −0.1 as the limit of non-inferiority

[a] Pooled estimate is from a fixed reader, random case analysis

[b] The estimated difference is relative to the commercial IR. Lower limits of the CI that are above −0.1 suggest the altered reconstruction technique is non-inferior to the commercial IR

Table 2

Patient-level estimates of sensitivity and specificity for the detection of bowel inflammation ($n = 24$ with inflammation, $n = 38$ without)

| Configuration | Reader | Sensitivity[a] | | Specificity[a] | |
		Estimate (%)	95% CI	Estimate (%)	95% CI
Commercial IR (ASIR/SAFIRE)	1	88	(69, 96)	87	(73, 94)
	2	79	(60, 91)	84	(70, 93)
	3	79	(60, 91)	89	(76, 96)
	Pooled[b]	82	(69, 95)	87	(80, 94)
ANLM	1	83	(64, 93)	76	(61, 87)
	2	88	(69, 96)	87	(73, 94)
	3	75	(55, 88)	89	(76, 96)
	Pooled[b]	82	(70, 93)	84	(77, 91)
FBP	1	92	(74, 98)	87	(73, 94)
	2	88	(69, 96)	84	(70, 93)
	3	75	(55, 88)	87	(73, 94)
	Pooled[b]	85	(73, 97)	86	(78, 94)

[a]Sensitivity and specificity are based on 24 patients with at least one inflamed segment and 38 patients without, respectively

[b]Pooled estimates of sensitivity and specificity have been adjusted for the correlation of readers within case through generalized estimating equations

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Timing and image quality summary statistics

Variable	Configuration	Mean	SD	p value[a]	post hoc comparisons[b]
Reading time, s					
	Commercial IR	286.7	135.1	0.89	N/A
	ANLM	292.4	116.6		
	FBP	282.3	141		
Quality rating (1 = non-diagnostic, … 5 = routine diagnostic)					
	Commercial IR	4.2	0.4	<0.0001	All pairs
	ANLM	4.3	0.4		
	FBP	3.7	0.5		
Sharpness (1 = very sharp, … 5 = non-diagnostic)					
	Commercial IR	2.0	0.3	<0.0004	b < a, c
	ANLM	1.8	0.3		
	FBP	2.0	0.3		
Noise (1 = less than usual, … 4 = increased noise affecting interpretation)					
	Commercial IR	2.5	0.3	<0.0001	All pairs
	ANLM	2.4	0.3		
	FBP	2.8	0.3		
Texture (0 = no noticeable change, … 3 = blotchiness or change affecting confidence)					
	Commercial IR	1.6	0.4	0.06	N/A
	ANLM	1.5	0.4		
	FBP	1.4	0.4		

Complete scales to assess image quality are available in "Appendix"

[a] p value is for a 2° of freedom test for any difference in mean values among configuration after adjustment for clustering (blocking) on case

[b] Post hoc comparisons are presented without adjustment for multiple testing and are indicated as group "a" for Commercial IR, "b" for ANLM, "c" for FBP