




Structured reporting of CT enterography for inflammatory bowel disease: effect on key feature reporting, accuracy across training levels, and subjective assessment of disease by referring physicians

Benjamin Wildman-Tobriner ¹, Brian C. Allen,¹ Mustafa R. Bashir,^{1,2} Morgan Camp,¹ Chad Miller,¹ Lauren E. Fiorillo,³ Alan Cubre,¹ Sanaz Javadi,⁴ Alex D. Bibbey,¹ Wendy L. Ehieli,¹ Nancy McGreal,⁵ Reinaldo Quevedo,⁵ Julie K. Thacker,⁶ Maciej Mazurowski,¹ Tracy A. Jaffe¹

¹Department of Radiology, Duke University Hospital, 2301 Erwin Rd, Durham, NC 27710, USA

²Center for Advanced Magnetic Resonance Development, Duke University Medical Center, 2301 Erwin Rd, Durham, NC 27710, USA

³Department of Radiology, Ohio State University Wexner Medical Center, 395 W. 12th Avenue, Columbus, OH 43210, USA

⁴Department of Radiology, University of Texas MD Anderson, 1400 Pressler Street, Unit 1459, Houston, TX 77030, USA

⁵Department of Gastroenterology, Duke University Hospital, 2301 Erwin Rd, Durham, NC 27710, USA

⁶Department of Surgery, Duke University Hospital, 2301 Erwin Rd, Durham, NC 27710, USA

Abstract

Purpose: To compare the content and accuracy of structured reporting (SR) versus non-structured reporting (NSR) for computed tomographic enterography (CTE) of inflammatory bowel disease (IBD).

Materials and methods: This IRB-approved, HIPAA-compliant, retrospective study included 30 adult subjects (15 male, 15 female; mean age 41.9 years) with IBD imaged with CTE. Nine radiologists (3 faculty, 3 abdominal imaging fellows, and 3 senior radiology residents) independently interpreted all examinations using both NSR and SR, separated by four weeks. Reports were assessed for documentation of 15 key reporting features and a subset of 5 features was assessed for accuracy. Thirty faculty reports (15 NSR [5 per reader] and 15 SR [5 per reader]) were randomly selected for review by three referring physicians, who independently rated quality metrics for each report.

Results: NSR documented the presence or absence of 8.2 ± 2.2 key features, while SR documented 14.6 ± 0.5

features ($p < 0.001$). SR resulted in increased documentation of 13 of 15 features including stricture ($p < 0.001$), fistula ($p < 0.001$), fluid collection ($p = 0.003$), and perianal disease ($p < 0.001$). Among a subset of five features, accuracy for diagnosing multifocal disease was minimally increased when using SR (76% NSR vs. 83% SR; $p = 0.01$), but accuracy for other features was not affected by report type. Referring physicians significantly preferred SR based on ease of information extraction ($p < 0.01$).

Conclusion: Structured reporting of CTE for IBD improved documentation of key reporting features for trainees and faculty, though there was minimal impact on accuracy. Referring physicians subjectively preferred the structured reports.

Key words: Structured report—Standardized report—Inflammatory bowel disease—Enterography—CT

Correspondence to: Benjamin Wildman-Tobriner; email: benjamin.wildman-tobriner@duke.edu

Radiology reports are an important part of assessing disease severity and response to therapy, and multiple radiology societies are focused on improving reporting

practices [1–6]. One proposed method involves standardizing report templates, also called structured reporting (SR). Potential benefits of SR include more consistent reporting of important scan features [3], better utilization of scan interpretations for research purposes [7, 8], and catering to referring clinician preference [9, 10]. Conversely, concerns regarding diagnostic accuracy, completeness, and time required to complete templates have been raised [11–13]; consequently, SR remains a divisive issue in many radiology practices.

SR has been found to be useful when documenting complex disease processes, such as local staging of pancreatic cancer [3]. Inflammatory bowel disease (IBD), another complex disease process, may also benefit from a standardized report template [14]. In particular, SR may be helpful for computed tomographic enterography (CTE), the CT gold standard imaging examination for diagnosis of and monitoring disease activity in IBD [15]. Cross-sectional imaging assists clinicians' assessment of response to therapy and aids in planning surgical intervention [16, 17], and SR may promote more consistent data exchange and ultimately facilitate communication between radiologists, gastroenterologists, and surgeons. In addition, SR has the potential to assist radiologists in remaining consistent across examinations, particularly when different radiologists are interpreting scans from the same patient over time.

Increased disease complexity has the potential not only to hinder communication across provider types and between radiologists, but it also proves challenging for radiology trainees. Although infrequent, errors made by radiology residents tend to occur when interpreting more complex examinations [18, 19]. It has been proposed that SR may help residents develop a search pattern and improve accuracy; however, results have been mixed [11, 20, 21].

As IBD is a complex process involving multiple bowel segments as well as findings outside the bowel, we hypothesized that using SR would result in a more complete assessment of the abdomen and pelvis and that SR might help trainees improve the consistency and accuracy of their reporting. Thus, the purpose of this investigation was to objectively compare report content and accuracy across multiple training levels of structured reports (SR) versus non-structured reports (NSR) for CTE examinations of patients with IBD.

Materials and methods

Our institutional review board approved this Health Insurance Portability and Accountability Act-compliant, retrospective study, and the requirement for informed consent was waived.

Patient population

Our institution is a tertiary referral center, and the picture archiving and communication system was searched

for CTE examinations performed between July 1, 2013 and July 30, 2015. A random selection of patients (18 years of age or older) with findings of active inflammatory bowel disease included in the impression on the finalized report was included. Patients with negative CTE examinations were excluded. Demographic data were reviewed in the electronic medical record, and the final patient population included 30 subjects (15 male and 15 female), aged 18–77 years (mean = 41.9 ± 14.7 years). All 30 subjects had a history of inflammatory bowel disease; 25 had Crohn's disease, 3 had ulcerative colitis, and 2 had unspecified disease (pathology suggesting IBD but without definite categorization).

Imaging protocol

All CT examinations were performed on 64-detector multi-detector CT scanners (Discovery HD750, GE Healthcare, Milwaukee, WI, USA or Somatom Definition Flash, Siemens Healthcare, Erlangen, Germany). Scan parameters were as follows: tube voltage 120 kVp; automated tube current modulation with a noise index of 22 (GE systems) or a reference mAs of 200 mAs (Siemens systems); and 500-ms gantry rotation with 0.625-mm collimation. Images were reconstructed as 5-mm contiguous axial images and 3-mm contiguous coronal reformatted images were created. Prior to scanning, patients ingested 450–900 mL of low-Hounsfield unit enteric contrast (VoLumen [0.1% barium suspension], Bracco Diagnostics, Princeton, NJ, USA) over a 60-min period, based on patient tolerance. Intravenous contrast enhancement was obtained with 150 mL iopamidol (300 mg Iodine/mL; Isovue-300, Bracco Diagnostics) injected at 3 mL/sec. Images were obtained during the enteric phase with a 45 s scan delay.

Report template

The structured reporting template for the CTE examination was derived from a template proposed by the Society of Abdominal Radiology Crohn's disease focus group and by Baker et al. with minor modifications [16]. The final version contained 14 key features of IBD assessment (Fig. 1). No drop-down lists or default responses were included; blank spaces were present for each item. Of note, the presence of steatorrhea (low-density fecal material) was included, as its symptoms can mimic those of active IBD [22, 23].

Imaging analysis

Nine radiologists independently reviewed the datasets: three fellowship-trained abdominal radiologists (faculty) with 6–11 years' post-fellowship experience, three abdominal imaging radiology fellows, and three senior

GASTROINTESTINAL:
 Evidence of prior surgery: [...]
 Sites of involvement:
 Stomach: [...]
 Small bowel disease:
 Disease location: [...]
 Length of disease: [...]
 Enhancement pattern: [...]
 Stricture: [...]
 Penetrating disease:
 Sinus tract: [...]
 Fistula: [...]
 Colon: [...]

Presence of steatorrhea: [...]

EXTRAIESTINAL:
 Mesenteric findings:
 Vasa recta: [...]
 Mesenteric adenopathy: [...]
 Fluid collections: [...]
 Perianal disease: [...]

MSK:
 Sacroiliitis: [...]
 Avascular necrosis: [...]

Other:
 Abdomen and pelvis: [...]
 Lower chest: [...]

Fig. 1. Structured report template for interpretation of CT enterography for patients with inflammatory bowel disease.

radiology residents (R3–R4). Each reader was first provided the 30 CTEs and asked to interpret them using a typical free text report. They were not instructed on any specific features to report, nor were they provided any definitions for feature criteria (i.e., they were not given definitions for what constitutes ‘stricture’). After a wash-out period of at least 4 weeks, they were provided the same 30 examinations, this time to interpret using the provided structured report. In total, the 30 CTE examinations were interpreted by all nine readers, for a total of 270 NSR and 270 SR. All reports (NSR and SR) were constructed using the PowerScribe 360 dictation system (Nuance Communications, Inc., Burlington, MA, USA).

Quantitative analysis

Both the NSR and SR were then assessed in consensus by a 4th year radiology resident and a fellowship-trained abdominal radiologist with 5 years’ post-fellowship experience for the documentation of the presence or absence of 15 key reporting features. In the NSR, mention of a key feature qualified as documentation, whether the finding was positive or negative (e.g., “no abscess”

counted the same as “abscess present”). Similarly, in the SR, documentation in the field of either the presence or the absence of a feature qualified as documentation.

Five of the key features were then analyzed for accuracy: multifocal disease, stricture, fistula, fluid collection, and perianal disease. The reference standard for each case was established in consensus by two unblinded abdominal radiologists with 5 and 12 years’ post-fellowship experience, respectively.

Reader experience

The participating radiologists (the readers) were administered a survey following the completion of the study. The survey featured four subjective questions regarding the radiologist’s experience with SR and his or her opinion of it moving forward. The questions were as follows: (1) Using the structured report was 1 = easy, 2 = difficult but got better with practice, or 3 = consistently bothersome and fatiguing. (2) Moving forward, how would structured reporting for CTE of IBD affect your productivity in the reading room: 1 = definite increase, 2 = slight increase, 3 = no change, 4 = slight decrease, or 5 = definite decrease. (3) Overall, what is your report type preference for reporting CTE of IBD: 1 = strongly prefer structured, 2 = slightly prefer structured, 3 = no preference, 4 = slightly prefer non-structured, or 5 = strongly prefer non-structured. (4) After participating in this study, how likely are you to use structured reporting for CT of other disease processes: 1 = definitely more likely, 2 = somewhat more likely, 3 = neutral, 4 = somewhat less likely, or 5 = definitely less likely.

Qualitative analysis

De-identified CTE reports from the three faculty readers were provided to three referring physicians (a pediatric gastroenterologist with 7 years of post-fellowship experience, a gastroenterology fellow, and a colorectal surgeon with 12 years of post-fellowship experience) who treat patients with IBD in their clinical practices. A total of 30 reports (15 NSR [5 per reader] and 15 SR [5 per reader]) were randomly selected for review by each clinician, who independently evaluated the reports based on three criteria: ease of information extraction (0 = difficult, 1 = needs some effort, or 2 = easy), clarity of patient anatomy (0 = unclear, 1 = mostly clear, or 2 = very clear), and ability to identify disease phenotype [17, 24] (1 = yes, 2 = no).

Statistical analysis

Results were summarized using descriptive statistics. For the quantitative analysis, a mixed-effects linear regression model (to account for the correlation between observations due to using the same cases/readers) was

used to determine whether there was a difference in the number of features per report between NSR and SR and across training levels. A paired *t* test was used to determine differences in each individual feature for NSR versus SR. A two-sample *f* test was performed to assess differences in the variance between the numbers of key features reported between report types.

To assess the association between report type and accuracy, we used generalized linear mixed-effects models with readers and cases as random effects. The same analysis was repeated to investigate such effect for residents only, fellows only, and faculty only.

To evaluate the agreement between readers, we used the Fleiss' Kappa coefficient. We used the bootstrap approach (10,000 bootstrap samples from cases) to estimate confidence intervals for the agreement and to determine whether the difference between the Fleiss' Kappa coefficients for the SR and NSR was statistically significant. Coefficients were then categorized according to the following scale: 0.01–.20 = slight agreement; 0.21–.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–1.0 = almost perfect agreement.

A χ^2 test was used to compare the referring clinician preference in regards to SR versus NSR. Descriptive statistics were calculated for the post-study survey of participating radiologists.

Statistical analysis was conducted using Microsoft Excel (Microsoft, Redmond, WA) and R version 3.2.1 (2015/06/18, www.R-project.org). *P* values of less than 0.05 were considered significant.

Results

Quantitative analysis

Interpretations using NSR documented the presence or absence of 8.2 ± 2.2 key features (range 4–14), while SR documented 14.6 ± 0.5 features (range 13–15) ($p < 0.001$). Increased reporting of both pertinent positive and pertinent negative features comprised this difference, with NSR documenting 5.1 ± 1.5 positive findings compared to 5.9 ± 1.7 for SR ($p < 0.001$), and 3.2 ± 1.8 negative findings for NSR compared to 8.7 ± 1.6 for SR ($p < 0.001$). There was also a decrease in the variability in the number of key features reported, with NSR having a standard deviation of 2.2 compared to 0.5 for SR ($p < 0.001$). When using NSR, there was a gradient from residents (7.5 ± 2.3 features) to faculty (9.1 ± 1.7 features), though this was not significant ($p = 0.32$). With SR, readers across training level reported approximately the same number of features (Fig. 2).

Multiple key features were consistently omitted from the NSR, and the use of SR resulted in increased documentation of many of these features. Specifically, 13 of 15 key features were documented significantly more in the SR including fluid collection and perianal disease.

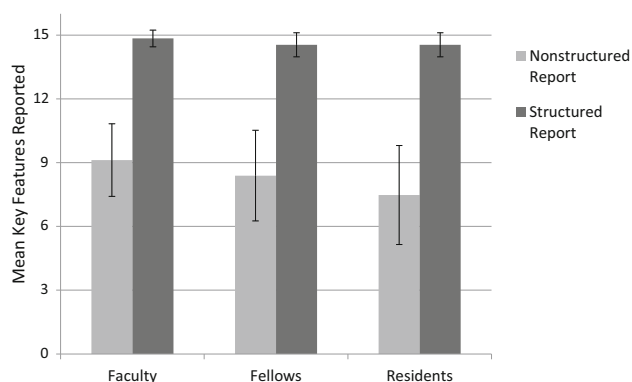


Fig. 2. Number of key features reported by report type and training level. All three levels of training showed a significant increase in key feature reporting when using the structured report.

The remaining key features that were included more frequently in the SR are outlined in Table 1. Disease location was mentioned with similar frequency between report types. Only wall thickening was mentioned less frequently in SR, as this was not specifically included in the template.

Accuracy was assessed for a subset of 5 key features (multifocal disease, stricture, fistula, fluid collection, and perianal disease) for both reporting styles (Table 2). In general, accuracy when using SR was not significantly improved compared to NSR. For reporting on fistula, fluid collection, and perianal disease, accuracy rates were not significantly different for all readers combined or within training levels. There was a small but statistically significant improvement in accuracy for describing multifocal disease, with accuracy increasing from 76% (205/270) with NSR to 83% (224/270) with SR ($p = 0.01$). This overall improvement was manifested by increased accuracy for faculty (84% (76/90) SR vs. 72% (65/90) NSR, $p = 0.01$) but not for fellows or residents. There was a trend toward increased accuracy for reporting on stricture (60% (163/270) SR vs. 54% (147/270) NSR, $p = 0.06$).

The analysis of interobserver variability for the subset of five features showed increased agreement using SR in some instances. For example, agreement was 'fair' for fistula in NSR ($k = 0.21$, 95% CI 0.10–0.31) and increased to 'substantial' for SR ($k = 0.62$, 95% CI 0.43–0.79) ($p < 0.001$). Agreement increased from 'slight' ($k = 0.01$, 95% CI –0.09 to 0.07) to 'fair' ($k = 0.29$, 95% CI –0.03 to 0.54) for perianal disease ($p < 0.001$), from 'slight' ($k = 0.15$, 95% CI 0.06–0.25) to 'fair' ($k = 0.35$, 95% CI 0.18–0.50) for stricture ($p < 0.001$), and from 'slight' ($k = 0.04$, 95% CI –0.03 to 0.14) to 'moderate' ($k = 0.59$, 95% CI –0.02 to 0.89) for fluid collection ($p = 0.25$). Agreement for multifocal

Table 1. Key features described in non-structured versus structured reports

Feature	Described in NSR ^a	Described in SR ^b	<i>p</i> value
Prior surgery	97/270 (36%)	268/270 (99%)	<0.001
Location of disease	269/270 (99%)	269/270 (99%)	1
Enhancement pattern	220/270 (81%)	258/270 (96%)	0.003
Wall thickening	237/270 (88%)	197/270 (73%)	<0.001
Length of disease	203/270 (75%)	262/270 (97%)	<0.001
Stricture	211/270 (78%)	268/270 (99%)	<0.001
Fistula	192/270 (71%)	270/270 (100%)	<0.001
Fibrofatty proliferation	18/270 (7%)	269/270 (99%)	<0.001
Vasa recta enlargement	95/270 (35%)	270/270 (100%)	<0.001
Enlarged lymph nodes	260/270 (96%)	270/270 (100%)	<0.001
Fluid collection	186/270 (69%)	270/270 (100%)	<0.001
Perianal disease	86/270 (41%)	270/270 (97%)	<0.001
Steatorrhea	19/270 (7%)	270/270 (53%)	<0.001
Sacroiliitis	112/270 (41%)	270/270 (100%)	<0.001
Avascular necrosis	43/270 (16%)	269/270 (99%)	<0.001
Average number of features per report	8.2/15	14.6/15	<0.001

^a NSR non-structured report. Number of times a feature was mentioned out of 270 total non-structured reports generated

^b SR structured report. Number of times a feature was mentioned out of 270 total structured reports generated

Table 2. Reporting accuracy between NSR^a and SR^b for a subset of key features, across training levels (numbers in percent accurate)

Feature	Faculty		Fellows		Residents		All Readers	
	NSR	SR	NSR	SR	NSR	SR	NSR	SR
Multifocal disease	72	84, <i>p</i> = 0.01	75	82, <i>p</i> = 0.14	80	82, <i>p</i> = 0.60	76	83, <i>p</i> = 0.01
Stricture	51	64, <i>p</i> = 0.01	60	58, <i>p</i> = 0.71	52	59, <i>p</i> = 0.18	54	60, <i>p</i> = 0.06
Fistula	92	91, <i>p</i> = 0.72	84	86, <i>p</i> = 0.79	78	82, <i>p</i> = 0.20	85	86, <i>p</i> = 0.51
Fluid collection	94	94, <i>p</i> = 0.99	93	94, <i>p</i> = 0.59	96	97, <i>p</i> = 0.58	94	95, <i>p</i> = 0.59
Perianal disease	96	94, <i>p</i> = 0.53	92	92, <i>p</i> = 0.99	95	92, <i>p</i> = 0.13	95	93, <i>p</i> = 0.24

^a NSR non-structured report. Number of times a feature was mentioned out of 270 total non-structured reports generated

^b SR structured report. Number of times a feature was mentioned out of 270 total structured reports generated

Table 3. Subjective assessment of reports by referring clinicians

Criterion	Mean rating for NSR	Number of NSR with highest rating (percent)	Mean rating for SR	Number of SR with highest rating (percent)	<i>p</i> value
Ease of information extraction ^a	1.2	19/45 (42%)	1.7	35/45 (78%)	<0.01
Clarity of anatomy ^b	1.7	32/45 (71%)	1.8	39/45 (89%)	0.07
Ability to identify disease phenotype ^c	0.87	39/45 (87%)	0.98	44/45 (98%)	0.05

^a Graded on a scale of 0–2 (0 = difficult, 1 = needs some effort, 2 = easy)

^b Graded on a scale of 0–2 (0 = unclear, 1 = mostly clear, 2 = very clear)

^c Graded on a 0–1 scale (0 = no, 1 = yes)

disease was ‘substantial’ and did not change ($k = 0.68$ NSR vs. $k = 0.70$ SR, $p = 0.59$).

not demonstrate a significant preference when evaluating reports for clarity of anatomy.

Qualitative analysis

The referring clinicians preferred SR with regard to ease of information extraction (mean score SR 1.7 vs NSR 1.2 on a scale of 0–2, $p < 0.01$) (Table 3). For SR, the clinicians rated 35 of 45 reports (78%) as “easy” regarding ease of information extraction, compared to only 19 of 45 reports (42%) for NSR. There was also a small but significant preference for SR when trying to identify disease phenotype: clinicians rated 44 out of 45 (98%) SR as being helpful for phenotype identification, compared to 39 out of 45 (87%) NSR. The clinicians did

Reader experience

When asked about their experience using SR, two of the nine readers (22%) rated it as “easy,” six (67%) rated it as “difficult but got better with practice,” and only one (11%) rated it as “consistently bothersome and fatiguing.” In regards to productivity for reporting CTE of IBD when using SR, four readers (44%) stated using SR would either “definitely” or “slightly” increase productivity, while the other five readers (56%) did not think their productivity would be affected. None of the readers felt SR would decrease their productivity. When asked

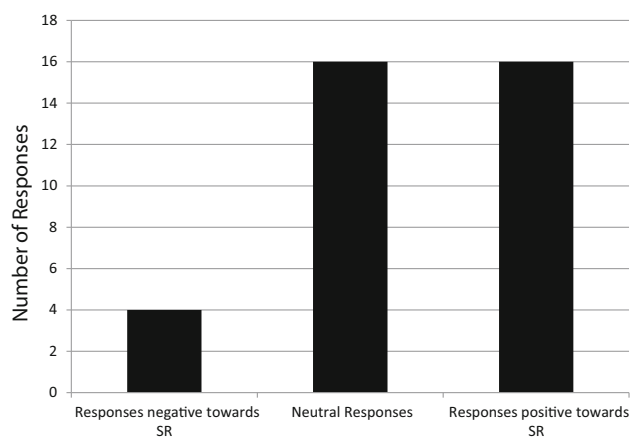


Fig. 3. Reader responses to a survey regarding structured reporting; the majority of responses were neutral or positive.

about using SR for reporting CT of other disease processes, five readers (56%) stated that after completing this project they were “definitely” or “somewhat” more likely to use SR. Only one reader (11%) said that completing this project made him or her “somewhat less likely” to use SR in the future. Out of 36 total responses, only four (11%) were negative toward SR (Fig. 3).

Discussion

SR is an advantageous reporting method to allow thorough documentation of complex disease processes such as IBD. The results of this study demonstrate that when using non-structured reports to describe the findings of inflammatory bowel disease on CTE, radiologists across all levels of training consistently omit numerous key descriptors and pertinent negative findings. However, when using structured reports, radiologists detailed nearly double the number of key features in a more consistent fashion, with both pertinent positive and negative findings reported more frequently. In addition, our referring physicians showed some subjective preference for SR. However, despite improved documentation of findings, overall accuracy and accuracy across training levels were minimally affected.

These data support the growing evidence that SR allows for increased reporting of key features of complex disease processes. Improved reporting using SR has already been described in multiphasic CT for pancreatic cancer [3] as well as in MRI for rectal cancer staging [25], both being complex diseases. In our study, increased reporting was manifest through higher numbers of both positive and negative findings. Radiologists generally identify positive findings regardless of report type; however, even a small increase in positive finding reporting could be clinically significant, and our radiologists reported nearly one full additional positive finding when using SR. On the other hand, reporting of pertinent

negative findings is also important [26, 27], particularly when describing complex disease processes like IBD. In our study, there was a nearly threefold increase in the number of pertinent negative findings described when using SR. Features such as fluid collection, fistula, stricture, and perianal disease were all described significantly more often in SR, even when they were absent. This improvement suggests that unless prompted some pertinent negative findings may be omitted from a report, which could lead to confusion or a sense of incompleteness in the report.

While more features were reported overall, we did not find a significant impact on accuracy when examining a subset of key features including stricture, fistula, fluid collection, or perianal disease. The lower accuracy rate for stricture suggests that the diagnosis is challenging, and when coupled with high interobserver variability, these data highlight the subjectivity and difficulty of reporting stricture on CT. In contrast, multifocal disease was a feature where the use of SR was associated with greater accuracy, though the increase was modest. Our findings in regards to accuracy are similar to those in prior works: Powell et al. found no increase in accuracy when readers used SR to interpret maxillofacial CT [28], and Lin et al. described mixed results regarding accuracy with the use of SR for cervical spine CT [20]. It appears that while more descriptors may be provided with SR for CTE, accuracy may not improve.

When using NSR, there was a gradient across training levels while assessing the number of key features reported. As one might predict, residents reported the fewest key features and faculty the most. This gradient was eliminated when using SR, with all three training levels reporting a similar number of key features. These results are in contradistinction to Johnson et al., who found that residents had decreased completeness of reports when using SR to report on brain MRIs [11]. Nonetheless, increased reporting of key features for trainees can have positive effects. First, using a template that highlights important disease characteristics can enhance learning and help trainees develop an approach for a complex disease. Second, reporting became more consistent. Interestingly, accuracy was not improved among trainees using SR, and in general trainee accuracy was similar to faculty accuracy. Upper-level resident performance has been shown to be similar to that of faculty physicians for other tasks, and these data provide additional support [18].

Subjective analysis by referring physicians suggested a slight preference for SR. The referrers found it easier to extract information from SR and rated SR marginally higher with regard to identifying disease phenotype. The ability to understand anatomy was rated as high for both report types. These findings are similar to studies that have demonstrated preference for templated reports [9, 10, 29]. Despite these results, it is important to recognize

that these are subjective measures and that attitudes may vary between physicians and across institutions. New templates should be reviewed with and tailored to specific groups of referrers. It is also important to note that the complexity of a given case may influence what report style is preferred. When presented with a report describing multifocal disease with certain aspects fluctuating over time, a templated report may prove challenging to read, whereas a relatively 'negative' report might be easy to read in a structured format. This point also applies to the interpreting radiologist, whose experience with SR or NSR may also differ depending on the complexity of a case.

Just as the referring physicians slightly preferred SR, the radiology readers in our study demonstrated some preference for SR. Overall, responses to subjective questions were largely positive or neutral toward SR, with no readers feeling that SR would decrease productivity. These results are in line with multiple studies suggesting radiologist preference for itemized reporting [29, 30].

While our data suggest that SR yields more key features with less variability, SR is not perfect. For example, wall thickening was reported less frequently in SR, possibly because it was not included as a discrete item on the template. Interestingly, wall thickness was documented in 88% of NSR, which means that radiologists using NSR often think about and subsequently describe the feature. The decrease using SR suggests that when using a template, interpreters may fail to comment on features not listed in front of them, features they might normally describe. It appears that the act of populating a standard report may change how readers report information, perhaps negatively in the case of this particular feature. As a result of these findings, we have added wall thickening to our institution's template.

There are limitations to this study. First, a relatively small number of studies were included and the number of positive findings within the dataset for certain key features was low, which affects accuracy data. Next, the use of a template and its inclusion of various features helps prompt readers to report on key features, which may introduce bias toward more key feature reporting for SR. Recall that bias could affect results since each reader reviewed each study twice. A minimum four-week break between reading sessions was used to minimize this bias. In addition, readers knew that their reports were part of a study rather than part of the daily clinical workflow; this too could affect reporting behavior. Another limitation is the reference standard used to establish report accuracy: while two experienced faculty radiologists helped establish the standard, they did not have any surgical or pathologic correlate for each individual feature. Finally, we did not record dictation times or any objective assessment of reader productivity. Future efforts could include studying readers and reports within a

true clinical workflow to better assess the impact on productivity.

Based on our results and the current literature, we feel that for assessment of complex disease processes where there is potential for miscommunication between referrer and radiologist, structured formats could provide benefit. In fact, our institution has incorporated templated reporting of CTE for IBD into its daily practice.

In conclusion, structured reporting of CTE for inflammatory bowel disease improved documentation of key reporting features for trainees and faculty, though there was minimal impact on accuracy. Referring physicians subjectively preferred the structured reports, and radiologists using structured reports did not view them negatively.

Compliance with ethical standards

Funding No funding was received for this study.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Statement of informed consent was not applicable since the manuscript does not contain any patient data.

References

1. European Society of Rad (2010) The future role of radiology in healthcare. *Insights Imaging* 1(1):2–11. doi:10.1007/s13244-009-0007-x
2. Radiological Society of North America (2016). RSNA radiology reporting initiative. Available at: <http://reportingwiki.rsna.org/>. Accessed November 30, 2016
3. Brook OR, Brook A, Vollmer CM, et al. (2015) Structured reporting of multiphasic CT for pancreatic cancer: potential effect on staging and surgical planning. *Radiology* 274(2):464–472. doi:10.1148/radiol.14140206
4. American College of Radiology (2014). Liver imaging reporting and data system. Available at: <https://www.acr.org/Quality-Safety/Resources/LIRADS>. Accessed February 1, 2017
5. Barentsz JO, Richenberg J, Clements R, et al. (2012) ESUR prostate MR guidelines 2012. *Eur Radiol* 22(4):746–757. doi:10.1007/s00330-011-2377-y
6. Radiological Society of North America (2016). Radreport. Available at: <http://www.radreport.org/>. Accessed November 30, 2016
7. Siström CL, Langlotz CP (2005) A framework for improving radiology reporting. *J Am Coll Radiol* 2(2):159–167. doi:10.1016/j.jacr.2004.06.015
8. Reiner BI, Knight N, Siegel EL (2007) Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol* 4(5):313–319. doi:10.1016/j.jacr.2007.01.015
9. Schwartz LH, Panicek DM, Berk AR, Li Y, Hricak H (2011) Improving communication of diagnostic radiology findings through structured reporting. *Radiology* 260(1):174–181. doi:10.1148/radiol.11101913
10. Siström CL, Honeyman-Buck J (2005) Free text versus structured format: information transfer efficiency of radiology reports. *AJR Am J Roentgenol* 185(3):804–812. doi:10.2214/ajr.185.3.01850804
11. Johnson AJ, Chen MY, Swan JS, Applegate KE, Littenberg B (2009) Cohort study of structured reporting compared with conventional dictation. *Radiology* 253(1):74–80. doi:10.1148/radiol.2531090138
12. Weiss DL, Langlotz CP (2008) Structured reporting: patient care enhancement or productivity nightmare? *Radiology* 249(3):739–747. doi:10.1148/radiol.2493080988249/3/739

13. Marcal LP, Fox PS, Evans DB, et al. (2015) Analysis of free-form radiology dictations for completeness and clarity for pancreatic cancer staging. *Abdom Imaging* 40(7):2391–2397. doi:[10.1007/s00261-015-0420-1](https://doi.org/10.1007/s00261-015-0420-1)
14. Wildman-Tobriner B, Allen BC, Davis JT, et al. (2016) Structured reporting of magnetic resonance enterography for pediatric Crohn's disease: effect on key feature reporting and subjective assessment of disease by referring physicians. *Curr Probl Diagn Radiol* . doi:[10.1067/j.cpradiol.2016.12.001](https://doi.org/10.1067/j.cpradiol.2016.12.001)
15. Park MJ, Lim JS (2013) Computed tomography enterography for evaluation of inflammatory bowel disease. *Clin Endosc* 46(4):327–366. doi:[10.5946/ce.2013.46.4.327](https://doi.org/10.5946/ce.2013.46.4.327)
16. Baker ME, Hara AK, Platt JF, Maglinte DD, Fletcher JG (2015) CT enterography for Crohn's disease: optimal technique and imaging issues. *Abdom Imaging* 40(5):938–952. doi:[10.1007/s00261-015-0357-4](https://doi.org/10.1007/s00261-015-0357-4)
17. Leyendecker JR, Bloomfeld RS, DiSantis DJ, et al. (2009) MR enterography in the management of patients with Crohn disease. *Radiographics* 29(6):1827–1846. doi:[10.1148/rg.29609551029/6/1827](https://doi.org/10.1148/rg.29609551029/6/1827)
18. Cooper VF, Goodhart LA, Nemcek AA Jr, Ryu RK (2008) Radiology resident interpretations of on-call imaging studies: the incidence of major discrepancies. *Acad Radiol* 15(9):1198–1204. doi:[10.1016/j.acra.2008.02.011](https://doi.org/10.1016/j.acra.2008.02.011)
19. Ruchman RB, Jaeger J, Wiggins EF 3rd, et al. (2007) Preliminary radiology resident interpretations versus final attending radiologist interpretations and the impact on patient care in a community hospital. *AJR Am J Roentgenol* 189(3):523–526. doi:[10.2214/AJR.07.2307](https://doi.org/10.2214/AJR.07.2307)
20. Lin E, Powell DK, Kagetsu NJ (2014) Efficacy of a checklist-style structured radiology reporting template in reducing resident misses on cervical spine computed tomography examinations. *J Digit Imaging* 27(5):588–593. doi:[10.1007/s10278-014-9703-2](https://doi.org/10.1007/s10278-014-9703-2)
21. Collard MD, Tellier J, Chowdhury AS, Lowe LH (2014) Improvement in reporting skills of radiology residents with a structured reporting curriculum. *Acad Radiol* 21(1):126–133. doi:[10.1016/j.acra.2013.10.007](https://doi.org/10.1016/j.acra.2013.10.007)
22. Hofmann AF, Poley JR (1972) Role of bile acid malabsorption in pathogenesis of diarrhea and steatorrhea in patients with ileal resection. I. Response to cholestyramine or replacement of dietary long chain triglyceride by medium chain triglyceride. *Gastroenterology* 62(5):918–934
23. Sandborn WJ (2009) How to avoid treating irritable bowel syndrome with biologic therapy for inflammatory bowel disease. *Dig Dis* 27(Suppl 1):80–84. doi:[10.1159/000268125](https://doi.org/10.1159/000268125)
24. Maglinte DD, Gourtsoyiannis N, Rex D, Howard TJ, Kelvin FM (2003) Classification of small bowel Crohn's subtypes based on multimodality imaging. *Radiol Clin North Am* 41(2):285–303
25. Sahni VA, Silveira PC, Sainani NI, Khorasani R (2015) Impact of a structured report template on the quality of MRI reports for rectal cancer staging. *AJR Am J Roentgenol* 205(3):584–588. doi:[10.2214/AJR.14.14053](https://doi.org/10.2214/AJR.14.14053)
26. European Society of R (2011) Good practice for radiological reporting. Guidelines from the European society of radiology (ESR). *Insights Imaging* 2(2):93–96. doi:[10.1007/s13244-011-0066-766](https://doi.org/10.1007/s13244-011-0066-766)
27. Kushner DC, Lucey LL, American College of R (2005) Diagnostic radiology reporting and communication: the ACR guideline. *J Am Coll Radiol* 2(1):15–21. doi:[10.1016/j.jacr.2004.08.005](https://doi.org/10.1016/j.jacr.2004.08.005)
28. Powell DK, Lin E, Silberzweig JE, Kagetsu NJ (2014) Introducing radiology report checklists among residents: adherence rates when suggesting versus requiring their use and early experience in improving accuracy. *Acad Radiol* 21(3):415–423. doi:[10.1016/j.acra.2013.12.004](https://doi.org/10.1016/j.acra.2013.12.004)
29. Bosmans JM, Weyler JJ, De Schepper AM, Parizel PM (2011) The radiology report as seen by radiologists and referring clinicians: results of the COVER and ROVER surveys. *Radiology* 259(1):184–195. doi:[10.1148/radiol.10101045](https://doi.org/10.1148/radiol.10101045)
30. Naik SS, Hanbidge A, Wilson SR (2001) Radiology reports: examining radiologist and clinician preferences regarding style and content. *AJR Am J Roentgenol* 176(3):591–598. doi:[10.2214/ajr.176.3.1760591](https://doi.org/10.2214/ajr.176.3.1760591)