GASTROINTESTINAL

Grading of Crohn's disease activity using CT, MRI, US and scintigraphy: a meta-analysis

C. A. J. Puylaert[1] · J. A. W. Tielbeek[1] · S. Bipat[1] · J. Stoker[1]

Abstract

Purpose To assess the grading of Crohn's disease activity using CT, MRI, US and scintigraphy.

Materials and methods MEDLINE, EMBASE and Cochrane databases were searched (January 1983–March 2014) for studies evaluating CT, MRI, US and scintigraphy in grading Crohn's disease activity compared to endoscopy, biopsies or intraoperative findings. Two independent reviewers assessed the data. Three-by-three tables (none, mild, frank disease) were constructed for all studies, and estimates of accurate, over- and under-grading were calculated/summarized by fixed or random effects models.

Results Our search yielded 9356 articles, 19 of which were included. Per-patient data showed accurate grading values for CT, MRI, US and scintigraphy of 86 % (95 % CI: 75–93 %), 84 % (95 % CI: 67–93 %), 44 % (95 % CI: 28–61 %) and 40 % (95 % CI: 16–70 %), respectively. In the per-patient analysis, CT and MRI showed similar accurate grading estimates ($P=0.8$). Per-segment data showed accurate grading values for CT and scintigraphy of 87 % (95 % CI: 77–93 %) and 86 % (95 % CI: 80–91 %), respectively. MRI and US showed grading accuracies of 67–82 % and 56–75 %, respectively.

Conclusions CT and MRI showed comparable high accurate grading estimates in the per-patient analysis. Results for US

✉ C. A. J. Puylaert
  c.a.puylaert@amc.uva.nl

[1] Academic Medical Center, Department of Radiology, University of Amsterdam, Meibergdreef 9, Amsterdam 1105AZ, The Netherlands

and scintigraphy were inconsistent, and limited data were available.

Key Points
• *CT and MRI have comparable high accuracy in grading Crohn's disease.*
• *Data on US and scintigraphy is inconsistent and limited.*
• *MRI is preferable over CT as it lacks ionizing radiation exposure.*

Introduction

Cross-sectional imaging techniques are widely used for diagnosis and evaluation of Crohn's disease. Numerous studies have evaluated the diagnostic accuracy of cross-sectional imaging techniques in patients with Crohn's disease, and a meta-analysis was published that investigated the diagnostic accuracy of computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US) and scintigraphy [1]. However, clinical monitoring and choice of therapy largely rely on grading of disease activity.

Clinical symptoms and inflammatory lesions can exist independently, so assessment of the bowel is essential in guiding therapy decisions [2]. If inflammation is present, it is important to distinguish between mild, moderate and severe disease, as medical management differs among these stages [3]. Ileocolonoscopy, the current reference standard for luminal Crohn's disease, is accurate for assessing mucosal abnormalities, but it has several drawbacks, as it is an invasive technique, is associated with the risk of bowel perforation, is incapable of assessing trans- and extraluminal disease, and is

🖄 Springer

limited to the colon and terminal ileum [4]. Video capsule endoscopy (VCE) is a well-tolerated and accurate alternative to ileocolonoscopy that allows assessment of the whole gastrointestinal tract, although it has shown lower specificity and bears the risk of capsule retention, which occurs in up to 13 % of patients with Crohn's disease [5].

Cross-sectional imaging techniques that could accurately grade disease severity would be preferable to ileocolonoscopy, as they are non-invasive and not limited to the colon and terminal ileum. Several studies have looked at the use of cross-sectional imaging for assessing the severity of Crohn's disease, but offered no comparison between imaging techniques, as no meta-analysis was performed [2, 6]. To our knowledge, only one such meta-analysis has been performed, but it evaluated only MRI and used a search period that ended in April 2007 [7]. This study showed that MRI correctly graded disease activity in 91 % of patients with frank (moderate-to-severe) disease. However, correct grading was limited in patients with disease in remission and with mild disease (62 % for both). Furthermore, no comparison with other imaging techniques was made and numerous articles on the grading of Crohn's disease using MRI have been published since 2007.

Our purpose was to systematically review and compare the accuracy of CT, MRI, US, scintigraphy and positron emission tomography–computed tomography (PET-CT) in grading Crohn's disease activity on a per-patient or per-segment basis as compared to endoscopy, biopsies or intraoperative findings by performing a meta-analysis. Furthermore, we aimed to investigate the degree of over- and under-grading for these imaging techniques.

Material and methods

This review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [8]. The review protocol was not published or registered in advance.

Literature search and strategy

We performed an electronic search in MEDLINE, EMBASE and Cochrane databases for studies examining the accuracy of CT, MRI, US, scintigraphy and PET (-CT) for grading Crohn's disease activity in human subjects. Search terms 'Crohn's disease' and 'inflammatory bowel disease' were combined using 'OR' and search terms for imaging modalities were combined using 'OR' as well. These two groups were combined using 'AND'. The search period was limited from January 1983 to March 2014. Details of the search strategy are provided in the electronic supplementary material (Appendix E1).

Study selection on title and abstract

All articles retrieved from the electronic search were assessed by one observer (CP). Non-relevant articles and articles in the form of a review, case report, comment or letter were excluded. Subsequently, the remaining titles and abstracts were independently assessed by two observers (CP, JT) to identify potentially eligible articles. In cases of uncertainty, articles were deemed potentially eligible and retrieved as full text.

Study selection on full text

The full texts of the remaining articles were retrieved. Two observers (CP, JT) independently reviewed all eligible articles for the following inclusion criteria: (a) ten or more patients were included (fewer were considered case-series); (b) CT, MRI, US, scintigraphy or PET (-CT) was used to grade Crohn's disease activity; (c) patients with clinically suspected inflammatory bowel disease (IBD) or known IBD/Crohn's disease were included; (d) endoscopy, biopsies or intraoperative findings were used as a reference test; (e) imaging features used for grading disease activity were defined; (f) raw data were available to construct 3×3 tables; (g) articles were written in English, Italian, Spanish, French, German or Dutch; and (h) patients with Crohn's disease could be analysed separately from other IBD patients. No patient age limits were applied. Articles in the form of a review, case report, conference abstract, comment or letter were excluded. In the case of duplicate publications, we excluded the studies with the lower number of patients. Disagreement regarding potential eligibility and inclusion was resolved by consensus. The observers were not blinded to author and journal names.

Study characteristics

Methodological characteristics Both reviewers extracted study characteristics independently for all included articles using a standardized form. To assess the quality of the study design, we used a modified Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS 2) tool [9, 10], as it separately assesses risk of bias in several methodological domains (patient selection, index test, reference test and patient flow) using a number of signalling questions (Table 1). Risk of bias for each domain was described as high, low or unclear. In addition, concerns regarding the applicability of the patient population, index and reference test to the review question were rated by the observers as high, low or unclear. Disagreements were resolved by discussion.

Patient characteristics The following patient characteristics were recorded: number of patients included, number of patients in the analysis, whether patients were

Table 1 Methodological characteristics from the QUADAS tool and their corresponding signalling questions [9, 10]. The risk of bias is determined for every domain using the signalling questions

Modified QUADAS Methodological Characteristics

Domain	Signalling questions (Yes, no, unclear)
Patient selection	Was a consecutive or random sample of patients enrolled? Was a case–control design avoided? Did the study avoid inappropriate exclusions?
Index test	Were the index test results interpreted without knowledge of the results of the reference test? If a threshold was used, was it pre-specified? Was the execution and interpretation (expertise, image analysis) of the index test described in sufficient detail to permit its replication? [a]
Reference test	Is the reference test likely to correctly classify the target condition? Were the reference test results interpreted without knowledge of the results of the index test? Was the execution and interpretation of the reference test described in sufficient detail to permit its replication? [a]
Patient flow	Was there an appropriate interval between index test(s) and reference test (>1 month)? Did all patients receive a reference test? Did all patients receive the same reference test? Were all patients included in the analysis?
Prospective / Retrospective [b]	Was the data collected after the research question was defined?

[a] This signalling question was added from QUADAS 1 [10]. We considered this question essential for quality assessment, while it is not part of QUADAS 2 [9]
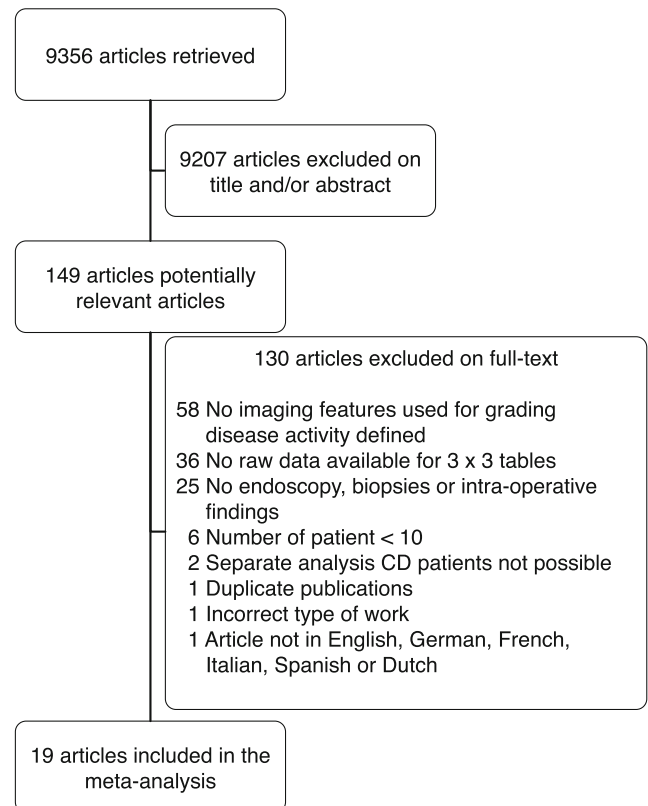
[b] This item was not part of the QUADAS tool

recruited consecutively, age characteristics, gender ratio, patient spectrum (i.e. known or suspected IBD or Crohn's disease) and other selection criteria for patient inclusion.

Imaging characteristics Imaging characteristics concerning type of equipment and basic specifications (type of scanner for CT, field strength and coil type for MRI, and transducer type for US), techniques used for evaluation (sequences for MRI, use of Doppler for US, labelling target and tracer type for scintigraphy), bowel preparation (fasting and/or laxatives), use of luminal and/or intravenous contrast medium, timing of post-contrast scans and use of spasmolytic drugs were extracted.

Reference test All reference tests (i.e. endoscopy, biopsies or intraoperative findings) used for analysis were recorded.

Imaging and reference test interpretation We recorded the following information regarding interpretation of imaging and reference tests: the interval in days between index and reference tests, bowel segments that were examined, grading criteria used for imaging and reference tests, imaging features used for evaluation of disease activity, and whether grading was performed on a per-patient and/or per-bowel-segment basis.

9356 articles retrieved

9207 articles excluded on title and/or abstract

149 articles potentially relevant articles

130 articles excluded on full-text

58 No imaging features used for grading disease activity defined
36 No raw data available for 3 x 3 tables
25 No endoscopy, biopsies or intra-operative findings
6 Number of patient < 10
2 Separate analysis CD patients not possible
1 Duplicate publications
1 Incorrect type of work
1 Article not in English, German, French, Italian, Spanish or Dutch

19 articles included in the meta-analysis

CD = Crohn's disease
Fig. 1 Flow diagram showing study selection

Fig. 2 QUADAS signalling questions (Table 1) per domain (from up to down: patient selection, index test, reference test and patient flow). The last column shows whether studies included patients prospectively.
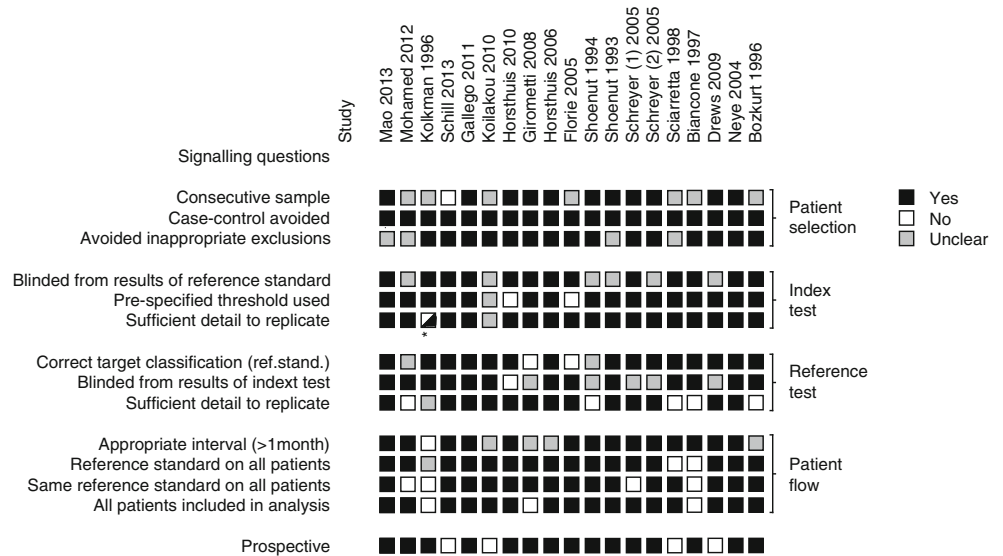
* For index tests CT and scintigraphy, the answer here were **no** and **yes**, respectively. Any other signaling questions regarding CT and scintigraphy for this study were answered identically.

Data extraction

Grading results for imaging and reference tests were extracted with the grading scales used in individual studies (i.e. three-, four-, or five-grade scales). From this data, three-by-three contingency tables comparing results from index and reference tests categorized as none, mild or frank disease were constructed for each study. These categories did not use predefined criteria, but were formed either by using the original grading from each study (in the case of a three-grade scale) or by merging certain grades to form a three-grade scale. If a four-grade scale was used (none, mild, moderate or severe disease), groups with moderate and severe disease were merged into frank disease. For five-grade scales, the second and third scales were grouped into mild disease and the fourth and fifth were grouped into frank disease. When studies used multiple reference tests, we used intraoperative findings as the reference standard. In other cases,

histological findings from biopsies were preferred over endoscopic findings. Because the imaging results in these studies were based on the most severe lesion, we considered histological data from biopsies as more lesion-specific and better resembling imaging results than endoscopic results.

Publication bias

To study publication bias, we followed the method by Deeks et al., as recommended in the Cochrane handbook for DTA reviews [11]. We first calculated effective sample sizes (ESS) for each study. We then performed linear regression analyses if enough datasets were available in a group ($n>5$), with the proportion of accurate grading per study as the independent variable and $1/\sqrt{ESS}$ as the dependent variable. A significant regression coefficient ($P<0.05$) was deemed sufficient to indicate publication bias.

Fig. 3 QUADAS risk of bias per domain and concerns regarding applicability for domains of patient selection, index test and reference test
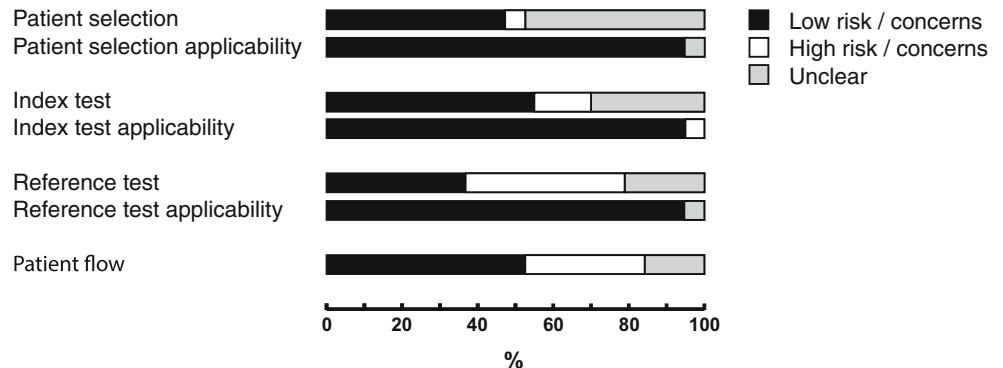
Table 2 Study characteristics

Study	Imaging modality	No. of patients included	No. of patients in analysis	Consecutive	Age, mean (range) or mean±SD	Male/female ratio	Patient spectrum	Inclusion criteria
Mao 2013 [17]	CT	32	32	Y	30 (18–51)[b]	22:10	Known CD	Suspected recurrence after ileocolic resection
Mohamed 2012 [16]	CT	26	26	?	43.4 (19–69)	18:8	Known CD	Referred to further assessment with CTE
Kolkman 1996 [18]	CT, SG	32	17	?	36 (17–65)	11:6	Known/suspected IBD	Suspected IBD or IBD exacerbations or suspected abdominal complications
Schill 2013 [29]	MRI	76	76	N	31.5 (16–76)[b]	40:36	Known CD	Patients scheduled for CD surgery
Gallego 2011 [28]	MRI	61	61	Y	36.1 (14–65)	29:32	Known CD	NA
Koilakou 2010 [27]	MRI	26	26	?	36.5 (22–69)[b]	16:13	Known CD	Patients with previous ileocolic resection
Horsthuis 2010 [26]	MRI	33	15	Y	14 (8–17)[a,b]	15:18[a]	Suspected IBD	Age 8–18 years
Girometti 2008 [25]	MRI	52	45	Y	42.5 (18–67)	23:29[a]	Known/suspected CD	Referred for CC with biopsy and MRI for relapse or suspected onset of CD
Horsthuis 2006 [24]	MRI	20	20	Y	36±13	7:13	Known CD	Scheduled for CC
Florie 2005 [21]	MRI	31	31	?	36±12	22:9	Known CD	Scheduled for ICC because of clinical suspicion of relapsing CD
Shoenut 1994 [20]	MRI	20	12	Y	42.6 (20–70)[a]	12:8[a]	Suspected IBD	NA
Shoenut 1993 [19]	MRI	28	19	Y	34.1 (20–58)	17:11[a]	Known IBD	Referred to MRI for evaluation and on medical therapy
Schreyer 2005 [22]	MRI	30	30	Y	29 (18–65)	8:22	Known CD	Routine small bowel MRI
Schreyer 2005 [23]	MRI	22	12	Y	33.4 (19–55)	5:7	Known/suspected IBD	NA
Drews 2009 [32]	US	32	32	N	38.8 (17–71)	14:18	Known CD	NA
Neye 2004 [31]	US	22	22	Y	33.7 (16–56)	9:13	Known CD	Referred to gastroenterologist
Bozkurt 1996 [30]	US	88	32	?	39 (16–87)[a]	48:40[a]	Suspected IBD	NA
Biancone 1997 [34]	SG	17	10	?	43±11[a]	9:8[a]	Known CD	Patients 6–12 months after ileocecal resection
Sciarretta 1998 [33]	SG	103	31	?	38.3 (15–78)[a]	54:49[a]	Suspected IBD	NA

CC colonoscopy, *CD* Crohn's disease, *CTE* computed tomography enterography, *IBD* inflammatory bowel disease, *ICC* ileocolonoscopy, *MRI* magnetic resonance imaging, *NA* not applicable, *Y* yes, *N* no, *?* unclear

[a] Values reflect the total number of patients included in their respective studies, not only the patients used in this analysis

[b] Median (range)

Table 4 MRI characteristics

Study	Field strength	Coil	Bowel preparation	Luminal contrast	Enterography (EG) / enteroclysis (EC)
Schill 2013 [29]	1.5 T	Body and spine array coils	NS	1.5–2 L Mannitol solution orally 45 min prior and 0.5–1 L 0.9 % NaCl rectally	EG
Gallego 2011 [28]	1.0 T	Body coil	8 hrs fasting	1.5 L PEG and mineral salts	EG
Koilakou 2010 [27]	1.5 T	NS	NS	100–150 mL/min 0.5 % methylcellulose solution	EC
Horsthuis 2010 [26]	3.0 T	Phased array coil	Metamucil in 250 mL water 4 hrs prior	NS	EG
Girometti 2008 [25]	1.5 T	Phased array coil	8 hrs fasting	2 L PEG	EG
Horsthuis 2006 [24]	3.0 T	Phased array body coil	Metamucil in 250 mL water 4 hrs prior	NS	EG
Florie 2005 [21]	1.5 T	NS	4 hrs fasting	1 L water 2 hrs prior	EG
Shoenut 1994 [20]	1.5 T	NS	NS	None	NA
Shoenut 1993 [19]	1.5 T	NS	NS	NS	NA
Schreyer 2005 [22]	1.5 T	Phased array body coil	12 hrs fasting	2 L Mannitol solution with carob seed 1 hr prior orally and 0.4–1.0 L 0.9 % NaCl rectally	NA
Schreyer 2005 [23]	1.5 T	Phased array body coil	Macrogol 3350	1.5 L Gd (5 mmol/L) mixture with water rectally	NA

Study	I.V. contrast	Spasmolytic agent	Post-contrast sequence timing	Sequences
Schill 2013 [29]	0.2 mL/kg Gd-DTPA	40 mg Buscopan iv	70 s	3D T2-SPACE, bSSFP, RARE, T1-FLASH, T1-FLASH (post-contrast), T1-FLASH with fat suppression (post-contrast),
Gallego 2011 [28]	0.1 mmol/kg Gd-DTPA	Buscopan iv	40 s, 70 s (used for RCE), 120 s, 180 s	bSSFP, interpolated 3D T1 with fat suppression (pre-/post-contrast), RARE
Koilakou 2010 [27]	0.1 mmol/kg Gd-DTPA	20 mg Buscopan iv	NS	Interpolated 3D T1 with fat suppression (post-contrast), SSFP, T2 with fat suppression, RARE
Horsthuis 2010 [26]	0.1 mL/kg Gadodiamide	Buscopan iv	NS	Interpolated 3D T1 with fat suppression (post-contrast), RARE
Girometti 2008 [25]	0.2 mL/kg Gd-DTPA	10 mg Buscopan iv	30 s, 45 s, 60 s, 75 s, 90 s, 150 s	bSSFP, cine, interpolated 3D T1 with fat suppression (pre-/post-contrast), RARE
Horsthuis 2006 [24]	0.05 mmol/kg Gadodiamide	20 mg Buscopan iv or 1 mg glucagon hydrochloride	70 s	bFFE, T2-TSE, T1-FFE with fat suppression (post-contrast)
Florie 2005 [21]	0.1 mmol/kg Gd-DTPA	20 mg Buscopan iv or 1 mg glucagon hydrochloride	NS	bSSFP, interpolated 3D T1 (pre-/post-contrast), out-of-phase fast low angle shot, RARE
Shoenut 1994 [20]	0.1 mmol/kg Gd-DTPA	NS	5 s, 30 s (used for RCE), 5 min, 10 min	T1-FLASH (pre-/post-contrast), T1 spin echo with fat suppression (post-contrast)
Shoenut 1993 [19]	0.1 mmol/kg Gd-DTPA	NS	5 s, 30 s (used for RCE), 10 min	

Table 4 (continued)

Study	I.V. contrast	Post-contrast sequence timing	Spasmolytic agent	Sequences
				T1-FLASH, T1-FLASH (post-contrast), T1 with fat suppression (pre-/post-contrast)
Schreyer 2005 [22]	0.2 mmol/kg Gd-DTPA	70 s	40 mg Buscopan iv	2D T1-FLASH, 2D & 3D T1-FLASH with fat suppression (post-contrast), bSSFP, RARE
Schreyer 2005 [23]	0.1 mmol/kg Gd-DTPA	NS	40 mg Buscopan iv	2D T1-FLASH, 2D T1-FLASH fat sat (post-contrast), 3D T1-FLASH, bSSFP, RARE,

bFFE balanced fast-field echo, *(b)SSFP* (balanced) steady-state free precession, *Gd(-DTPA)* gadolinium(-diethylenetriaminepentaacetic acid), *iv* intravenous, *MRI* magnetic resonance imaging, *NaCl* sodium chloride, *NS* not specified, *PEG* polyethylene glycol, *RARE* rapid acquisition with refocusing echoes, *RCE* relative contrast enhancement, *T* Tesla, *T1-FLASH* T1-weighted fast low-angle shot, *TSE* turbo spin-echo

Table 5 US characteristics

Study	Transducer type+frequency	Bowel preparation	Luminal contrast	I.V. contrast	Doppler+type
Drews 2009 [32]	Linear 5–12 MHz (neoterminal ileum) and convex 2–5 MHz (entire abdomen)	NS	NS	NS	Power Doppler
Neye 2004 [31]	Linear 5–12 Mhz and dynamic sector scanner 4–7 MHz	NS	NS	NS	Pulsed Doppler and Power Doppler
Bozkurt 1996 [30]	Linear 7.5 MHz	NS	NS	NS	NA

MHz megahertz, *NA* not applicable, *NS* not specified, *US* ultrasound

articles evaluating PET-CT were found that met our criteria.

Study characteristics

Methodological characteristics Evaluation of the imaging tests was performed blinded from the reference test in 13 studies [17, 18, 21, 22, 24–30, 33, 34]. The reference test was performed blinded to the imaging results in 12 studies [16, 17, 19, 21, 24, 26–30, 33, 34]. The remaining studies did not specify whether observers were blinded to other results [20, 23, 31, 32]. Fifteen of the studies included patients prospectively [16–26, 28, 30, 31, 34]. Signalling questions for the QUADAS tool were answered with 'yes' in 78.9 % of cases (Fig. 2). Patient selection and index test domains showed less risk of bias than reference test and patient flow domains. Concern about applicability of patient selection and index and reference tests was generally low (Fig. 3).

Patient characteristics A total of 549 patients were included (75 for CT, 347 for MRI, 86 for US, and 58 for scintigraphy). The mean study size was 29 patients (range, 10–76). Study characteristics are presented in Table 2. In ten of the studies, patients were recruited consecutively [17, 19, 20, 22–26, 28, 31]. Studies included patients with clinically suspected IBD, known IBD/Crohn's disease, or a combination of both (12, 4, and 3 studies, respectively).

Imaging characteristics Imaging equipment and specifications are presented in Tables 3, 4, 5 and 6. Bowel preparation (fasting and/or laxatives) was used in eight studies (1 CT, 7 MRI) [17, 21–26, 28]. Luminal contrast medium was used in ten studies (3 CT, 7 MRI) [16–18, 21–23, 25, 27–29], of which one used enteroclysis [27]. Intravenous contrast medium was used in 13 studies (2 CT, 11 MRI) [16, 17, 19–29].

Reference test Endoscopy, biopsies and intraoperative findings were used in 11, 8 and 4 studies, respectively (Table 7). Three studies recorded results for both endoscopy and histology from biopsies, for which we used the histological data in our analysis [30, 33, 34].

Imaging and reference test interpretation Thirteen of the studies used an interval of less than one month between imaging and reference test [17, 19–23, 26, 28, 29, 31–34]. The imaging features most commonly used for evaluation were bowel wall thickness and post-contrast enhancement (or tracer uptake for scintigraphy), which were both used in 17 studies (Table 7). The reference test and imaging criteria for each study are presented in Tables 8 and 9.

Table 6 Scintigraphy characteristics

Study	Labelling target	Tracer	Amount of tracer	Scans	Criteria used for image analysis
Kolkman 1996 [18]	Antigranulocyte antibodies	Tc-99 m HMPAO	NS	2 scans (at 1 hrs and 4 hrs)	Uptake of tracer compared to bone marrow and liver
Biancone 1997 [34]	Leukocytes	Tc-99 m HMPAO	185 MBq	2 scans (at 30 min and 3 hrs)	Uptake of tracer compared to bone marrow and liver
Sciarretta 1998 [33]	Leukocytes	Tc-99 m HMPAO	370–555 MBq	3 scans (at 30 min, 2–2.5 hrs and 24 hrs)	Uptake of tracer compared to bone marrow and liver

MBq megabecquerel, *NS* not specified, *Tc-99 m HMPAO* technetium hexamethylpropyleneamine oxime

Table 7 Imaging and reference test interpretation

Study	Imaging modality	Reference test used in analysis	Analysis per patient/ per segment	Time interval (days) index & reference test
Mao 2013 [17]	CT	ICC	Patient	<=7
Mohamed 2012 [16]	CT	B, SS	Patient	<=7 [a]
Kolkman 1996 [18]	CT, SG	B, ICC, SS	Segment	1–50
Schill 2013 [29]	MRI	SS	Patient	<=28
Gallego 2011 [28]	MRI	ICC	Patient	<=15
Koilakou 2010 [27]	MRI	ICC	Patient	NS
Horsthuis 2010 [26]	MRI	EGD, ICC	Patient	<=14
Girometti 2008 [25]	MRI	B	Patient	NS
Horsthuis 2006 [24]	MRI	ICC	Patient	1–48
Florie 2005 [21]	MRI	ICC	Patient	<=14
Shoenut 1994 [20]	MRI	B	Patient	<=3
Shoenut 1993 [19]	MRI	CC, SS	Patient	<=7
Schreyer 2005 [22]	MRI	ICC	Segment	<=7
Schreyer 2005 [23]	MRI	ICC	Segment	1
Drews 2009 [32]	US	B	Patient	<=5
Neye 2004 [31]	US	ICC	Segment	<=3
Bozkurt 1996 [30]	US	B	Segment	NS
Biancone 1997 [34]	SG	B	Patient	<=14
Sciarretta 1998 [33]	SG	B	Segment	<=7

Study	Grading scale index test	Grading scale reference test	Imaging features used for grading disease activity
Mao 2013 [17]	0–3	i0–i4 (Rutgeerts score)	Bowel wall thickness, post-contrast enhancement, mucosal irregularities/hyperdensities, mural stratification, stenosis and prestenotic dilatation and extraluminal findings (lymph nodes, abscesses, fistulas, comb sign, creeping fat)
Mohamed 2012 [16]	Mild, moderate, severe	Mild, moderate, severe	Bowel wall thickness, post-contrast enhancement, extraluminal findings (lymph nodes, abscesses, fistulas, comb sign, creeping fat, edema)
Kolkman 1996 [18]	0–3 (CT), 0–4 (SG)	0–3	Bowel wall thickness, T1 enhancement and pattern, ulceration, double-halo sign and extraluminal findings (creeping fat, mesenteric fibrovascular strands) (CT). Uptake of tracer compared to bone marrow and liver (SG)
Schill 2013 [29]	B1, B2, B3 (Montreal class.)	B1, B2, B3 (Montreal class.)	Target sign, T2 mural signal intensity, inflammatory mass, stenosis with prestenotic dilatation and extraluminal findings (lymph nodes, abscesses, fistulas, comb sign)
Gallego 2011 [28]	None, mild, moderate/severe	0–3 (SES-CD)	Bowel wall thickness and edema, T1 enhancement, mucosal abnormalities, inflammatory mass, motility, stenosis and extraluminal findings (lymph nodes, fistulas)

Table 7 (continued)

Study	Grading scale index test	Grading scale reference test	Imaging features used for grading disease activity
Koilakou 2010 [27]	0–3	i0–i4 (Rutgeerts score)	Bowel wall thickness, T1 enhancement, T2 mural signal, mucosal irregularities, infiltrate, edema, stenosis and prestenotic dilatation, extraluminal findings (abscesses, fistulas)
Horsthuis 2010 [26]	None, mild, moderate, severe (subjective)	None, mild, moderate, severe (subjective)	Bowel wall thickness, T1 enhancement, stenosis and prestenotic dilatation.
Girometti 2008 [25]	None, mild, moderate/severe	None, mild, moderate/severe	Bowel wall thickness, T1 enhancement, mucosal abnormalities, inflammatory mass, mesenteric involvement, motility, stenosis and extraluminal findings (lymph nodes, fistulas)
Horsthuis 2006 [24]	None, mild, moderate, severe (subjective)	None, mild, moderate, severe (subjective)	Bowel wall thickness, T1 enhancement, ulceration, length of diseased segment, cobblestoning, extraluminal findings (lymph nodes, abscesses, fistulas, comb sign and creeping fat)
Florie 2005 [21]	None, mild, moderate, severe (subjective)	None, mild, moderate, severe (subjective)	Bowel wall thickness, T1 enhancement, stenosis, target sign, cobblestoning
Shoenut 1994 [20]	Mild, moderate, severe	Mild, moderate, severe (subjective)	Bowel wall thickness, T1 enhancement, length of diseased segment
Shoenut 1993 [19]	Mild, moderate, severe	Mild, moderate, severe	Bowel wall thickness, T1 enhancement, length of diseased segment
Schreyer 2005 [22]	0–2	0–2	Bowel wall thickness, T1 enhancement, stenosis, lymph nodes, local injection for inflammation assessment
Schreyer 2005 [23]	0–2	0–2	Bowel wall thickness, T1 enhancement, lymph nodes, mesenteric injection
Drews 2009 [32]	0–4	0–4	Vascularization and thickness of the bowel wall, preservation of five-layer structure, length of diseased segment
Neye 2004 [31]	0–3	0–3	Vascularization and thickness of the bowel wall
Bozkurt 1996 [30]	0–2 (subjective)	0–2 (subjective)	Bowel wall thickness, echogenicity of the bowel wall, smoothness of boundaries, visibility of individual bowel wall layers
Biancone 1997 [34]	0–3	0–3 (subjective)	Uptake of tracer compared to bone marrow and liver
Sciarretta 1998 [33]	0–3	0–3 (subjective)	Uptake of tracer compared to bone marrow and liver

B biopsies, *CC* colonoscopy, *CT* computed tomography, *EGD* esophagogastroduodenoscopy, *ICC* ileocolonoscopy, *MRI* magnetic resonance imaging, *NS* not specified, *SES-CD* simple endoscopic score for Crohn's disease, *SG* scintigraphy, *SS* surgical specimens, *TI* terminal ileum, *US* ultrasound

[a] Time interval was not specified for patients undergoing surgery

Table 8 Original reference test criteria and categorization for this study

Study	None	Mild	Moderate
Mao et al. histological score (Rutgeerts score) [17]	i0: No lesions	i1: Less than 5 aphthous lesions	i2: More than 5 aphthous lesions with normal mucosa between the lesions or skip areas of larger lesions or lesions confined to ileocolonic anastomosis
Mohamed et al. histological score (subjective) [16]	-	Mild	
Kolkman et al histological score [18]	0: No abnormalities, or plain fibrosis	1: Some infiltration of polymorphonuclear leukocytes, no ulceration	
Schill et al. surgical score (based on Montreal classification) [29]	-	B1: Non-stricturing and non-penetrating	
Gallego et al. endoscopic score (SES–CD) [a] [28]	0–2 points: Inactive	3–6 points: Mild disease	
Koilakou et al. histological score (Rutgeerts score) [18]	i0: No lesions	i1: Less than 5 aphthous lesions	i2: More than 5 aphthous lesions with normal mucosa between the lesions or skip areas of larger lesions or lesions confined to ileocolonic anastomosis
Horsthuis et al. endoscopic score (subjective) [26]	No disease	Mild disease	
Girometti et al. histological score (subjective) [25]	No disease (or chronic, quiescent disease)	Mild disease	
Horsthuis et al. endoscopic score (subjective) [24]	No disease	Mild disease	
Florie et al. endoscopic score (subjective) [21]	No disease	Mild disease	
Shoenut et al. endoscopic score [20]	-	Mild: Mucosal erythema, friability and granularity	
Shoenut et al. histological score (subjective) [19]	-	Mild disease	
Schreyer et al. endoscopic score [22]	0: No findings	1: Erythema, decreased or absent vascular pattern, friability of the mucosa, single or multiple aphthous lesions, and small ulcers	
Schreyer et al. endoscopic score [23]	0: No findings	1: Erythema, decreased or absent vascular pattern, friability of the mucosa, single or multiple aphthous lesions, and small ulcers	2: Mild active inflammation
Drews et al. histological score [32]	0: No inflammation	1: Chronic non-active inflammation	
Neye et al. endoscopic score [31]	0: No lesions	1: Aphtes	
Bozkurt et al. histological score (subjective) [30]	0	1	
Biancone et al. histological score (subjective) [34]	0	1	
Sciarretta et al. histological score (subjective) [33]	0	1	

Study	Moderate	Severe
Mao et al. histological score (Rutgeerts score) [17]	i3: Diffuse aphthous ileitis with diffusely inflamed mucosa	i4: Diffuse inflammation with already large ulcers, nodules, and/or narrowing
Mohamed et al. histological score (subjective) [16]	Moderate	Severe
Kolkman et al histological score [18]		

Table 8 (continued)

Study	Moderate	Severe
Schill et al. surgical score (based on Montreal classification) [29]	2: Moderate infiltration of polymorphonuclear leukocytes, some ulceration present B2: Stricturing	3: Severely ulcerated with massive infiltration of polymorphonuclear leukocytes B3: Penetrating
Gallego et al. endoscopic score (SES-CD) [a] [28]	≥7 points: Moderate/severe disease	
Koilakou et al. histological score (Rutgeerts score) [18]	i3: Diffuse aphthous ileitis with diffusely inflamed mucosa	i4: Diffuse inflammation with already large ulcers, nodules, and/or narrowing
Horsthuis et al. endoscopic score (subjective) [26]	Moderate disease	Severe disease
Girometti et al. histological score (subjective) [25]	Moderate-to-severe disease	
Horsthuis et al. endoscopic score (subjective) [24]	Moderate disease	Severe disease
Florie et al. endoscopic score (subjective) [21]	Moderate disease	Severe disease
Shoenut et al. endoscopic score [20]	Moderate: Marked edema, linear or patchy ulceration	Severe: Coalescing ulceration, exudative colitis
Shoenut et al. histological score (subjective) [19]	Moderate disease	Severe disease
Schreyer et al. endoscopic score [22]	2: Presence of large ulcerous lesions, nodules, and/or narrowing	
Schreyer et al. endoscopic score [23]	2: Presence of spontaneous bleeding, and large ulcerous lesions, nodules, and/or narrowing	
Drews et al. histological score [32]	3: Moderate active inflammation	4: Severe active inflammation
Neye et al. endoscopic score [31]	2: Aphtes and ulcers<50 %	3: Aphtes and ulcers>50 %
Bozkurt et al. histological score (subjective) [30]	2	
Biancone et al. histological score (subjective) [34]	2	3
Sciarretta et al. histological score (subjective) [33]	2	

[a] 0–3 points are given for the following features: size of ulcers (0: none, 1: aphthous ulcers (0.1–0.5 cm), 2: large ulcers (0.5–2 cm), 3: very large ulcers (>2 cm)), ulcerated surface (0: none, 1: <10 %, 2: 10–30 %, 3: >30 %), affected surface (0: none, 1: <50 %, 2: 50–75 %, 3: >75 %) and presence of narrowing (0: none, 1: single, can be passed, 2: multiple, can be passed, 3: cannot be passed)

Table 9 Original imaging criteria and categorization for this study

Study	None	Mild	Severe
Mao et al. CT score [17]	0: No findings	1: Minor mucosal irregularities with slight wall thickening and mural contrast enhancement	
Mohamed et al. CT score [16]	-	Mild: Mucosal hyperenhancement	
Kolkman et al CT score [18]	0: No thickening of the bowel wall, normal mesentery	1: Thickened bowel wall, homogenous aspect, no enhancement with intravenous contrast, no double-halo sign	
Kolkman et al scintigraphic score [18]	0: No activity	1: Uptake less than bone marrow	2: Uptake equal to bone marrow
Schill et al. MRI score (based on Montreal classification) [29]	-	B1: Non-stricturing and non-penetrating	
Gallego et al. MRI score[a] [28]	0–1 points: No disease	2–6 points: Mild disease	
Koilakou et al. MRI score [18]	0:	1:	
Horsthuis et al. MRI score (subjective) [26]	No disease	Mild disease	
Girometti et al. MRI score[b] [25]	0–1 points: No disease	2–6 points: Mild disease	
Horsthuis et al. MRI score (subjective) [24]	No disease	Mild disease	
Florie et al. MRI score (subjective) [21]	No disease	Mild disease	
Shoenut et al. MRI score [20]	-	Mild: ≤70 % contrast-enhancement in the most diseased segment (by wall thickness and length)	
Shoenut et al. MRI score [19]	-	Mild: Length of diseased segment<5 cm, bowel wall thickness<5 mm, contrast-enhancement<50 %	
Schreyer et al. MRI score [22]	0: No criteria	1: One of the following criteria: bowel wall thickening, bowel stenosis, increased contrast media uptake, enlarged local lymph nodes and local injection for inflammation assessment	
Schreyer et al. MRI score [23]	0: No criteria	1: One of the following criteria: bowel wall thickening with contrast enhancement enlarged local lymph nodes and mesenteric injection	
Drews et al. US score [32]	0: Bowel wall thickness 3–4 mm with preserved five-layer structure, no increased vascularity	1: Bowel wall thickness>4 mm, no increased vascularity	2: Grade 1 plus short stretches of increased vascularity
Neye et al. US score [31]	1: 0 vessels/cm² and bowel wall thickness<5 mm	2: 0 vessels/cm²+bowel wall thickness>5 mm or 1-2 vessels/cm²+bowel wall thickness<5 mm	
Bozkurt et al. US score [30]	0: Normal bowel wall with an echo-poor layer of ≤4 mm with a smooth boundary. Only the ventral wall visualizable to gaseous distention	1: Bowel wall thickness>4 mm with individual layers visible	
Biancone et al. scintigraphic score [34]	0: No labeling	1: Less than bone marrow	
Sciarretta et al. histological score (subjective) [33]	0: No uptake	1: Less than bone marrow	

Study	Severe	
Mao et al. CT score [17]	2: Mucosal hyperdensity with distinct bowel wall thickening, no stenosis, or stenosis without prestenotic dilatation	3: Major mucosal abnormalities, distinct bowel wall thickening with target sign and extravisceral signs such as perienteric stranding, comb sign, fibrofatty proliferation, stenosis with prestenotic dilatation and/or the presence of complications
Mohamed et al. CT score [16]	Moderate: Abnormal mucosal enhancement and wall thickening (>3 mm)	Severe: Abnormal mucosal enhancement, wall thickening (>3 mm) and one or more extra-enteric manifestations

Table 9 (continued)

Study	Moderate	Severe
Kolkman et al CT score [18]	2: Thickened bowel wall, enhancement with intravenous contrast or double-halo sign, ulceration, or mesenteric fibrofatty proliferation	3: Thickened bowel wall, enhancement with intravenous contrast, ulceration, and mesenteric fibrovascular strands (edema of the mesenteric fat, engorged vasa recta, lymphadenopathy, fistula, abscess)
Kolkman et al scintigraphic score [18]	3: Uptake higher than bone marrow, but less than liver	4: Uptake equal or higher than liver
Schill et al. MRI score (based on Montreal classification) [29]	B2: Stricturing	B3: Penetrating
Gallego et al. MRI score[a] [28]	≥7 points: Moderate/severe disease	3:
Koilakou et al. MRI score [18]	2:	2:
Horsthuis et al. MRI score (subjective) [26]	Moderate disease	Severe disease
Girometti et al. MRI score[b] [25]	≥7 points: Moderate/severe disease	
Horsthuis et al. MRI score (subjective) [24]	Moderate disease	Severe disease
Florie et al. MRI score (subjective) [21]	Moderate disease	Severe disease
Shoenut et al. MRI score [20]	Moderate: 71–119 % contrast-enhancement in the most diseased segment (by wall thickness and length)	Severe: ≥120 % contrast-enhancement in the most diseased segment (by wall thickness and length)
Shoenut et al. MRI score [19]	Moderate: Length of diseased segment>5 cm, bowel wall thickness 0.5-1 cm, contrast-enhancement<100 %	Severe: Length of diseased segment>5 cm, bowel wall thickness>1 cm, contrast-enhancement>100 %
Schreyer et al. MRI score [22]	2: Two or more of the following criteria: bowel wall thickening, bowel stenosis, increased contrast media uptake, enlarged local lymph nodes and local injection for inflammation assessment	
Schreyer et al. MRI score [23]	2: Two or more of the following criteria: bowel wall thickening with contrast enhancement enlarged local lymph nodes and mesenteric injection	
Drews et al. US score [32]	3: Grade 2 plus longer stretches of increased vascularity	4: Grade 3 plus vascularity extending into surrounding mesentery
Neye et al. US score [31]	3: 1–2 vessels/cm^2+bowel wall thickness>5 mm or>2 vessels/cm^2+bowel wall thickness<5 mm	4: > 2 vessels/cm^2+bowel wall thickness>5 mm
Bozkurt et al. US score [30]	2: Bowel wall thickness with poorly defined individual layers and decreased echogenicity	
Biancone et al. scintigraphic score [34]	2: More than bone marrow, less than liver	3: Equal or more than liver
Sciarretta et al. histological score (subjective) [33]	2: More than bone marrow, less than liver	3: Equal or more than liver

[a] 0–2 points are given for the following features: bowel wall thickness (0: < 3 mm, 1: 3–4 mm, 2: > 4 mm), relative enhancement (0: <70 %, 1: 70–100 %, 2: >100 %), motility (0: normal, 1: reduced, 2: absent), percentage stenosis (0: ≤60 %, 1: >60 %), bowel wall edema (0: absent, 1: present), mucosal abnormalities (0: absent, 1: present), lymph nodes (0: absent, 1: present), fistulae or sinus tracts (0: absent, 1: present), inflammatory masses (0: absent, 1: present)

[b] 0–2 points are given for the following features: bowel wall thickness (0: < 3 mm, 1: 3–4 mm, 2: > 4 mm), wall–contrast enhancement (0: <70 %, 1: 70–100 %, 2: >100 %), percentage stenosis (0: <50 %, 1: 50–80 %, 2: >80 %), mucosal abnormalities (0: absent, 1: present), layered wall enhancement (0: absent, 1: present), peristalsis (0: present, 1: absent), distensibility (0: present, 1: absent), mesenteric involvement (0: absent, 1: present), pathologic lymph nodes (n>3) (0: absent, 1: present), fistulae or sinus tracts (0: absent, 1: present), inflammatory masses (0: absent, 1: present)

Publication bias

Linear regression analysis on MRI per-patient data showed a regression coefficient of 0.4 (95 % CI: −0.9 to 0.9), with no significant relationship between accurate grading and $1/\sqrt{ESS}$ ($P=0.09$). Data in other groups were deemed insufficient for performing linear regression analyses.

Data analysis

Results from our data analysis are presented in Table 10. Three-by-three contingency tables for each study can be found in the supplementary materials (Appendix E3).

Per-patient Data was provided on a per-patient basis in 13 studies (evaluating CT in 2, MRI in 9, US in 1 and scintigraphy in 1) (Fig. 4). I^2 values for overall grading accuracy for groups with more than one dataset were as follows: 67.7 % (95 % CI: 42.6–81.8 %) for CT, and 73.9 % (95 % CI: 56.2−84.4 %) for MRI.

CT and MRI data were pooled for each modality ($I^2<$ 75 %). US and scintigraphy were not pooled, as only one dataset was available for each modality. CT, MRI, US and scintigraphy showed accurate grading estimates of 86 % (95 % CI: 75–93 %), 84 % (95 % CI: 67–93 %), 44 % (95 % CI: 28–61 %) and 40 % (95 % CI: 16–70 %), respectively. CT and MRI showed similar overall grading accuracy ($P=0.8$), both higher than US ($P=0.0001$ and $P=0.001$, respectively) and scintigraphy ($P=0.003$ and $P=0.01$, respectively). CT and MRI showed similar over-grading ($P=0.8$) and under-grading ($P=0.5$). Both showed less under-grading than US ($P=0.002$ and $P=0.003$, respectively) and scintigraphy ($P=0.0005$ and $P=0.001$, respectively).

Per-segment Data were provided on a per-segment basis in seven articles, of which one evaluated both CT and scintigraphy, two evaluated MRI, two evaluated US, and two evaluated

scintigraphy, respectively (Fig. 4). I^2 values were 86.3 % (95 % CI: 66.4–94.4 %) for MRI, 91.5 % (95 % CI: 79.1–96.6 %) for US, and 0 % for scintigraphy. MRI and US data were not pooled, as data were too heterogeneous ($I^2 \geq 75$ %). Data on CT were also not pooled, as only one dataset was available. The overall grading accuracy was 87 % (95 % CI: 77–93 %) for CT and 86 % (95 % CI: 80–91 %) for scintigraphy. CT and scintigraphy showed similar overall grading accuracy ($P=0.8$), over-grading ($P=0.2$) and under-grading ($P=0.5$). Accuracy for MRI and US ranged from 67 to 82 % and 56 to 75 %, respectively.
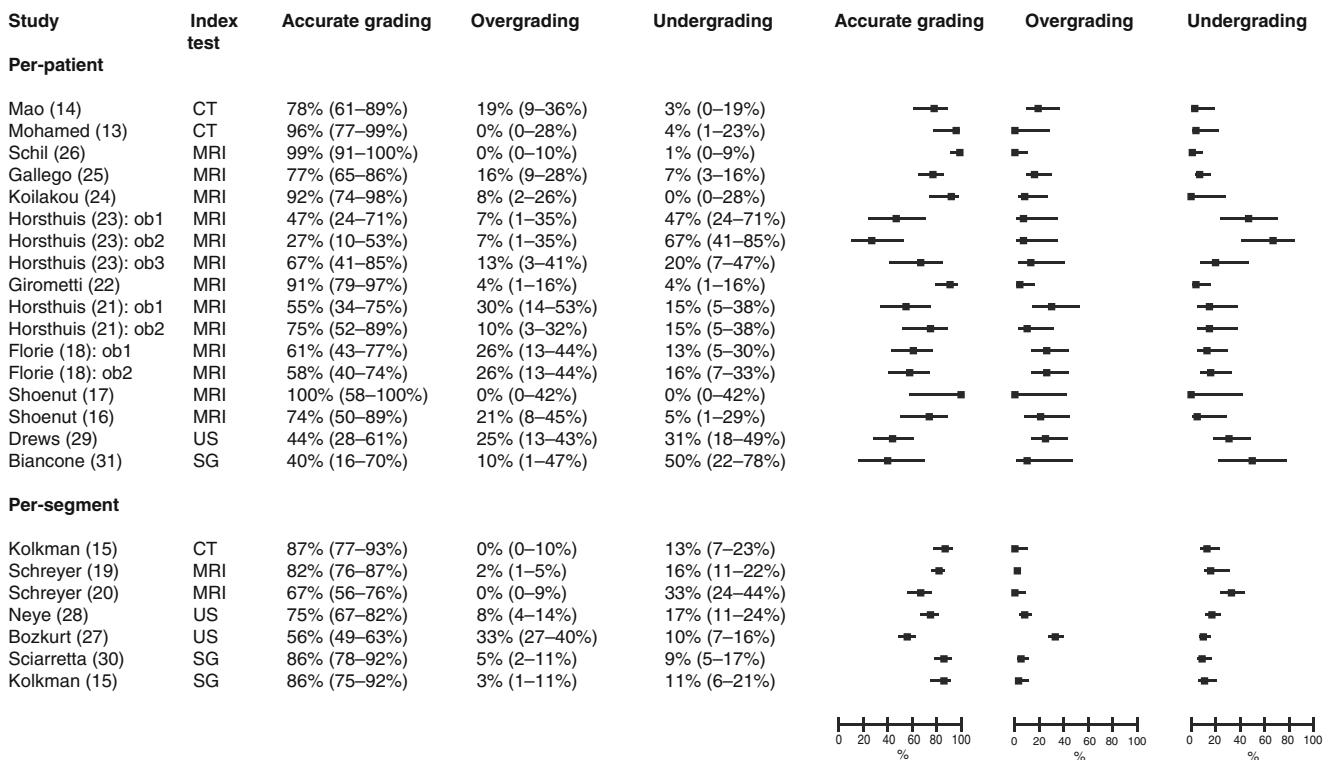
Discussion

In this study, we have shown that MRI and CT are highly accurate for grading Crohn's disease activity. These findings are important, as cross-sectional imaging plays an increasing role in the assessment of Crohn's disease activity, and there has been ongoing debate regarding the modality that should be the preferred choice [35–37]. Several studies have compared two or more modalities in the same patient group [38–41], but they have had relatively small sample sizes or only evaluated the terminal ileum.

CT and MRI showed similar accuracy in grading Crohn's disease activity (86 % and 84 % on a per-patient basis, respectively), and no significant differences in accuracy were seen between these two modalities. Data on over- and under-grading showed similar results for CT and MRI, further strengthening our conclusion of their comparability. Scintigraphy showed high accuracy of 86 % and 86 % for the studies using per-segment data, while accuracy of 40 % was reported in per-patient data. However, per-patient data for scintigraphy was reported in only one study, and with a small sample size ($n=10$) [34]. Furthermore, scintigraphy had the least number of included patients ($n=58$) in our meta-analysis. US showed low accuracy of 44 % in the per-patient data and 75 % and

Table 10 Comparison table with results for imaging tests from the 3×3 data analysis and corresponding P values

	Accurate grading	Over-grading	Under-grading
Per-patient (13 datasets)			
CT (n=2) vs MRI (n=9)	0.86 vs 0.84 (P=0.8)	0.10 vs 0.09 (P=0.8)	0.03 vs 0.06 (P=0.5)
CT (n=2) vs US (n=1)	0.86 vs 0.44 (P=0.0001)	0.10 vs 0.25 (P=0.07)	0.03 vs 0.31 (P=0.002)
CT (n=2) vs SG (n=1)	0.86 vs 0.40 (P=0.003)	0.10 vs 0.10 (P=1.0)	0.03 vs 0.50 (P=0.0005)
MRI (n=9) vs US (n=1)	0.84 vs 0.44 (P=0.001)	0.09 vs 0.25 (P=0.03)	0.06 vs 0.31 (P=0.003)
MRI (n=9) vs SG (n=1)	0.84 vs 0.40 (P=0.01)	0.09 vs 0.10 (P=0.9)	0.06 vs 0.50 (P=0.001)
US (n=1) vs SG (n=1)	0.44 vs 0.40 (P=0.8)	0.25 vs 0.10 (P=0.3)	0.31 vs 0.50 (P=0.3)
Per-segment (3 datasets) [a]			
CT (n=1) vs SG (n=2)	0.87 vs 0.86 (P=0.8)	0.00 vs 0.04 (P=0.2)	0.13 vs 0.10 (P=0.5)

[a] Data on MRI and US were not pooled and included in the comparison, as the data were too heterogeneous ($I^2 > 75$ %)

Study	Index test	Accurate grading	Overgrading	Undergrading	Accurate grading	Overgrading	Undergrading
Per-patient							
Mao (14)	CT	78% (61–89%)	19% (9–36%)	3% (0–19%)			
Mohamed (13)	CT	96% (77–99%)	0% (0–28%)	4% (1–23%)			
Schil (26)	MRI	99% (91–100%)	0% (0–10%)	1% (0–9%)			
Gallego (25)	MRI	77% (65–86%)	16% (9–28%)	7% (3–16%)			
Koilakou (24)	MRI	92% (74–98%)	8% (2–26%)	0% (0–28%)			
Horsthuis (23): ob1	MRI	47% (24–71%)	7% (1–35%)	47% (24–71%)			
Horsthuis (23): ob2	MRI	27% (10–53%)	7% (1–35%)	67% (41–85%)			
Horsthuis (23): ob3	MRI	67% (41–85%)	13% (3–41%)	20% (7–47%)			
Girometti (22)	MRI	91% (79–97%)	4% (1–16%)	4% (1–16%)			
Horsthuis (21): ob1	MRI	55% (34–75%)	30% (14–53%)	15% (5–38%)			
Horsthuis (21): ob2	MRI	75% (52–89%)	10% (3–32%)	15% (5–38%)			
Florie (18): ob1	MRI	61% (43–77%)	26% (13–44%)	13% (5–30%)			
Florie (18): ob2	MRI	58% (40–74%)	26% (13–44%)	16% (7–33%)			
Shoenut (17)	MRI	100% (58–100%)	0% (0–42%)	0% (0–42%)			
Shoenut (16)	MRI	74% (50–89%)	21% (8–45%)	5% (1–29%)			
Drews (29)	US	44% (28–61%)	25% (13–43%)	31% (18–49%)			
Biancone (31)	SG	40% (16–70%)	10% (1–47%)	50% (22–78%)			
Per-segment							
Kolkman (15)	CT	87% (77–93%)	0% (0–10%)	13% (7–23%)			
Schreyer (19)	MRI	82% (76–87%)	2% (1–5%)	16% (11–22%)			
Schreyer (20)	MRI	67% (56–76%)	0% (0–9%)	33% (24–44%)			
Neye (28)	US	75% (67–82%)	8% (4–14%)	17% (11–24%)			
Bozkurt (27)	US	56% (49–63%)	33% (27–40%)	10% (7–16%)			
Sciarretta (30)	SG	86% (78–92%)	5% (2–11%)	9% (5–17%)			
Kolkman (15)	SG	86% (75–92%)	3% (1–11%)	11% (6–21%)			

CT = Computed Tomography, MRI = Magnetic Resonance Imaging, US = Ultrasound, SG = Scintigraphy, ob = observer

Fig. 4 Accurate grading, over- and under-grading per study on a per-patient and per-segment basis

56 % for studies in the per-segment data. However, a relatively small number of patients ($n=86$) were included. In addition, no eligible studies evaluated luminal or intravenous contrast medium for US. The use of intravenous contrast appears to be a particularly promising technique, and may increase the accuracy of US. However, no robust reference standard or appropriate grading scale were used in these studies. We considered the possibility of performing subgroup and covariate analyses on the differences in technique, imaging criteria, reference methods and methodological criteria, but the results of these analyses would not be meaningful given the limited amount of available data. We examined MRI imaging features in three studies with the highest accuracy values. The following MRI features were used by at least two of these studies: bowel wall thickness, T1 enhancement and pattern, T2 mural signal intensity, mucosal abnormalities, presence of inflammatory mass, stenosis (with pre-stenotic dilatation), lymph nodes, abscesses, and fistulas [25, 27, 29].

The observed heterogeneity of the grading criteria for the index and reference tests in the studies that we included, our adjustment to construct 3×3 tables, and the differences in available data between imaging modalities were the major limitations of this meta-analysis. Although the grading criteria for index and reference tests differed by study, and different imaging features were used, the studies included showed considerable overlap in the use of imaging features and grading criteria. No generally accepted scoring systems exist for

imaging of Crohn's disease. To construct 3×3 tables from original 4×4 data, we merged moderate and severe disease into one group. Our decision to merge these grades was based on five articles [22, 23, 25, 28, 30] that had originally used 3×3 tables; two of these studies explicitly stated that their highest grade represented moderate and severe disease combined [25, 28]. The remaining three studies [22, 23, 30] used similar grading criteria. Another limitation was the heterogeneity of grading results, which we examined using I^2 statistics. Following those results, some of the datasets could not be pooled. In our conclusions, we took into account the greater availability of data for MRI compared to CT, US and scintigraphy. Furthermore, US and scintigraphy studies showed varying results, hampering our ability to arrive at a firm conclusion. There was only one head-to-head comparison study, which compared CT and scintigraphy in 17 patients [18].

We selected three reference standards for this meta-analysis [35]. Intraoperative findings served as the gold standard for assessing Crohn's disease. We also included endoscopy and endoscopic biopsies as reference standards, although they are not ideal, as they are incapable of assessing proximal ileum, jejunum and extraluminal disease, which could have led to incorrect estimation of disease activity. On the other hand, surgery is performed only in select patients, whereas endoscopy is applied across a wider spectrum. For our analysis, we gave precedence to results from biopsies over endoscopic results, but we recognize that this was a controversial choice, as

there is no widespread consensus on which is the better reference standard. The number of studies included could have been increased if VCE and/or double-balloon enteroscopy (DBE) were also used as a reference standard. We chose not to include these studies because interpretation of VCE and DBE has not yet been standardized, and so this would further increase heterogeneity in our study. A growing number of studies are using correlative statistics to examine quantitative scoring systems [42]. Because we used an ordinal outcome measure, we could not include these studies. Nevertheless, a meta-analysis focused on this type of data would be very useful. Finally, only patients with suspected IBD or known Crohn's disease were included, possibly introducing observer bias, leading to over-grading of disease activity.

Assessment of study quality using the QUADAS tool showed overall moderate quality of the studies included in this meta-analysis. The domains of reference test and patient flow showed the highest risk of bias, while patient selection and index test domains showed the lowest. Concern about the applicability of patient selection and index and reference tests was generally low.

Recently, Vermeire et al. stated that MR enterography had become the reference standard for assessing small and large bowel disease activity [43]. Based on our results, we can agree with this statement. Considering the radiation exposure from CT, it is not appropriate for repeated examinations, even with present-day reduced ionizing radiation exposure per examination, although it still has an important role in the acute setting [44]. Compared to endoscopy, MRI is non-invasive and able to investigate trans- and extramural disease, making it possible to evaluate both the small bowel and colon in one examination. Steps are being taken to come to a more uniform evaluation of MRI in Crohn's disease, which may improve accuracy [42, 45]. Furthermore, the versatility of MRI may be advantageous with new sequences being studied.

In conclusion, CT and MRI can both be used to grade disease activity in Crohn's disease, while no conclusions can be made on US and scintigraphy due to the limited and inconsistent data.

References

1. Horsthuis K, Bipat S, Bennink RJ, Stoker J (2008) Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: meta-analysis of prospective studies. Radiology 247:64–79

2. Panes J, Bouzas R, Chaparro M et al (2011) Systematic review: the use of ultrasonography, computed tomography and magnetic resonance imaging for the diagnosis, assessment of activity and abdominal complications of Crohn's disease. Aliment Pharmacol Ther 34:125–145

3. Travis SP, Stange EF, Lemann M et al (2006) European evidence based consensus on the diagnosis and management of Crohn's disease: current management. Gut 55:i16–i35

4. Hommes DW, van Deventer SJ (2004) Endoscopy in inflammatory bowel diseases. Gastroenterology 126:1561–1573

5. Fletcher JG, Fidler JL, Bruining DH, Huprich JE (2011) New concepts in intestinal imaging for inflammatory bowel diseases. Gastroenterology 140:1795–1806

6. Rimola J, Ordas I, Rodriguez S, Ricart E, Panes J (2012) Imaging indexes of activity and severity for Crohn's disease: current status and future trends. Abdom Imaging 37:958–966

7. Horsthuis K, Bipat S, Stokkers PC, Stoker J (2009) Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. Eur Radiol 19:1450–1460

8. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6, e1000097

9. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536

10. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25

11. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893

12. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560

13. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723

14. Bipat S, Zwinderman AH, Bossuyt PM, Stoker J (2007) Multivariate random-effects approach: for meta-analysis of cancer staging studies. Acad Radiol 14:974–984

15. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56:1129–1135

16. Mohamed AM, Amin SK, El-Shinnawy MA, Elfouly A, Baki AH (2012) Role of CT enterography in assessment of Crohn's disease activity: correlation with histopathologic diagnosis. Egypt J Radiol Nucl Med 43:353–359

17. Mao R, Gao X, Zhu ZH et al (2013) CT enterography in evaluating postoperative recurrence of Crohn's disease after ileocolic resection: complementary role to endoscopy. Inflamm Bowel Dis 19: 977–982

18. Kolkman JJ, Falke TH, Roos JC et al (1996) Computed tomography and granulocyte scintigraphy in active inflammatory bowel disease. Comparison with endoscopy and operative findings. Dig Dis Sci 41:641–650

19. Shoenut JP, Semelka RC, Silverman R, Yaffe CS, Micflikier AB (1993) Magnetic resonance imaging in inflammatory bowel disease. J Clin Gastroenterol 17:73–78

20. Shoenut JP, Semelka RC, Magro CM, Silverman R, Yaffe CS, Micflikier AB (1994) Comparison of magnetic resonance imaging and endoscopy in distinguishing the type and severity of inflammatory bowel disease. J Clin Gastroenterol 19:31–35

21. Florie J, Horsthuis K, Hommes DW et al (2005) Magnetic resonance imaging compared with ileocolonoscopy in evaluating disease severity in Crohn's disease. Clin Gastroenterol Hepatol 3: 1221–1228

22. Schreyer AG, Golder S, Scheibl K et al (2005) Dark lumen magnetic resonance enteroclysis in combination with MRI colonography for whole bowel assessment in patients with Crohn's disease: first clinical experience. Inflamm Bowel Dis 11: 388–394

23. Schreyer AG, Rath HC, Kikinis R et al (2005) Comparison of magnetic resonance imaging colonography with conventional colonoscopy for the assessment of intestinal inflammation in patients with inflammatory bowel disease: a feasibility study. Gut 54: 250–256

24. van Gemert-Horsthuis K, Florie J, Hommes DW et al (2006) Feasibility of evaluating Crohn's disease activity at 3.0 Tesla. J Magn Reson Imaging 24:340–348

25. Girometti R, Zuiani C, Toso F et al (2008) MRI scoring system including dynamic motility evaluation in assessing the activity of Crohn's disease of the terminal ileum. Acad Radiol 15:153–164

26. Horsthuis K, de Ridder L, Smets AM et al (2010) Magnetic resonance enterography for suspected inflammatory bowel disease in a pediatric population. J Pediatr Gastroenterol Nutr 51:603–609

27. Koilakou S, Sailer J, Peloschek P et al (2010) Endoscopy and MR enteroclysis: equivalent tools in predicting clinical recurrence in patients with Crohn's disease after ileocolic resection. Inflamm Bowel Dis 16:198–203

28. Gallego JC, Echarri AI, Porta A, Ollero V (2011) Ileal Crohn's disease: MRI with endoscopic correlation. Eur J Radiol 80:e8–e12

29. Schill G, Iesalnieks I, Haimerl M et al (2013) Assessment of disease behavior in patients with Crohn's disease by MR enterography. Inflamm Bowel Dis 19:983–990

30. Bozkurt T, Rommel T, Stabenow-Lohbauer U, Langer M, Schmiegelow P, Lux G (1996) Sonographic bowel wall morphology correlates with clinical and endoscopic activity in Crohn's disease and ulcerative colitis. Eur J Ultrasound 4:27–33

31. Neye H, Voderholzer W, Rickes S, Weber J, Wermke W, Lochs H (2004) Evaluation of criteria for the activity of Crohn's disease by power Doppler sonography. Dig Dis 22:67–72

32. Drews BH, Barth TF, Hanle MM et al (2009) Comparison of sonographically measured bowel wall vascularity, histology, and disease activity in Crohn's disease. Eur Radiol 19:1379–1386

33. Sciarretta G, Furno A, Mazzoni M, Basile C, Malaguti P (1993) Technetium-99 m hexamethyl propylene amine oxime granulocyte scintigraphy in Crohn's disease: diagnostic and clinical relevance. Gut 34:1364–1369

34. Biancone L, Scopinaro F, Ierardi M et al (1997) 99mTc-HMPAO granulocyte scintigraphy in the early detection of postoperative asymptomatic recurrence in Crohn's disease. Dig Dis Sci 42: 1549–1556

35. Panes J, Bouhnik Y, Reinisch W et al (2013) Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. J Crohns Colitis 7: 556–585

36. Masselli G, Gualdi G (2013) CT and MR enterography in evaluating small bowel diseases: when to use which modality? Abdom Imaging 38:249–259

37. Grand DJ, Harris A, Loftus EV Jr (2012) Imaging for luminal disease and complications: CT enterography, MR enterography, small-bowel follow-through, and ultrasound. Gastroenterol Clin N Am 41:497–512

38. Siddiki HA, Fidler JL, Fletcher JG et al (2009) Prospective comparison of state-of-the-art MR enterography and CT enterography in small-bowel Crohn's disease. AJR Am J Roentgenol 193:113–121

39. Lee SS, Kim AY, Yang SK et al (2009) Crohn disease of the small bowel: comparison of CT enterography, MR enterography, and small-bowel follow-through as diagnostic techniques. Radiology 251:751–761

40. Jensen MD, Nathan T, Rafaelsen SR, Kjeldsen J (2011) Diagnostic accuracy of capsule endoscopy for small bowel Crohn's disease is superior to that of MR enterography or CT enterography. Clin Gastroenterol Hepatol 9:124–129

41. Jensen MD, Kjeldsen J, Rafaelsen SR, Nathan T (2011) Diagnostic accuracies of MR enterography and CT enterography in symptomatic Crohn's disease. Scand J Gastroenterol 46:1449–1457

42. Rimola J, Rodriguez S, Garcia-Bosch O et al (2009) Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. Gut 58:1113–1120

43. Vermeire S, Ferrante M, Rutgeerts P (2013) Recent advances: personalised use of current Crohn's disease therapeutic options. Gut 62:1511–1515

44. Peloquin JM, Pardi DS, Sandborn WJ et al (2008) Diagnostic ionizing radiation exposure in a population-based cohort of patients with inflammatory bowel disease. Am J Gastroenterol 103: 2015–2022

45. Steward MJ, Punwani S, Proctor I et al (2012) Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. Eur J Radiol 81:2080–2088