Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review

Horsthuis, K.; Bipat, S.; Stokkers, P.C.F.; Stoker, J.

DOI
10.1007/s00330-008-1287-0

Publication date
2009

Document Version
Final published version

Published in
European Radiology

Eur Radiol (2009) 19: 1450–1460
DOI 10.1007/s00330-008-1287-0

Karin Horsthuis
Shandra Bipat
Pieter C. F. Stokkers
Jaap Stoker

Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review

Received: 7 July 2008
Accepted: 27 November 2008
Published online: 3 February 2009

K. Horsthuis · S. Bipat · J. Stoker
Department of Radiology,
Academic Medical Center,
Amsterdam, The Netherlands

P. C. F. Stokkers
Department of Gastroenterology,
Academic Medical Center,
Amsterdam, The Netherlands

K. Horsthuis (✉)
Academic Medical Center,
Meibergdreef 9,
1105, AZ,
Amsterdam, The Netherlands
e-mail: k.horsthuis@amc.uva.nl
Tel.: +31-20-5669111
Fax: +31-20-5669119

Abstract To systematically review the evidence on the accuracy of MRI for grading disease activity in Crohn's disease (CD). The MEDLINE, EMBASE, CINAHL and Cochrane databases were searched for studies on the accuracy of MRI in grading CD compared to a predefined reference standard. Two independent observers scored all relevant data. Three disease stages were defined: remission, mild and frank disease. The accuracy rates of MRI per disease stage were calculated by means of a random-effects model. Seven studies were included from a search resulting in 253 articles. In total 140 patients (16 patients in remission, 29 with mild disease and 95 with frank disease) were used for data analysis. MRI correctly graded 91% (95% CI: 84–96%) of patients with frank disease, 62% (95% CI: 44–79) of patients with mild disease and 62% (95% CI: 38–84) of patients in remission. MRI more often overstaged than understaged disease activity; MRI overstaged disease activity in 38% of patients in remission, mostly as mild disease. Overstaging of mild disease was observed in 21%, understaging in 17%. MRI correctly grades disease activity in a large proportion of patients with frank disease. For patients in remission or with mild disease, MRI correctly stages disease activity in many patients (62%).

Keywords Magnetic resonance imaging · Crohn's disease · Disease activity · Systematic review

Introduction

Before the initiation of medical or surgical therapy for symptomatic Crohn's disease (CD), it is crucial to assess whether inflammatory activity is present, because even though the CD may be in remission, symptoms of coexisting irritable bowel syndrome (IBS) may mimic active disease. It also is important to distinguish bowel obstruction due to inflammation from stenosis due to residual fibrotic stenosis as these respectively warrant medical therapy or surgical therapy. Furthermore, if inflammatory activity is present, it is important to distinguish between mild, moderate or severe disease as medical management differs among the disease stages [1, 2].

The reference standard for diagnosing active CD and staging disease activity is endoscopy [3]. However, with standard endoscopic techniques only part of the bowel can be visualized, while the low patient acceptance forms another drawback of this technique.

Many studies have advocated the use of computed tomography (CT) for abdominal evaluation in patients with CD, as it is an accurate and patient-friendly technique [4–8]. However, during an abdominal CT examination patients are exposed to considerable radiation doses (mean cumulative effective dose is 36.1 mSv; however, more than 75 mSv can be obtained) [9].

As assessment of disease activity is often necessary repeatedly, the excess lifetime cancer mortality risk

attributable to radiation exposure will increase when abdominal CT is used for CD evaluation. It has been estimated that about 1.5 to 2.0% of all cancers in the US may be attributable to the radiation from CT studies [10]. In contrast, magnetic resonance imaging (MRI) is an investigation that does not require the use of ionizing radiation. As it also is a non-invasive technique, MRI is increasingly used for abdominal evaluation in patients with CD [11–13]. However, while MRI has been shown to be accurate in diagnosing active CD [14, 15], the accuracy of MRI in staging disease activity is not so clear yet. As MRI is inferior to colonoscopy in the detection of subtle mucosal detail, MRI might provide false-negative results in patients with mild, superficial CD. This hypothesis is supported by findings from several studies in which false-negative MRI results were seen in patients with active, mostly mild CD [16–19]. However, in other studies disease activity was overestimated on MRI [20–22].

Thus, the purpose of our study was to systematically review the accuracy of MRI in staging disease activity in CD by performing a meta-analysis.

Materials and methods

Search strategy and study eligibility

A computer-assisted search was performed of the MED LINE, EMBASE, CINAHL and Cochrane databases to identify papers reporting the accuracy of MRI in staging CD activity. In MEDLINE and EMBASE, we used "Crohn disease (MeSH)" and "Magnetic resonance imaging (MeSH)" as search terms. For searching the CINAHL and Cochrane databases, we used "Crohn disease" and "Magnetic resonance imaging" as free text words. The search period was restricted from 1990 through April 2007. No age limits or language restrictions were applied.

Titles and/or abstracts of all retrieved papers were checked by one observer (KH) to determine eligibility for inclusion. Reference lists of review articles and eligible studies were checked manually to identify other relevant papers. Hand searching of major journals was not performed. Only data that were presented as full-text articles were eligible for inclusion. As field strength of most MRI systems currently used in clinical practice is ≥1.0 T, we decided to exclude papers in which MRI field strength was ≤0.5 T. All eligible articles were retrieved as full-text articles.

Study selection

Two reviewers (KH and SB) independently checked all retrieved articles to check whether they satisfied the following criteria: (1) they provided data on disease activity of CD; (2) MRI was used to evaluate CD; (3) findings at histopathology, colonoscopy and/or intra-operative findings were used as the reference standard; (4) positive criteria were defined for MRI (i.e., criteria described to stage disease activity); (5) data were available to fill out cross-tabs (for calculation of agreement in staging disease).

If all criteria were met, the article was included in the study. Disagreement between the two reviewers regarding inclusion was resolved by consensus. The authors of the primary research were approached for additional information, if neccessary.

Study characteristics

Both reviewers independently assessed study characteristics of the included studies and extracted relevant data, described in detail below, by using a standardized form. No blinding of authors' information, authors' affiliation or journal title was performed. Inconsistencies in assessment of the included studies were resolved by consensus.

Patient characteristics

The following patient characteristics were recorded: (1) number of patients; (2) sex ratio distribution; (3) mean age (range); (4) part of the gastrointestinal tract examined.

Study quality assessment

To assess study quality characteristics, the QUADAS tool was used as a guideline. The QUADAS tool has been developed for reviewers to evaluate the quality of studies and especially studies of diagnostic accuracy [23, 24]. The following characteristics were assessed:

(1) Whether the spectrum of patients was representative of the patients who will receive MRI in practice;
(2) If selection criteria were clearly described;
(3) Whether the time period between the MRI and the reference standard was short enough to be reasonably sure that the condition did not change between the two tests;
(4) Whether all patients received verification using a reference standard;
(5) Whether the execution of the MRI was described in sufficient detail to permit its replication (we considered the MRI description as sufficient if information was provided about the following imaging features: magnetic field strength; type of coil used, bowel preparation used, and sequences used for evaluation; the use of intravenous and/or luminal contrast medium);
(6) Whether the execution of the reference standard was described in sufficient detail to permit its replication

(we considered the reference standard described as sufficient if the criteria used for diagnosing the different disease stages were defined);
(7) Whether the MRI results and the reference test results were evaluated independently;
(8) Whether interpretation of the MRI results was independent of clinical information.

Imaging features

The following imaging features were recorded for MRI, if available: (1) magnetic field strength; (2) coil used (body or surface); (3) bowel preparation and type of bowel preparation (bowel cleansing, fasting and/or diet, use of spasmolytic medication); (4) amount and type of intravenous and/or luminal contrast medium (enteroclysis, oral and/or rectal contrast medium) if administrated; (5) sequences used for disease evaluation.

Imaging criteria used for staging disease activity

For each study the imaging criteria that were used to stage CD on MRI (e.g., pathological bowel wall thickening, pathological bowel wall enhancement and stenosis) were noted.

Reference standard

The verification method used (surgery, histopathology and/or colonoscopy) was recorded for each study.

Data extraction

For each study, 3×3 (remission, mild, frank) or 4×4 (remission, mild, moderate, severe) contingency tables were extracted from the articles, depending on the way of reporting.

Data analysis

An overall analysis was performed for the 3×3 data. For this approach, 4×4 tables were reconstructed to 3×3 tables by grouping moderate and severe disease together as frank disease. For the 3×3 data, analysis was performed using a multivariate random-effects approach [25] performed by

Table 1 Patient characteristics of the included studies

Study	Patient spectrum	Selection criteria	Patients (n) *	Male:female ratio	Mean age (range)	GI-tract examined
Shoenut 1994	Suspected IBD	Consecutive patients with suspected IBD presenting for the first time with symptoms of IBD	20 (**12 CD**, 6 UC, 2 IC)	12:8	42.6 (20–70)	Colon and terminal ileum
Durno 2000	Known IBD	Children and adolescents undergoing colonoscopic evaluation for IBD	15 (**9 CD**, 4 UC, 1 IC)	NA	14.1 (7–17)	Colon and terminal ileum
Laghi 2003	Suspected CD	Consecutive children referred to the pediatric gastroenterology unit	75 (**26 CD**; 18 UC; 11 IC; 20 controls)	NA	Median 14 (12–17) for CD patients	Terminal ileum
Florie 2005	Known CD	Patients scheduled for colonoscopy because of clinical suspicion of relapsing CD	31 (**31 CD**)	22:9	36 (18–60)	Colon and small bowel
Schreyer Gut 2005	Highly suspected or known IBD	Consecutive patients scheduled for colonoscopy to assess disease activity or pathological changes of the colon	22 (**12 CD**; 10 UC)	11:11	Median 38 (19–71)	Colon and terminal ileum
Schreyer Inflamm Bowel Dis 2005	Known CD	Consecutive patients assigned to a routine MR enterography of the small bowel	30 (**30 CD**)	8:22	Median 29 (18–65)	Colon and small bowel
Van Gemert 2006	Known CD	Scheduled to undergo colonoscopy	20 (**20 CD**)	7:13	36 (22–58)	Colon and small bowel

*Numbers in bold indicate numbers of CD patients per study
IBD: Inflammatory bowel disease
CD: Crohn's Disease
UC: Ulcerative colitis
IC: Indeterminate colitis
NA: Not available

using a Bayesian algorithm [26] in the Winbugs program. Summary estimates were calculated. If studies reported data for multiple independent observers, we used the data leading to the lowest Aikaike information criterion (AIC) value to calculate summary estimates; a lower AIC value indicates a better fit of the data [27].

Analysis on 4×4 tables could not be performed due to the limited amount of data per stage. The results of the indivual studies are described.

Results

Search strategy and study selection

The search strategy resulted in 253 articles; 36 were found to be eligible after reading the abstract and were retrieved as full text for further analysis. Finally, seven papers [17, 18, 28–32] fulfilled all inclusion criteria and were used for data extraction and data analysis (Appendix 1). There was no disagreement regarding inclusion between the two reviewers.

Study characteristics

Inconsistencies in assessments of the included studies among reviewers were resolved by consensus. Of the 56 items scored (8 per study), for 6 items, inconsistencies existed.

Patient characteristics

In three of the seven included studies, only patients with CD were evaluated [17, 31, 32]. In the other four studies, both patients with CD and with ulcerative colitis were

Table 2 Study design characteristics

Study	Representative spectrum patients	Selection criteria	Time interval †	Verification ‡	Execution MRI §	Execution reference test ¶	Evaluation MRI II	Evaluation reference test #	Clinical information *	Reference standard
Shoenut	Yes	Yes	3 days	Complete	No	Yes	NA	NA	NA	Histopathology
Durno	Yes	Yes	2 days	Complete	No	No	Yes	NA	NA	Colonoscopy
Laghi	Yes	No	NA	Complete	Yes	Yes	Yes	NA	NA	Colonoscopy
Florie	Yes	Yes	2 weeks	Complete	Yes	Yes	Yes	NA	No	Colonoscopy
Schreyer Gut 2005	Yes	Yes	Same day	Complete	Yes	Yes	Yes	NA	NA	Colonoscopy intra-operative findings (+ histopathology)
Schreyer Inflamm Bowel Dis 2005	Yes	Yes	1 week (in 29/30 pts)	Complete	Yes	Yes	NA	NA	NA	Colonoscopy intra-operative findings (+ histopathology)
Van Gemert	Yes	Yes	Median 5 days (1–48 days)	Complete	Yes	Yes	Yes	NA	No	Colonoscopy

†Time interval between MRI and reference standard
NA, not described
‡Verification of the included patients by reference standard (complete or incomplete)
§Execution of MRI described sufficiently
MRI: magnetic field; bowel preparation; sequences; luminal/IV contrast
Yes: Sufficiently described
No: Not sufficiently described
¶Reference standard: criteria for staging disease activity defined or not
Yes: Mentioned in study
No: Not mentioned in study
II: Evaluation of MRI performed blinded from the reference standard
Yes: Mentioned in study
NA: Information concerning blinding not available
#Evaluation of the reference standard performed independently of MRI
NA: Information concerning blinding not available
*Clinical information available during interpretation of the imaging test
No: Clinical information not available during interpretation of the imaging examination
NA: Not described in the study

Table 3 Imaging features and criteria

Study	Magnetic field strength	Bowel preparation	Luminal contrast	Intravenous contrast	Coil	Criteria used for disease assessment
Shoenut 1994	1. 5 T	NA	NA	0.1 mmol/kg Gadolinium-DTPA (Magnevist)	NA	Percentage of contrast enhancement Wall thickening
Durno 2000	1.5 T	NA	NA	0.1 mmol/kg Gadolinium-DTPA (Magnevist)	NA	Percentage of contrast enhancement
Laghi 2003	1.5 T	Overnight fast	Oral: PEG solution (10 ml/kg bodyweight)	0.1 mmol/kg Gadolinium-DTPA (Magnevist)	Phased array body coil	Bowel wall thickening Parietal contrast enhancement
Florie 2005	1.5 T	Fasting 4 h Buscopan 20 mg iv	Oral: 1 l water	0.1 mmol/kg Gadolinium-DTPA (Magnevist)	NA	Bowel wall thickening Bowel wall enhancement Stenosis Target sign Cobblestoning Extraintestinal findings
Schreyer Gut 2005	1.5 T	KleanPrep Buscopan 40 mg iv	Rectal: 1.5 l (1.1–1.8 l) of a gadolinium-water mixture (5 mol/l Gadolinium-DTPA)	0.1 mmol/kg Gadolinium-DTPA (Magnevist)	Phased array body coil	Bowel wall thickening Contrast enhancement
Schreyer Inflamm Bowel Dis 2005	1.5 T	Fasting ≥12 h Buscopan 40 mg iv	Oral: 2 l tap water (25 g Mannitol+5 g carob seed per liter) Rectal: application of 0.9% NaCl 400–1,000 ml	0.2 mmol/kg Gadolinium-DTPA (Magnevist)	Phased array body coil	Bowel wall thickening Bowel stenosis Increased contrast media uptake Enlarged local lymph nodes Local injection
Van Gemert 2006	3.0 T	Fasting 4 h Buscopan 20 mg iv	Oral: Minimum of 1 l Metamucil-solution (13.6 g/l)	0.05 mmol/kg Gadodiamide (Omniscan	Phased array body coil	Bowel wall enhancement Bowel wall thickening Length of pathological bowel Stenosis Ulceration Cobblestoning Extraintestinal findings

NA: Not available

included [18, 28–30]. For our analyses we only used the data on CD. In two studies [29, 30] children were included; in the other studies, only adult patients were included (Table 1).

Study design characteristics

Selection criteria were described in six of the seven studies. In four of the studies patients were eligible for inclusion if they were scheduled for a colonoscopy. Hardly any clinical and laboratory data were provided in detail. Verification of results was complete in all studies, but in some of the patients the entire bowel could not be examined. In the studies evaluating disease activity per bowel segment, only the segments that were inspected at colonoscopy were used for comparison with the MRI findings. The criteria used to determine the presence of CD on the reference standard were not uniformly described. Information on whether the reference test was evaluated independently from the index test was not reported (Table 2).

Imaging features and imaging criteria used for diagnosis

In six studies the magnetic field strength was 1.5 T; in one study the field strength was 3.0 T. The bowel preparation, the use of luminal contrast medium and the type of coil that was used were not reported adequately in all studies. The type, concentration and amount of the intravenous contrast medium was reported in all studies (Table 3).

The one criterion considered indicative of disease in all studies was pathological bowel wall enhancement, while bowel wall thickening was used as parameter in six studies. However, different appraisals were used to determine pathological bowel wall enhancement; in the older studies percentages of contrast enhancement were used (post- and precontrast MRI), with higher ratios indicating more severe disease [28, 29]. In other studies subjective enhancement was used to stage disease [17, 18, 30–32]. With regard to bowel wall thickening, cutoff values to indicate the different stages of disease were provided only in one study [30]. All other imaging criteria (e.g., presence of stenosis, lymphadenopathy) were inconsistently used.

Data extraction

Data were reported on a per-patient basis in five studies [28–32] and on a per-segment basis in two studies [17, 18]. For the studies reporting segmental data, we grouped the available segmental data per patient to enable data analysis on a per-patient basis; only bowel segments that were inspected endoscopically or surgically were included, and the most severe segmental score was used for analysis. In four studies distinction was made among remission, mild, moderate and severe disease [28, 29, 31, 32]; in two studies only remission, mild and frank disease were distinguished [17, 18]. In one study numerical scores from 0 to 4 were given with stages 1 and 2 representing mild disease and stages 3 and 4 representing frank disease [30]. For each study, 3×3 (remission, mild, frank) contingency tables were constructed.

Table 4 Summary estimates for 3×3 data on per-patient basis

Study	Reference standard	Remission (n= 16)			Mild (n=29)			Frank (n=95)		
	MRI	Remission	Mild	Frank	Remission	Mild	Frank	Remission	Mild	Frank
Shoenut		0	0	0	0	1	0	0	0	11
Durno		0	0	0	0	0	2	0	1	6
Laghi		0	0	0	0	4	0	0	0	22
Florie: obs1		6	3	1	0	4	4	0	4	9
Florie: obs2*		5	4	1	2	3	3	1	2	10
Schreyer		1	0	0	0	1	0	1	1	26
Schreyer		0	0	0	2	2	0	0	1	7
Gemert: obs1		0	5	0	0	8	1	0	3	3
Gemert: obs2*		4	1	0	1	7	1	0	2	4
Summary estimates		0.62 (0.38–0.84)	0.31 (0.12–0.55)	0.07 (0.00–0.22)	0.17 (0.06–0.33)	0.62 (0.44–0.79)	0.21 (0.08–0.37)	0.02 (0.00–0.06)	0.07 (0.03–0.14)	0.91 (0.84–0.96)

Proportions in parentheses are confidence intervals
For calculation of the summary estimates, the results of observer 2 from the study of Florie and the results of observer 2 of the study of van Gemert were included. The model with these results had the lowest AIC value, indicating the best fit of the data

Summary estimates

The 3× 3 data

For calculating the summary estimates on the 3×3 data, we grouped moderate and severe disease together as frank disease from studies reporting 4×4 tables [28, 29, 31, 32].

In addition, in two studies [31, 32] results were provided for two observers (see Table 4). In both studies the data obtained by the second observer led to the lowest AIC value. Therefore, we used the results of observer 2 in both studies for calculating the summary estimates (Table 4). Data of in total 140 patients were used for the 3×3 data analysis with 16 patients in remission, 29 with mild disease and 95 with frank disease.

MRI correctly staged frank disease in a large proportion of patients (91%). Correct staging of mild disease by MRI occurred in 62% (95% CI: 38–84%) of the patients, and this estimate has a broad confidence interval, indicating the heterogeneity within the results. For remission, correct staging by MRI occurred in 62% (95% CI: 44–79%) of patients, also with heterogeneous results.

MRI overstaged disease activity in 37% of patients in remission, mostly as mild disease (31%). Overstaging of mild disease as frank disease was observed in 21% and understaging in 17%.

The 4× 4 data

In four studies distinction was made among remission, mild, moderate and severe disease [28, 29, 31, 32]. In two of these studies [31, 32], results were provided for two observers. However, due to the low number of data, analysis could not be performed to present summary estimates. In total 72 patients were evaluated in these four studies: 15 patients in remission, 20 with mild disease, 21 with moderate disease and 16 with severe disease. The results of the individual studies are reported in Table 5.

Discussion

MRI was highly accurate for diagnosing patients with frank disease. MRI more often overstaged than understaged disease activity in CD, but in most of these patients radiological staging and disease staging by the reference standard differed one grade.

An explanation for the inaccuracy in staging of patients with mild disease and patients in remission of MRI compared with the reference standard is the relative inexperience with evaluation of abdominal MRI for CD. Although bowel wall enhancement and bowel wall thickening are recognized as important parameters that indicate CD, no strict cutoff points have been defined yet to differentiate between the different stages of disease. This is reflected by the variation in definitions used in the different studies. In all included studies the subjective evaluation of the observers was very important for staging. Even in the studies wherein cutoff points were clearly described to differentiate among the different stages of disease, the radiologist had to subjectively define which bowel loop to use for assessment of enhancement and thickening.

Also, more patients were included with frank disease than with mild disease, while patients in remission were least often included. Frank disease is often easier to diagnose than mild disease or remission, as in this disease stage the parameters indicative of disease are most pronounced.

Another explanation for inaccuracy of MRI in staging is the fact that MRI and the reference standard are essentially different methods. With ileocolonoscopy only the lumen and the inner surface of the bowel wall can be assessed, while tissue sampling for histopathological examination only provides mucosal specimens. Meanwhile, on MRI the entire bowel wall with all its layers and the extraintestinal abdomen (e.g., the mesenteric vessels, mesenteric lymph nodes, mesenteric fat) are evaluated. As CD is a transmural disease, the extent of inflammatory or fibrotic changes might be better assessed on MRI than by inspection of the mucosal surface. A good next step would therefore be to

Table 5 Individual study results for 4×4 data on a per-patient basis

Reference standard	Remission (n=15)				Mild (20)				Moderate (21)				Severe (16)			
MRI	Rem	Mild	Mod	Sev	Rem	Mild	Mod	Sev	Rem	Mild	Mod	Sev	Rem	Mild	Mod	Sev
Shoenut	0	0	0	0	0	1	0	0	0	0	3	3	0	0	0	5
Durno	0	0	0	0	0	0	1	1	0	0	1	3	0	1	1	1
Florie:ob1	6	3	1	0	0	4	3	1	0	3	3	0	0	1	5	1
Florie:ob2	5	4	1	0	2	3	1	2	1	2	2	1	0	0	1	6
Gemert:ob1	0	5	0	0	0	8	1	0	0	3	2	0	0	0	1	0
Gemert:ob2	4	1	0	0	1	7	1	0	0	2	2	1	0	0	1	0

Rem = remission, Mild = mild disease, Mod = moderate disease, Sev = severe disease

compare MRI results with surgical pathology as in this manner all bowel wall layers can be examinated.

We only determined the ability of MRI to grade disease activity for the colon and terminal ileum, while CD can also be localized in the small bowel. We decided to limit our meta-analysis to findings in the colon and terminal ileum, as no reference standard was available for grading disease activity of the small bowel. The investigation that has often been used for evaluation of small bowel CD in the past (i.e., small bowel barium examination) is increasingly considered to be an imperfect reference standard. Comparative studies of MRI with established superior reference tests for the small bowel, such as double-balloon endoscopy (DBE) or video capsule endoscopy (VCE), are very scarce [33] as these endoscopic techniques were not commercially available until very recently and are only limitedly available at present. Also, for VCE or DBE the assessment of the severity of CD of the small bowel is not standardized yet.

A limitation of our analysis is the fact that we grouped moderate and severe disease together as frank disease. Information about the ability of MRI to differentiate between moderate and severe disease is discarded in this manner. However, we decided to put these data together in order to provide a more robust statement regarding the accuracy of MRI for disease activity, as only a limited amount of data was available. We provided these data to show the limited number of studies and the extreme heterogeneity in results between studies.

Another limitation is that although we accepted only colonoscopic, histopathological and/or surgical results as reference standard, the criteria for determination of disease activity on the reference standard were not identical between studies. Therefore, activity assessment on the reference standard might not have been consistent between studies. This might have influenced pooled accuracy estimates of MRI for staging disease activity. However, all three reference methods are reliable and are often used for assessment.

We decided not to perform subgroup analysis on the differences in technique, MR imaging criteria used or reference methods used as conclusions from subgroup analysis would not be very reliable due to the limited amount of data available. Therefore, we can not draw conclusions on the influence of the aforementioned differences for staging disease.

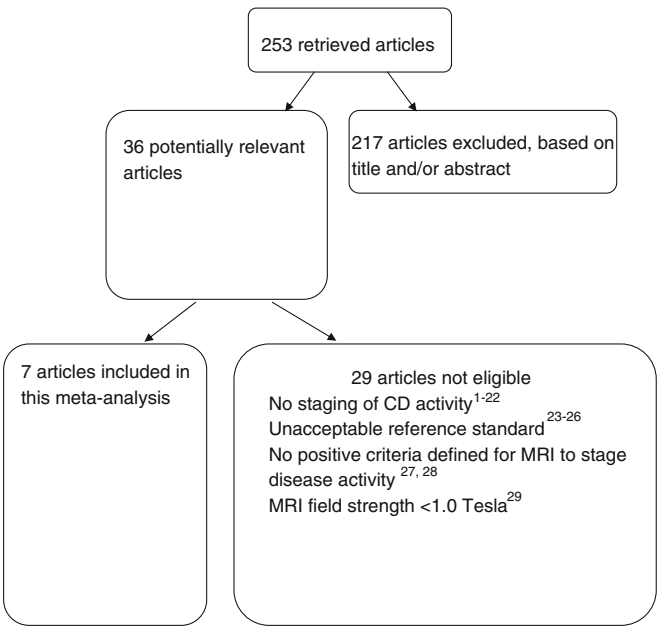
Before MRI can be implemented in routine clinical practice for the evaluation of CD, more research should be done on the reproducibility of MRI of the small bowel and colon. In our meta-analysis only two studies looked at interobserver agreement, and both reported moderate kappa values [31, 32]. As an imaging technique should be both accurate and reproducible, more studies are required to determine the role of MRI in clinical practice.

Also, before MRI can be used as a valid alternative for colonoscopy in the assessment of CD activity, it should become clear which imaging criteria are consistent with the different stages of CD. If standardized criteria were available internationally, larger trials would be possible, while comparison among studies would also be simplified. For that purpose, a more standardized technical imaging approach would be advisable as well. Future research should therefore focus on standardization of preparation, imaging technique and more uniform imaging criteria used for diagnosis of disease, in addition to including larger numbers of patients.

It would be interesting to see how other imaging techniques commonly used for evaluation of CD (i.e., computed tomography, ultrasonography) would perform in staging disease activity. Data on staging disease activity in CD are lacking for these techniques; by using the same inclusion criteria as we described above, only one article on power Doppler sonography [34] would be eligible for analysis (data not shown).

In conclusion, MRI can be used for staging disease activity in CD as with MRI most patients with frank disease are correctly diagnosed. However, in patients with disease in remission and mild disease, correct staging is limited.

Appendix 1: Search results



No staging of CD activity (n=22)

(1) Gualdi GF, Volpe A, Polettini E, Casciani E, Minervini S (1994) Computerized tomography and magnetic resonance in the evaluation of patients with Crohn disease. Their role in the identification, assessment of extent and management of the disease. *Article in Italian*. Clin Ter 144:545–551

(2) Gualdi GF, Polettini E, Minervini S (1994) Computerized tomography and magnetic resonance in Crohn's disease. *Article in Italian*. Ann Ital Chir 65:275–278
(3) Hansmann HJ, Kosa R, Düx M, et al. (1997) Hydro-MRI of chronic inflammatory bowel disease. *Article in German*. Fortschr Röntgenstr 167:132–138
(4) Holzknecht N, Helmberger T, Von Ritter C, Gauger J, Faber S, Reiser M (1998) MRI of the small intestine with rapid MRI sequences in Crohn disease after enteroclysis with oral iron particles. *Article in German*. Radiologe 38:29–36
(5) Low RN, Francis IR, Politoske D, Bennett M (2000) Crohn's disease evaluation: comparison of contrast-enhanced MR imaging and single-phase helical CT scanning. J Magn Reson Imaging 11:127–135
(6) Rieber A, Wruk D, Potthast S, et al. (2000) Diagnostic imaging in Crohn's disease: comparison of magnetic resonance imaging and conventional imaging methods. Int J Colorectal Dis 15:176–181
(7) Koh DM, Miao Y, Chinn RJ, et al (2001) MR imaging evaluation of the activity of Crohn's disease. AJR Am J Roentgenol 177:1325–1332
(8) Albert J, Scheidt T, Basler B, et al (2002) Magnetic resonance imaging in diagnosis and follow-up of Crohn's Disease–Is conventional enteroclysis still necessary? *Article in German*. Z Gastroenterol 40:789–794
(9) Miao YM, Koh DM, Amin Z, et al (2002) Ultrasound and magnetic resonance imaging assessment of active bowel segments in Crohn's disease. Clin Radiol 57:913–918
(10) Neurath MF, Vehling D, Schunk K, et al (2002) Noninvasive assessment of Crohn's disease activity: a comparison of 18F-fluorodeoxyglucose positron emission tomography, hydromagnetic resonance imaging, and granulocyte scintigraphy with labeled antibodies. Am J Gastroenterol 97:1978–1985
(11) Potthast S, Rieber A, Von Tirpitz C, Wruk D, Adler G, Brambs HJ. (2002) Ultrasound and magnetic resonance imaging in Crohn's disease: a comparison. Eur Radiol 12:1416–1422
(12) Ganten M, Encke J, Flosdorff P, Grüber-Hoffmann B, Erb G, Hansmann J (2003) Follow up of Crohn's disease under therapy with hydro-MRI. *Article in German*. Radiologe 43:26–33
(13) Magnano G, Granata C, Barabino A, et al (2003) Polyethylene glycol and contrast-enhanced MRI of Crohn's disease in children: preliminary experience. Pediatr Radiol 33:385–391
(14) Darbari A, Sena L, Argani P, Oliva-Hemker JM, Thompson R, Cuffari C (2004) Gadolinium-enhanced magnetic resonance imaging: a useful radiological tool in diagnosing pediatric IBD. Inflamm Bowel Dis 10:67–72
(15) Narin B, Ajaj W, Gohde S, et al (2004) Combined small and large bowel MR imaging in patients with Crohn's disease: a feasibility study. Eur Radiol 14:1535–1542
(16) Ajaj W, Lauenstein TC, Langhorst J, et al (2005) Small bowel hydro-MR imaging for optimized ileocecal distension in Crohn's disease: should an additional rectal enema filling be performed? J Magn Reson Imaging 22:92–100
(17) Ajaj WM, Lauenstein TC, Pelster G, et al (2005) Magnetic resonance colonography for the detection of inflammatory diseases of the large bowel: quantifying the inflammatory activity. Gut 54:257–263
(18) Godefroy C, Pilleul F, Dugougeat F, et al (2005) Value of contrast-enhanced MR enterography in pediatric Crohn's disease: preliminary study. *Article in French*. J Radiol 86:1685–1692
(19) Hohl C, Haage P, Krombach GA, et al (2005) Diagnostic evaluation of chronic inflammatory intestinal diseases in children and adolescents: MRI with true-FISP as new gold standard? *Article in German*. Fortschr Röntgenstr 177:856–863
(20) Röttgen R, Herzog H, Lopez-Hänninen E, Cho CH, Felix R, Schröder RJ (2005) Combination of dynamic MR enteroclysis (Sellink) and MR colonography to diagnose Crohn's disease. *Article in German*. Fortschr Röntgenstr 177:1131–1138
(21) Sempere GA, Martinez Sanjuan V, Medina Chulia E, et al (2005) MRI evaluation of inflammatory activity in Crohn's disease. AJR Am J Roentgenol 184:1829–1835
(22) Negaard A, Sandvik L, Mulahasanovic A, Berstad AE, Klow NE (2006) Magnetic resonance enteroclysis in the diagnosis of small-intestinal Crohn's disease: diagnostic accuracy and inter- and intra-observer agreement. Acta Radiol 47:1008–1016

Unacceptable reference standard (n=4)

(23) Low RN, Sebrechts CP, Politoske DA, et al (2002) Crohn disease with endoscopic correlation: single-shot fast spin-echo and gadolinium-enhanced fat-suppressed spoiled gradient-echo MR imaging. Radiology 222:652–660
(24) Pilleul F, Godefroy C, Yzebe-Beziat D, Dugougeat-Pilleul F, Lachaux A, Valette PJ (2005) Magnetic resonance imaging in Crohn's disease. Gastroenterol Clin Biol 29:803–808
(25) Florie J, Wasser MN, Arts-Cieslik K, Akkerman EM, Siersema PD, Stoker J (2006) Dynamic contrast-enhanced MRI of the bowel wall for assessment of disease activity in Crohn's disease. AJR Am J Roentgenol 186:1384–1392
(26) Maccioni F, Bruni A, Viscido A, Colaiacomo MC, Cocco A, Montesani C et al (2006) MR imaging in patients with Crohn disease: value of T2- versus T1-weighted gadolinium-enhanced MR sequences with use of an oral superparamagnetic contrast agent. Radiology 238:517–530

No positive criteria defined for MRI to stage disease activity (n=2)

(27) Rottgen R, Herzog H, Lopez-Haninnen E, Felix R (2006) Bowel wall enhancement in magnetic resonance colonography for assessing activity in Crohn's disease. Clin Imaging 30:27–31

(28) Schunk K, Reiter S, Kern A, Orth T, Wanitschke R (2001) Hydro-MRI in inflammatory bowel diseases: a comparison with colonoscopy and histology. *Article in German*. Fortschr Röntgenstr 173:731–738

MRI field strength <1.0 T (n=1)

(29) Madsen SM, Thomsen HS, Munkholm P, et al (2002) Inflammatory bowel disease evaluated by low-field magnetic resonance imaging. Comparison with endoscopy, 99mTc-HMPAO leucocyte scintigraphy, conventional radiography and surgery. Scand J Gastroenterol 37:307–316

References

1. Travis SPL, Stange EF, Lémann M et al (2006) European evidence based consensus on the diagnosis and management of Crohn's disease: current management. Gut 55(Suppl 1):i16–i35
2. Baumgart DC, Sandborn WJ (2007) Inflammatory bowel disease: clinical aspects and established and evolving therapies. Lancet 369:1641–1657
3. Hommes DW, van Deventer SJ (2004) Endoscopy in inflammatory bowel diseases. Gastroenterology 126:1561–1573
4. Booya F, Fletcher JG, Huprich JE et al (2006) Active Crohn disease: CT findings and interobserver agreement for enteric phase CT enterography. Radiology 241:787–795
5. Bodily KD, Fletcher JG, Solem CA et al (2006) Crohn Disease: mural attenuation and thickness at contrast-enhanced CT Enterography–correlation with endoscopic and histologic findings of inflammation. Radiology 238:505–516
6. Paulsen SR, Huprich JE, Fletcher JG et al (2006) CT enterography as a diagnostic tool in evaluating small bowel disorders: review of clinical experience with over 700 cases. Radiographics 26:641–657
7. Hara AK, Leighton JA, Heigh RI et al (2006) Crohn disease of the small bowel: preliminary comparison among CT enterography, capsule endoscopy, small-bowel follow-through, and ileoscopy. Radiology 238:128–134
8. Chiorean MV, Sandrasegaran K, Saxena R et al (2007) Correlation of CT enteroclysis with surgical pathology in Crohn's disease. Am J Gastroenterol 102:2541–2550
9. Desmond AN, O'Regan K, Curran C et al (2008) Crohn's disease: factors associated with exposure to high levels of diagnostic radiation. Gut 57:1524–1529
10. Brenner DJ, Hall EJ (2007) Computed tomography–an increasing source of radiation exposure. N Engl J Med 357:2277–2284
11. Stange EF, Travis SPL, Vermeire S et al (2006) For the European Crohn's and Colitis Organisation (ECCO). European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. Gut 55 (Suppl 1):i1–i15
12. Maccioni F, Bruni A, Viscido A et al (2006) MR imaging in patients with Crohn disease: value of T2- versus T1-weighted gadolinium-enhanced MR sequences with use of an oral superparamagnetic contrast agent. Radiology 238:517–530
13. Low RN, Sebrechts CP, Politoske DA et al (2002) Crohn disease with endoscopic correlation: single-shot fast spin-echo and gadolinium-enhanced fat-suppressed spoiled gradient-echo MR imaging. Radiology 222:652–660
14. Horsthuis K, Bipat S, Bennink RJ, Stoker J (2008) Ultrasonography, magnetic resonance imaging, scintigraphy, and computed tomography for diagnosis of inflammatory bowel disease: a meta-analysis of prospective studies. Radiology 247:64–79
15. Ryan ER, Heaslip IS (2008) Magnetic resonance enteroclysis compared with conventional enteroclysis and computed tomography enteroclysis: a critically appraised topic. Abdom Imaging 33:34–37
16. Albert J, Scheidt T, Basler B et al (2002) Magnetic resonance imaging in diagnosis and follow-up of Crohn's Disease–Is conventional enteroclysis still necessary? *Article in German*. Z Gastroenterol 40:789–794
17. 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
18. 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
19. Ajaj WM, Lauenstein TC, Pelster G et al (2005) Magnetic resonance colonography for the detection of inflammatory diseases of the large bowel: quantifying the inflammatory activity. Gut 54:257–263
20. Narin B, Ajaj W, Gohde S et al (2004) Combined small and large bowel MR imaging in patients with Crohn's disease: a feasibility study. Eur Radiol 14:1535–1542
21. Pilleul F, Godefroy C, Yzebe-Beziat D, Dugougeat-Pilleul F, Lachaux A, Valette PJ (2005) Magnetic resonance imaging in Crohn's disease. Gastroenterol Clin Biol 29:803–808
22. Sempere GA, Martinez Sanjuan V, Medina Chulia E et al (2005) MRI evaluation of inflammatory activity in Crohn's disease. AJR Am J Roentgenol 184:1829–1835
23. 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

24. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J (2006) Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 6:9
25. 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
26. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337
27. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Control 19:716
28. 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
29. Durno CA, Sherman P, Williams T, Shuckett B, Dupuis A, Griffiths AM (2000) Magnetic resonance imaging to distinguish the type and severity of pediatric inflammatory bowel diseases. J Pediatr Gastroenterol Nutr 30:170–174
30. Laghi A, Borrelli O, Paolantonio P et al (2003) Contrast enhanced magnetic resonance imaging of the terminal ileum in children with Crohn's disease. Gut 52:393–397
31. 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
32. 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
33. Triester SL, Leighton JA, Leontiadis GI et al (2006) A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with non-stricturing small bowel Crohn's disease. Am J Gastroenterol 101:954–964
34. 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