**Diagnostic accuracy for the extent and activity of newly diagnosed and relapsed Crohn’s disease: a multicentre prospective comparison of magnetic resonance enterography and small bowel ultrasound –The METRIC study**

Stuart A Taylor1 FRCR, Susan Mallett2, Gauraang Bhatnagar1, Rachel Baldwin-Cleland3, Stuart Bloom4, Arun Gupta3, John Hamlin5, Ailsa Hart3, Antony Higginson6, Ilan Jacobs7, Sara McCartney4, Anne Miles8, Steve Morris9, Charles Murray10, Andrew Plumb1, Richard Pollok11, Shonit Punwani1, Laura Quinn2, Manuel Rodriquez-Justo12, Zainib Shabir13,Andrew Slater14, Damian Tolan15, Simon Travis16, Alastair Windsor17, Peter Wylie18, Ian Zealley19, Steve Halligan1, on behalf of the METRIC investigators\*

1. Centre for Medical Imaging, University College London, Charles Bell House, 43-45 Foley street, W1W 7TS
2. Institute of Applied Health Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
3. Intestinal Imaging Centre, St Mark's Hospital, Harrow, UK
4. Department of Gastroenterology, University College Hospital, 235 Euston Road, London, NW1 2BU, UK
5. Leeds Gastroenterology Institute, Leeds General Infirmary, Great George Street, Leeds, LS1 3EX, UK
6. Department of Radiology, Portsmouth Hospitals NHS Trust, Portsmouth, UK
7. X
8. Department of Psychological Sciences, Birkbeck University of London, London, United Kingdom
9. Applied Health Research, University College London, London, United Kingdom
10. Department of Gastroenterology and Endoscopy, Royal Free London NHS Foundation Trust, London, UK
11. Department of Gastroenterology, St George's Hospital, London SW17 0QT, United Kingdom
12. Department of Histopathology, University College Hospital, 235 Euston Road, London, NW1 2BU, UK
13. Comprehensive Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, Holborn, London
14. Department of Radiology, John Radcliffe Hospital, Oxford, UK.
15. Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals Trust, Beckett Street, Leeds, LS9 7TF, UK
16. Translational Gastroenterology Unit, Oxford University Hospitals, Oxford, UK
17. Department of Surgery, University College Hospital, 235 Euston Road, London, NW1 2BU, UK
18. Department of Radiology, Royal Free London NHS Foundation Trust, London, UK
19. Department of Radiology, Ninewells Hospital, Dundee, DD1 9SY, Scotland, UK

Corresponding author

Stuart A Taylor, Centre for Medical Imaging, University College London, Charles Bell House, 43-45 Foley street, W1W 7TS. Email stuart.taylor1@nhs.net. Tel 07960169321

\*METRIC investigators

Jade Dyer, Pranitha Veeramalla,

Richard Ellis, Fergus Thursby-Pelham, Richard Beable, Nicola Gibbons, Claire Ward

Anthony O’Connor, Hannah Lambie, Rachel Hyland, Nigel Scott, Roger Lapham, Doris Quartey, Deborah Scimshaw

Helen Bungay, Maggie Betts, Simona Fourie

Niall Power

Rajapandian Ilangovan, Uday Patel, Evgenia Mainta, Phillip Lung, Ian Johnston, Mani Naghibi, Morgan Moorghen, Adriana Martinez, Francois Porte

Christopher Alexakis, James Pilcher, Anisur Rahman, Jonny Vlahos, Rebecca Greenhalgh, Anita Wale

Teresita Beeston, Wivijin Piga, Joey Clemente, Farooq Rahman, Simona de Caro, Shameer Metha, Rosa Vega, Roman Jastrub , Harbir Sidhu, Hameed Rafiee

Mairead Tennent, Caron Innes, Craig Mowat, Gillian Duncan

Research in context

Evidence before this study

Cross sectional imaging is fundamental for diagnosis and management of Crohn’s disease and is replacing barium fluoroscopic techniques which have been the bedrock of small bowel imaging for many years. The dissemination of cross sectional imaging has however occurred despite a paucity of supportive data from prospective multi centre studies recruiting consecutive and unselected patients . Emphasis is placed on magnetic resonance imaging enteography (MRE) and enteric ultrasound (US) as they avoid ionising radiation. Clinical uptake of US has been hampered by concerns over diagnostic accuracy and perceived high levels of inter-observer variation. MRE is a more recent innovation necessitating access to comparatively restricted high technology imaging platforms. We searched PubMed and Embase for articles between Jan 1st 1990 and Dec 1st 2017 without language restriction. We used MeSH and full-text search for “Crohn’s disease”, “magnetic resonance imaging”, “ultrasound”, and “diagnostic accuracy”. Emphasis was placed on meta-analyses and systematic reviews using appropriate search restrictions. We found a number of meta-analyses, which in general suggest MRE and US are comparable for detecting small bowel Crohn’s disease. However, the primary literature has limitations. Most studies are small single centre explanatory trials recruiting less than 50 patients. Tests are rarely compared in the same patients, introducing bias caused by differences between patients and disease phenotype, and use inconsistent reference standards. Many also score poorly on the Quality Assessment of Diagnostic Accuracy Studies tool.

Added value of this study

This is the largest prospective multicenter study to date comparing the diagnostic accuracy of MRE and US for the presence, extent and activity of enteric Crohn’s disease, using a construct reference standard incorporating 6 months of patient follow up. We used a pragmatic trial design to better assess test performance in routine clinical practice, and used the preferred methodology for diagnostic accuracy studies, by comparing tests in the same patients. Both tests achieved high accuracy for detecting and localising small bowel Crohn’s disease, but sensitivity and specificity for small bowel disease presence and extent was significantly greater for MRE than US.

Implications of all the available evidence

MRE and US are both accurate for detecting the presence and extent of small bowel Crohn’s disease, but MRE is superior.

Abstract

Background

Magnetic resonance enteography (MRE) and enteric ultrasound (US) are used variably to image Crohn’s disease. We compared their diagnostic accuracy for presence, extent and activity of enteric disease

Methods

This pragmatic multicenter cohort study recruited from 8 UK hospitals. Eligible patients were 16 years or older, newly diagnosed with Crohn’s disease, or had established disease with suspected relapse. Patients underwent both MRE and US in addition to standard clinical investigations. Discrepancy between MRE and US for small bowel (SB) disease presence triggered an additional investigation, if not already available. The primary outcome was difference in per patient sensitivity for SB disease extent (correct identification and segmental localisation) against a construct reference standard (panel diagnosis). Accuracy for SB and colonic disease presence and activity were secondary outcomes.

Findings

284 patients completed the study (133 new diagnosis, 151 relapse). MRE sensitivity for SB disease extent (80% [95% CI 72 to 86]) and presence (97% [91 to 99]) were significantly greater than US (70% [62 to 78], 92% [84 to 96]); a 10% (1 to 18), and 5% (1 to 9), difference respectively. MRE specificity for SB disease extent (95% [85 to 98]) was also significantly greater than US (81% [64 to 91]). Sensitivity for active SB disease was significantly greater for MRE than US (96% [92 to 99] vs. 90% [82 to 955], difference 6% (2 to 11). Across all patients, there were no significant differences in accuracy for colonic disease extent or presence. Accuracy in newly diagnosed and relapse patients was generally similar, although US had significantly greater sensitivity for colonic disease in newly diagnosed patients (67% [47 to 81) vs. 47% (31 to 64), difference 20% (1 to 39).

Interpretation

MRE has higher diagnostic accuracy for the extent and activity of small bowel Crohn’s disease than US.

Funding

NIHR Health Technology Assessment Program

Introduction

Small bowel imaging is fundamental for comprehensive phenotyping of Crohn’s disease and essential to direct therapeutic strategy1. Barium fluoroscopy has long been the bedrock of small bowel investigation, providing detailed mucosal evaluation2. However, in recent years enthusiasm for barium has dwindled, and it is replaced increasingly by cross sectional imaging, namely computed tomography enterography (CTE), magnetic resonance imaging enteography (MRE), and ultrasound (US). Advocates stress that these techniques evaluate the bowel wall and beyond, complimenting endoscopic visualisation. As barium fluoroscopy is abandoned, dissemination of cross sectional imaging has been relatively uncontrolled, despite a paucity of supportive data from methodologically sound prospective multi-centre studies. This lack of robust evidence is concerning given the pivotal role assumed by small bowel imaging over the lifetime of patients with Crohn’s disease.

Of the available modalities, MRE and US are preferred3 since they avoid irradiating4 generally young patients who require repeat imaging. Enteric US is longer established5, requires little patient preparation, and the technology is widely available. However, questions remain over accuracy, particularly in the proximal bowel and deep pelvis 6, and perceived inter-observer variability 7. Conversely, MRE is a more recent innovation8, requires oral contrast and access to high technology imaging platforms, which are comparatively restricted in many health care settings.

Although meta-analyses suggest MRE and US are largely equivalent for diagnosing and staging Crohn’s disease6,9-20, the primary literature is of questionable quality. The large majority of studies are small, single centre, 21 17,20 and few compare tests directly in the same patient, despite this being advocated as optimal methodology for diagnostic accuracy studies22. Also, very few utilise a construct reference standard paradigm (panel diagnosis), which incorporates concepts of diagnostic test validation based on patient outcomes, and has distinct methodical advantages when a single reference standard is elusive23.

In order to redress this, we conducted a prospective multicenter study to elucidate and then directly compare the diagnostic accuracy of MRE and US for small bowel Crohn’s disease against a construct reference standard incorporating follow up. To reflect normal clinical practice, we recruited both newly diagnosed patients and those with established disease in whom luminal relapse was suspected.

Methods

Study Design

The METRIC study is a multicentre, prospective cohort study comparing the diagnostic accuracy of MRE and enteric ultrasound US for the presence, extent and activity of small bowel Crohn’s disease in newly diagnosed patients, or patients with established disease and suspected relapse (ISRCTN03982913). The trial achieved National Health Service research ethics committee approval in September 2013 (13/SC/0394) and was conducted in accordance with the principles of Good Clinical Practice (GCP). The trial was supervised by University College London’s Clinical Trials Unit and independent Data Monitoring and Trial Steering Committees. All patients recruited gave written informed consent. The full study protocol has been published previously24 and can be found online.

Patients and recruitment sites

Patients were recruited from 8 UK National Health Service (NHS) teaching and general hospitals, representative of institutions likely to implement MRE and US for patient management (appendix table 1). All sites had an established inflammatory bowel disease service treating >150 patients annually and were already performing MRE and US as part of usual clinical practice.

Patients were eligible for the new diagnosis subgroup if they had been diagnosed with Crohn’s disease in the 3 months preceding recruitment based on conventional diagnostic criteria, or where Crohn’s disease was strongly suspected based on imaging or endoscopic features but pending final diagnosis. Eligible patients had already undergone colonoscopy or were awaiting it at recruitment. Patients in whom the final diagnosis was not Crohn’s disease were excluded subsequently.

Patients were eligible for the suspected luminal relapse subgroup if they had established Crohn’s disease (for greater than 3 months) and high clinical suspicion of luminal relapse based on objective markers of inflammatory activity (CRP >8mg/l or faecal calprotectin >100mcg/g), and/or symptoms suggestive of luminal stenosis (including obstructive symptoms such as colicky abdominal pain, vomiting), and/or abnormal endoscopy suggesting relapse. Eligible patients for both arms were aged ≥16. Patients were ineligible if they were pregnant or had any contraindication to MRI.

Suitable patients were identified from outpatient clinics, multi-disciplinary team meetings and in-patient wards by members of the local research team, who took informed consent. A screening log detailed all patients approached to take part in the study, and reasons for non-participation, if applicable. Patient demographics and baseline clinical data were collated (for example age, sex, Montreal classification [relapse subgroup only], disease/symptom duration, current medication and surgical history).

Interventions

Patients underwent MRE and US in addition to any other enteric imaging or endoscopic investigations performed during their usual clinical care.

MRE was undertaken according to local standard clinical protocols (including the choice of oral contrast agent) on either 1.5T or 3T MRI platforms. A minimum dataset of sequences was acquired (appendix table 2). US was performed by radiologists or sonographers using standard platforms and both curvilinear and high-resolution probes, without oral or intravenous contrast agents.

Across all sites, 28 practitioners interpreted MRE and US studies (27 radiologists interpreted MRE and US and 1 sonographer performed and interpreted US). All radiologists were affiliated with the British Society of Gastrointestinal and Abdominal Radiology (BSGAR) with declared subspecialty interest in gastrointestinal radiology and were either consultant grade or post Fellowship of the Royal College of Radiologists, with at least one year of sub-specialty training in gastrointestinal radiology. The sonographer had undergone formal training according to their sites’ local polices and was performing enteric US routinely. Before study commencement, a two-day hands-on workshop for investigators was held to standardise US technique and agree description of enteric findings.

MRE and US were interpreted by two different practitioners blinded to the findings of the other, and to all other imaging, endoscopic and clinical data other than the cohort to which the patient was recruited (i.e. new diagnosis or relapse) and surgical history (since this information would normally be provided on the clinical request). Using case report forms, practitioners noted the presence and activity of Crohn’s disease in the small bowel and colon, together with any extra-enteric complications, using established criteria6,16,25. The segmental location of any disease was also recorded24; disease sites separated by >3cm of normal bowel within a particular segment were recorded separately. Diagnostic confidence for disease presence was scored from 1 to 6, grouped into normal (levels 1,2) equivocal (levels 3,4) and abnormal (levels 5,6). A formal clinical report was then generated as per usual clinical practice.

The nature and findings of all other small bowel imaging or endoscopies performed as part of usual care were collected by members of the local research team. These tests were performed and interpreted according to usual clinical practice at local sites, without blinding. A case report form recorded colonoscopic findings specifically.

If there was discrepancy between MRE and US for the presence or location of small bowel disease, an “arbiter” small bowel investigation was performed if patients had not already undergone additional small bowel imaging as part of usual care. Discrepancy was defined as terminal ileal disease reported on one of MRE or US in the absence of endoscopic visualisation, and/or disease reported in the small bowel upstream of the terminal ileum on one of MRE or US. The nature of the additional test was left to local discretion and could include for example barium follow through, CTE or capsule endoscopy. Repeat targeted unblinded MRE or US was also permitted to resolve discrepancies.

Where possible, CRP calprotectin and the Harvey Bradshaw index were collected at recruitment and repeated between 10 and 20 weeks later.

 Reference standard

We used the construct reference standard paradigm (panel diagnosis) incorporating the concept of clinical test validation, i.e. whether test results are meaningful in practice23. Specifically, we followed patients’ clinical course for 6 months to assess the impact of MRE and US findings on clinical decision making and patient outcomes. Each recruitment site convened a series of consensus panels consisting of at least one local gastroenterologist and two radiologists (one local and one from another site); a histopathologist was available to the panel if required and a member of the trial management group attended to ensure uniformity of process. For each individual patient, the panel considered the images and results of all small bowel investigations (including MRE and US) and all additional information accrued over the follow-up period including endoscopies, surgical findings**,** histopathology, HBI, CRP, calprotectin (and changes thereof), and clinical course. The panel recorded its opinion as to whether small bowel or colonic Crohn’s disease was present, and, if so, whether disease was active. Disease could only be categorised as active if there was at least one objective marker of this [(i) ulceration as seen at endoscopy and/or (ii) measured CRP >8 mg/l and/or (iii) measured calprotectin >250 and/or (iv) histopathological evidence of acute inflammation based on biopsy or surgery within 2 months of study imaging].

Outcomes

The primary outcome was the per patient difference in sensitivity between MRE and US for correct identification and localisation of small bowel Crohn’s disease, irrespective of activity, i.e. the extent of small bowel disease. To be true positive for disease extent, the index test had to correctly locate both the presence and segmental location of disease. Secondary outcomes included specificity for disease extent, sensitivity and specificity for small bowel disease presence, the difference in per patient sensitivity and specificity for colonic disease presence and extent, and identification of active disease. Most outcomes were reported for the new diagnosis and suspected luminal relapse subgroups individually, and for the terminal ileum and colon using colonoscopy as a standalone reference standard (when available), due to its robustness for identifying disease.

Statistical analysis

We estimated that a sample size of 210 patients with small bowel disease would give 90% power to detect a significant (10%) sensitivity difference between MRE and US for disease extent24. Assuming a 70% prevalence of small bowel disease and 10% loss to follow up/non-Crohn’s disease diagnosis, gave a target sample size of 334 patients across both cohorts (167 new diagnosis and 167 relapse).

Disease reported as equivocal was treated as positive in the analysis. The primary outcome was calculated per patient. Secondary outcomes for bowel segments were based on all segments, excluding those resected at baseline.

Direct comparison of sensitivity and specificity differences between MRE and US were calculated using multilevel bivariate models, from paired data using xtmelogit in STATA 14.0 [College Station, Texas 77845 USA]. When models did not converge due to small numbers of patients, McNemar’s comparison of paired proportions was used to obtain univariable estimates and exact 95% CI were calculated. Statistical significance was based on 95% CI.

Role of funder

The primary funder (the National Institute for Health Research) stipulated a diagnostic accuracy study using a cohort design but no funders were involved in the collection, analysis, or interpretation of data, nor in the writing or submitting of this report. The corresponding author had full access to all data and final responsibility for the decision to submit for publication

Results

Recruitment commenced December 2013 and completed September 2016. Overall, 518 patients were assessed for eligibility of whom 183 were excluded (figure 1). Of the 335 patients who entered the trial, 51 were excluded subsequently (20 male, mean age 33 years [range 17 to 72]); 31 did not have Crohn’s disease, 2 were lost to follow up, 10 did not undergo MRE and/or US, 6 withdrew consent, and 2 newly diagnosed patients underwent surgery without colonoscopy. This gave a final cohort of 284 (133 new diagnosis and 151 relapse), (figure 1, table 1). One hundred and fifty-four (54%) were women. Based on the reference standard, 233 (82%) patients had small bowel Crohn’s disease (thereby meeting sample size stipulations), which was active in 209 (90%) (table 2). One hundred and twenty-nine (45%) had colonic disease, which was active in 126 (98%). Twenty-one had enteric fistulae, and 7 had intraabdominal abscess.

In total, 53 patients (24 new diagnosis, 29 relapse) were discrepant for small bowel disease presence or location of whom 48 (91%) had an additional small bowel imaging test available to the consensus panel. The range of imaging, endoscopic and biochemical data available to the consensus panels is shown in appendix table 3.

MRE sensitivity for small bowel disease extent (i.e. presence and correct segmental location), was 80% (95% CI 72 to 86) compared to 70% (62 to 78) for US, a difference of 10% (1 to 18), which was statistically significant (table 3, figure 2). MRE specificity for small bowel disease extent was also significantly greater than US: 95% (85 to 98) vs. 81% (64 to 91) respectively, a difference of 14% (1 to 27).

MRE sensitivity for small bowel disease presence, regardless of location was 97% (91 to 99)*,* significantly greater than US (92% [84 to 96]), a difference of 5% (1 to 9), (figure 2).

There were no significant differences in sensitivity or specificity between MRE and US for colonic disease extent or presence (table 3, figure 2).

The detection rate for individual small bowel and colonic segments is given in appendix table 4. Although the study was not powered to detect differences on a segmental level, MRE was significantly more sensitive than US for ileal (84% [67 to 93] vs 56% [38 to 73) and rectal disease (44% [32 to 58] vs. 22% [13 to 35].

Sensitivities of MRE and US for small bowel disease presence and extent in the new and relapse patient cohorts were very similar to those estimated across all patients (table 4), although differences were not statistically significant (the study was not powered to detect differences in the cohorts). US however had significantly greater sensitivity for colonic disease presence than MRE in the new patient cohort (67% [49 to 81] vs. 47% [31 to 64]), a difference 20% (1 to 39). For both MRE and US, sensitivity for colonic disease tended to be higher in the relapse patient cohort (table 4), although the estimated sensitivity for colonic disease extent was poor for both.

MRE sensitivity for active small bowel disease was 96% (92 to 99) compared to 90% (82 to 95) for US, a significant difference of 6% (2 to 11) (table 5). Specificity for active small bowel disease and accuracy for active colonic disease were not significantly different between tests (table 5).

Sensitivity and specificity for active disease split by patient cohort were very similar to those estimated across all patients (appendix table 5).

MRE detected 5/7 (71%) abscesses and 18/21 (86%) patients with enteric fistulae compared to 3/7 (43%) and 11/21 (52%) for US respectively.

Against a colonoscopic standard of reference (available in 186 patients), MRE had a sensitivity of 97% (91 to 99), for terminal ileal disease presence compared to a sensitivity of 91% (79 to 97) for US, a difference of 6% (-1 to 12) (appendix table 6). Sensitivity for colonic disease presence was modest for both MRE and US (41% [26 to 58] and 49% (33 to 65]). These differences were not statistically significant (the study was not powered to detect differences based on a colonoscopic standard of reference alone).

Discussion

At the time of writing, METRIC is the largest prospective multicenter study directly comparing diagnostic accuracy of MRE and US for the presence, extent and activity of enteric Crohn’s disease in the same patients. Although both techniques have variably replaced barium fluoroscopy in clinical practice, robust evidence supporting this transition has been modest until now. Our results provide this evidence: We found both MRE and US highly accurate for detecting small bowel Crohn’s disease, achieving 96% and 92% sensitivity respectively. Barium fluoroscopy has long been advocated as a sensitive test for mucosal disease inaccessible to endoscopy, although support is limited to a handful of small studies2 and accuracy is increasingly questioned26. Conversely, against a rigorous ileo-colonoscopic reference standard, we found MRE and US achieved 97% and 91% sensitivity for terminal ileal disease, strongly supporting their utility as first line investigation and positioning them as competitive and viable diagnostic alternatives to invasive ileo colonoscopy.

Of the two, we found MRE had significantly higher sensitivity and specificity than US for small bowel extent, and higher sensitivity for disease presence. Overall, there was no significant difference in diagnostic accuracy for colonic disease (which was consistently lower than for small bowel disease), although US had greater sensitivity than MRE in newly diagnosed patients.

Our primary outcome combined those aspects necessary to stage small bowel Crohn’s disease correctly, i.e. is disease present and, if so, where? Both presence and extent dictate subsequent therapeutic strategy. For example, the finding of additional proximal small bowel disease may tip the balance towards medical rather than surgical intervention in the face of otherwise limited terminal ileal disease. As expected, sensitivity for disease extent was lower than that for disease detection alone.

We found detection rates at the upper end of estimates from prior meta-analyses6,9-20. Dong et al12 estimated sensitivity and specificity of US at 88% and 97% respectively, while Liu et al17 reported corresponding figures of 86% and 93% for MRE. However, the primary literature is markedly heterogeneous, which impacts on validity of point estimates. Most studies are single centre, typically recruit fewer than 50 patients, and are methodologically poor 17,21. Direct comparison of diagnostic tests in the same patients is advocated as the optimal methodology for diagnostic accuracy studies22 as differences are attributable directly to the tests and not differences between participants or study methods. Such head-to-head comparisons are rare in the literature. For example in their recent metanalysis, Greenup and al found just one of 33 included studies compared MRE and US directly in the same patients15. Reference standards may also be applied inconsistently, with endoscopy, surgery, and imaging all variably employed. For example, in a comparative study with US, Castiglione et al used MRE without any additional reference standard in many recruits 27. The potential for incorporation bias is self-evident.

We used the construct reference standard paradigm (panel diagnosis), which incorporates multiple data sources with clinical outcome 23. Although such an approach does have limitations, including potential panel bias, it is considered a very robust methodology for diagnostic accuracy studies where a single external reference standard is elusive23. To reduce incorporation bias, patients without supplementary small bowel imaging underwent a third small bowel investigation whenever discrepancy between MRE and US arose. It is notable than when our analysis was limited to an ileo-colonoscopic reference standard, any differences in accuracy between MRE and US closely mirrored those found using the consensus panel reference.

We recruited approximately equally from two clinically distinct patient cohorts; new diagnoses of Crohn’s disease and those with established disease, suffering relapse. Both are clinically distinct and important, and may manifest with differing disease phenotypes; prevalence of sticturing and penetrating disease increases with time28. Noting that METRIC was not powered to detect differences between these two cohorts, we found that sensitivity for small bowel disease was similar, although specificity tended to be a little lower in those with relapse. Conversely, sensitivity for colonic disease was higher in the relapse cohort, but was still poor for colonic disease extent (around 30%).

In newly diagnosed patients US achieved significantly greater sensitivity for colonic disease than MRE (although this was still modest at 67%). Optimised colonic evaluation using MRE requires purgation and fluid distension29, which are both omitted from standard MRE protocols, whereas, in general, US relies on evaluating the manually compressed uncleansed colon wall. Accuracy for both techniques in the colon still falls short of colonoscopy however.

METRIC was conceived as a large pragmatic trial30 since the literature is replete with small explanatory studies. We recruited from a range of hospital settings, both teaching and district general, and used local imaging protocols to enhance generalisability. The 28 practitioners all declared a specialist interest in gastrointestinal radiology and were representative of those reporting NHS small bowel imaging in terms of training and experience. We specifically avoided using a small number of highly experienced practitioners since they would not represent a national workforce. Imaging was interpreted according to local clinical practice so as to mirror “real world” procedures and enhance generalisability of our results. We acknowledge that blinding practitioners to individual patient history does not mirror usual clinical practice, but this precaution was necessary to isolate diagnostic test accuracy as far as possible. Recruited patients were, as far as possible, representative of those undergoing MRE and US in daily practice, although we did exclude pregnant women and patients with contraindications to MRI. Our results are therefore highly likely to extrapolate to the NHS and similar health care settings.

Diagnostic accuracy is clearly paramount when patients are investigated but patient experience, interobserver variability, and cost effectiveness are also of great importance, and will be reported elsewhere.

In summary, we found that both US and MRE achieve excellent diagnostic accuracy for the extent and activity of small bowel Crohn’s disease in both newly diagnosed patients and those suffering relapse. Both tests are valid first-line investigations and are competitive with ileocolonoscopy for evaluating the terminal ileum. MRE is the preferred radiological investigation when available because its sensitivity and specificity exceed US significantly.

References

1. Peyrin-Biroulet L, Fiorino G, Buisson A, Danese S. First-line therapy in adult Crohn's disease: who should receive anti-TNF agents? *Nature reviews Gastroenterology & hepatology* 2013; **10**(6): 345-51.

2. Bernstein CN, Boult IF, Greenberg HM, van der Putten W, Duffy G, Grahame GR. A prospective randomized comparison between small bowel enteroclysis and small bowel follow-through in Crohn's disease. *Gastroenterology* 1997; **113**(2): 390-8.

3. Panes J, Bouhnik Y, Reinisch W, et al. Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. *Journal of Crohn's & colitis* 2013; **7**(7): 556-85.

4. Desmond AN, O'Regan K, Curran C, et al. Crohn's disease: factors associated with exposure to high levels of diagnostic radiation. *Gut* 2008; **57**(11): 1524-9.

5. Sonnenberg A, Erckenbrecht J, Peter P, Niederau C. Detection of Crohn's disease by ultrasound. *Gastroenterology* 1982; **83**(2): 430-4.

6. Panes J, Bouzas R, Chaparro M, et al. 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. *Alimentary pharmacology & therapeutics* 2011; **34**(2): 125-45.

7. Fraquelli M, Sarno A, Girelli C, et al. Reproducibility of bowel ultrasonography in the evaluation of Crohn's disease. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 2008; **40**(11): 860-6.

8. Shoenut JP, Semelka RC, Magro CM, Silverman R, Yaffe CS, Micflikier AB. Comparison of magnetic resonance imaging and endoscopy in distinguishing the type and severity of inflammatory bowel disease. *Journal of clinical gastroenterology* 1994; **19**(1): 31-5.

9. Ahmed O, Rodrigues DM, Nguyen GC. Magnetic Resonance Imaging of the Small Bowel in Crohn's Disease: A Systematic Review and Meta-Analysis. *Canadian journal of gastroenterology & hepatology* 2016; **2016**: 7857352.

10. Choi M, Lim S, Choi MG, Shim KN, Lee SH. Effectiveness of Capsule Endoscopy Compared with Other Diagnostic Modalities in Patients with Small Bowel Crohn's Disease: A Meta-Analysis. *Gut and liver* 2017; **11**(1): 62-72.

11. Church PC, Turner D, Feldman BM, et al. Systematic review with meta-analysis: magnetic resonance enterography signs for the detection of inflammation and intestinal damage in Crohn's disease. *Alimentary pharmacology & therapeutics* 2015; **41**(2): 153-66.

12. Dong J, Wang H, Zhao J, et al. Ultrasound as a diagnostic tool in detecting active Crohn's disease: a meta-analysis of prospective studies. *European radiology* 2014; **24**(1): 26-33.

13. Fraquelli M, Colli A, Casazza G, et al. Role of US in detection of Crohn disease: meta-analysis. *Radiology* 2005; **236**(1): 95-101.

14. Giles E, Barclay AR, Chippington S, Wilson DC. Systematic review: MRI enterography for assessment of small bowel involvement in paediatric Crohn's disease. *Alimentary pharmacology & therapeutics* 2013; **37**(12): 1121-31.

15. Greenup AJ, Bressler B, Rosenfeld G. Medical Imaging in Small Bowel Crohn's Disease-Computer Tomography Enterography, Magnetic Resonance Enterography, and Ultrasound: "Which One Is the Best for What?". *Inflammatory bowel diseases* 2016; **22**(5): 1246-61.

16. Horsthuis K, Bipat S, Stokkers PC, Stoker J. Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. *European radiology* 2009; **19**(6): 1450-60.

17. Liu W, Liu J, Xiao W, Luo G. A Diagnostic Accuracy Meta-analysis of CT and MRI for the Evaluation of Small Bowel Crohn Disease. *Academic radiology* 2017; **24**(10): 1216-25.

18. Qiu Y, Mao R, Chen BL, et al. Systematic review with meta-analysis: magnetic resonance enterography vs. computed tomography enterography for evaluating disease activity in small bowel Crohn's disease. *Alimentary pharmacology & therapeutics* 2014; **40**(2): 134-46.

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

20. Puylaert CA, Tielbeek JA, Bipat S, Stoker J. Grading of Crohn's disease activity using CT, MRI, US and scintigraphy: a meta-analysis. *European radiology* 2015; **25**(11): 3295-313.

21. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology* 2003; **3**: 25.

22. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of internal medicine* 2013; **158**(7): 544-54.

23. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health technology assessment (Winchester, England)* 2007; **11**(50): iii, ix-51.

24. Taylor S, Mallett S, Bhatnagar G, et al. METRIC (MREnterography or ulTRasound in Crohn's disease): a study protocol for a multicentre, non-randomised, single-arm, prospective comparison study of magnetic resonance enterography and small bowel ultrasound compared to a reference standard in those aged 16 and over. *BMC gastroenterology* 2014; **14**: 142.

25. Tolan DJ, Greenhalgh R, Zealley IA, Halligan S, Taylor SA. MR enterographic manifestations of small bowel Crohn disease. *Radiographics : a review publication of the Radiological Society of North America, Inc* 2010; **30**(2): 367-84.

26. Triester SL, Leighton JA, Leontiadis GI, et al. A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with non-stricturing small bowel Crohn's disease. *The American journal of gastroenterology* 2006; **101**(5): 954-64.

27. Castiglione F, Mainenti PP, De Palma GD, et al. Noninvasive diagnosis of small bowel Crohn's disease: direct comparison of bowel sonography and magnetic resonance enterography. *Inflammatory bowel diseases* 2013; **19**(5): 991-8.

28. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001; **49**(6): 777-82.

29. Rimola J, Ordas I, Rodriguez S, et al. Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. *Inflammatory bowel diseases* 2011; **17**(8): 1759-68.

30. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ (Clinical research ed)* 2015; **350**: h2147.

Tables

Table 1

**Demographics of final study cohort**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **New diagnosis [n (%)]****N=133** | **Relapse [n (%)]****N=151** |
|  |  |  |  |
| Sex | Male  | 69 (52) | 61 (40) |
|  | Female | 64 (48) | 90 (60) |
| Age (years) | 16-25  | 49 (37) | 46 (30) |
|  | 26-35 | 32 (24) | 36 (24) |
|  | 36-45 | 18 (14) | 28 (19) |
|  | >45 | 34 (26) | 41 (27) |
| Disease duration | <1 year | NA | 5 (3) |
|  | 1-5 years | NA | 45 (30) |
|  | 6-10 years | NA | 39 (26) |
|  | >10 years | NA | 62 (41) |
| Disease location (Montreal classification)a | L1 | NA | 56 (37) |
|  | L2 | NA | 17 (11) |
|  | L3 | NA | 74 (49) |
|  | L4 | NA | 4 (3) |
| Disease Behaviour (Montreal classification) | B1 | NA | 80 (53) |
|  | B1p | NA | 4 (3) |
|  | B2 | NA | 52 (34) |
|  | B2p | NA | 1 (1) |
|  | B3 | NA | 12 (8) |
|  | B3p | NA | 2 (1) |
| Medication | None | 62 (47) | 32 (21) |
|  | 5-ASA | 21 (16) | 26 (17) |
|  | Steroids  | 48 (36) | 28 (19) |
|  | Immunomodulators | 16 (12) | 75 (50) |
|  | Anti-TNF antibodies | 5 (4) | 42 (28) |
| Previous enteric resection | Yes | 1 (1)b | 72 (48) |

aMontreal classification not collected for new diagnosed patients

b surgical resection for inflammatory mass 1 year prior to Crohn’s disease diagnosis

**Table 2**

## **Patient characteristics: disease presence and activity-consensus reference standard**

|  |  |  |
| --- | --- | --- |
|  | New diagnosis [n (%)] | Suspected relapse [n (%)] |
|  |  |  |
| **Disease presence** |  |  |
| Small bowel disease present | 111 (83) | 122 (81) |
| Colonic disease present | 77 (58) | 52 (34) |
|  |  |  |
| Isolated small bowel disease present | 56 (42) | 85 (56) |
| Isolated colonic disease present | 22 (17) | 15 (10) |
| Both small bowel and colonic disease present | 55 (41) | 37 (25) |
| Total number patients with disease present | 133 (100) | 137 (91) |
|  |  |  |
| Median number of involved small bowel segments [median (IQR),max ] | 11 to 1), 4 | 1 (1 to 1), 3 |
| Median number of involved colonic segments [median (IQR), max ] | 1 (0 to 3), 6 | 0 (0 to 1), 6 |
|  |  |  |
| **Disease activity** |  |  |
| Small bowel disease active | 104 (94) | 105 (86) |
| Colonic disease active | 76 (99) | 50 (96) |
| Total number patients with disease active | 130 (98) | 121 (88) |
|  Criteria for activitya  |  |  |
|  *Ulceration at endoscopy* | 71 (55) | 26 (21) |
|  *CRP >8* | 47 (36) | 57 (47) |
|  *Calprotectin >250* | 41 (32) | 43 (36) |
|  *Histological evidence of activity* | 100 (77) | 36 (30) |

a Patients could meet more than one criteria for disease activity

Table 3

### Per patient sensitivity and specificity for disease presence and extent against the consensus reference standard. Both patient cohorts combined.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Sensitivity % (CI 95%) |  | Specificity % (CI 95%) |
|  | Number of disease positive patientsa | MRE | US | Difference(P value) | Number disease negative patientsa | MRE | US | Difference(P value) |
| Small bowel disease extentb | 233 | 80(72 to 86) | 70(62 to 78) | 10(1 to 18) p=0.027 | 51 | 95(85 to 98) | 81(64 to 91) | 14(1 to 27)p=0.039 |
| Small bowel disease presence | 233 | 97(91 to 99) | 92(84 to 96) | 5(1 to 9)p=0.025 | 51 | 96(86 to 99) | 84(65 to 94) | 12(0 to 25)p=0.054 |
| Colonic disease extentb | 129 | 22(14 to 32) | 17(10 to 27) | 5(-5 to 15) p=0.332 | 155 | 93(87 to 97) | 93(87 to 97) | 0(-5 to 5)p=1.000 |
| Colonic disease presence | 129 | 64(50 to 75) | 73(59 to 83) | -9(-23 to 5) p=0.202 | 155 | 96(90 to 98) | 96(90 to 98) | 0(-3 to 3)p=1.000 |
| Small bowel and colonic disease extentb | 270 | 44(36 to 54) | 29(21 to 38) | 16(6 to 25) p=0.002 | 14 | 80(42 to 96) | 61(23 to 89) | 19(-20 to 59)p=0.337 |
| Small bowel and colonic disease presencec | 270 | 78(70 to 85) | 71(62 to 79) | 7(-2 to 15) p=0.117 | 14 | 80(42 to 96) | 61(23 to 89) | 19(-20 to 59)p=0.335 |

aby consensus reference standard

b agreement with reference standard for disease presence and segmental location

c agreement with reference standard for disease presence in patients with disease in the small bowel, colon or both

Table 4

**Sensitivity and specificity for disease presence and extent against the consensus reference standard, according to patient cohort.**

|  |  |  |
| --- | --- | --- |
|  | New diagnosisN=133 | Suspected relapseN=151 |
|  | DP,DNa | Sensitivity % (CI 95%) | Specificity % (CI 95%) | DP,DNa | Sensitivity % (CI 95%) | Specificity % (CI 95%) |
|  |  | MRE | US | Difference(P value) | MRE | US | Difference(P value) |  | MRE | US | Difference(P value) | MRE | US | Difference(P value) |
| Small bowel disease extentb | 111,22 | 77(66 to 86) | 66(54 to 77) | 11(-2 to 24)p=0.099 | 98(82 to 100) | 88(64 to 97) | 10(-5 to 24) P=0.195 | 122,29 | 82(72 to 89) | 74(62 to 83) | 8(-3 to 19) P=0.141 | 92(74 to 98) | 75(50 to 90) | 17(-3 to 37) P=0.099 |
| Small bowel disease presence | 111,22 | 96(89 to 99) | 92(82 to 96) | 4(-1 to 10)p=0.148 | 99(84 to 100) | 91(65 to 98) | 8(-5 to 21) p=0.238 | 122,29 | 97(91 to 99) | 92(82 to 96) | 5(0 to 11)p=0.063 | 94(76 to 99) | 78(50 to 92) | 16(-4 to 36)p=0.111 |
| Colonic disease extentb | 77,56 | 17(9 to 30) | 9(4 to 19) | 8(-2 to 19) p=0.115 | 93(82 to 98) | 92(80 to 97) | 1(-7 to 10) p=0.752 | 52,99 | 31(17 to 48) | 33(19 to 51) | -2(-22 to 17) p=0.817 | 93(85 to 97) | 94(86 to 97) | -1(-7 to 5) p=0.804 |
| Colonic disease presence | 77,56 | 47(31 to 64) | 67(49 to 81) | -20(-39 to -1)p=0.043 | 96(86 to 99) | 95(84 to 98) | 1(-5 to 7) p=0.738 | 52,99 | 84(67 to 94) | 80(61 to 91) | 4(-11 to 20) p=0.589 | 96(88 to 98) | 95(89 to 99) | -1(-5 to 4) p=0.791 |
| Small bowel and colonic disease extentb | 133,0 | 33(22 to 46) | 20(12 to 30) | 13(1 to 26) p=0.029 | N/A | N/A | N/A | 137,14 | 56(43 to 68) | 40(28 to 52) | 16(2 to 31) p=0.027 | 80(42 to 96) | 61(24 to 88) | 19(-20 to 59) p=0.339 |
| Small bowel and colonic disease presencec | 133,0 | 65(52 to 76) | 66(53 to 77) | -1(-15 to 13) p=0.877 | N/A | N/A | N/A | 137,14 | 88(79 to 93) | 76(64 to 85) | 12(2 to 22) p=0.018 | 80(42 to 96) | 61(23 to 89) | 19(-20 to 59) p=0.336 |

a Disease positive (DP), Disease negative (DN) by consensus reference standard

bagreement with reference standard for disease presence and segmental location

c agreement with reference standard for disease presence in patients with disease in the small bowel, colon or both

**Table 5**

**Per patient sensitivity and specificity for the presence of active disease versus the consensus reference standard. Both patient cohorts combined.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Sensitivity % (CI 95%)** |  | **Specificity % (CI 95%)** |
|  | **Number of patients with active disease ˆ** | **MRE** | **US** | **Difference****(P value)** | **Number of patients with inactive disease ˆ** | **MRE** | **US** | **Difference****(P value)** |
| Active small bowel disease | 209 | 96(92 to 99) | 90(82 to 95) | 6(2 to 11)p=0.010 | 75 | 83(68 to 92) | 77(60 to 88) | 6(-8 to 20)p= 0.376 |
| Active Colonic disease | 126 | 63(48 to 76) | 66(51 to 79) | -3(-18 to 13)p=0.735 | 158 | 97(91 to 99) | 98(94 to 99) | -1(-4 to 1)p=0.304 |
| Active Small bowel and colonic disease\*\*  | 251 | 77(68 to 85) | 66(56 to 75) | 11(1 to 21)p=0.024 | 33 | 28(10 to 56) | 28(10 to 56) | 0 (-26 to 26)p=1.000 |

ˆ by consensus reference standard

\* agreement with reference standard for disease active

\*\* agreement with reference standard for active disease presence in patients with disease in the small bowel, colon or both

**Figures**

**183 Excluded**

58 declined participation

28 failed to respond to invitation

22 non-Crohn’s diagnosis

20 Unable to complete MRE and/or US in timely fashion

13 Not meet study eligibility criteria (relapse cohort) based on low CRP

8 contraindication to MRE

7 not able give informed consent

5 Previous recruitment or declined approach

4 moved/lived far away

4 Proceeded straight to surgery prior to colonoscopy (new diagnosis cohort)

2 Newly diagnosed more than 3 months previously

2 under 16 years old

10 unknown

**Reference standard**

133 Consensus panel at 6 months

**Reference standard**

151 Consensus panel at 6 months

**Index tests**

151 MRE and US

**Index tests**

133 MRE and US

**51 Withdrawals**

31 final diagnoses other than Crohn’s

5 did not undergo MRE

3 did not undergo US

2 did not undergo MRE or US

3 withdrew consent

3 no longer wished to participate in follow up

2 lost to follow up

2 Underwent surgery without colonoscopy

133 Newly diagnosed participants

151 Suspected relapse participants

**284 Included participants**

335 Recruited participants

518 Screened participants

**Figure 1. Patient flow diagram**

**Figure 2 MRE and US sensitivity and specificity for small bowel and colonic disease extent and presence against the consensus reference standard**

**Appendix**

**Table 1. Recruitment sites and recruitment totals**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Recruitment site** | **Total patients screened** | **Total patients recruited****(new diagnosis)** **[n (%) of total]** | **Total patients recruited (suspected relapse)** **[n (%) of total]** | **Total patients recruited****(both patient cohorts** **[n (%) of total]** | **Total patients withdrawn****[n (%) of total]** | **Total patients in final study cohort** **(new diagnosis)****(% of total)** | **Total patients in final study cohort (suspected relapse)****(% of total)** | **Total patients in final study cohort****(both cohorts)****(% of total)** |
| University College London Hospital | 177 | 66 (39) | 69 (41) | 135 (40) | 19 (36) | 52 (39) | 64 (42) | 116 (41) |
| St Marks Hospital, Harrow | 78 | 8 (5) | 16 (10) | 24 (7) | 4 (8) | 5 (4) | 15 (10) | 20 (6) |
| Royal Free Hospital, London | 7 | 1 (1) | 2 (1) | 3 (1) | 1 (2) | 1 (1) | 1 (1) | 2 (1) |
| Queen Alexandra Hospital, Portsmouth | 66 | 32 (19) | 27 (16) | 59 (18) | 9 (18) | 28 (20) | 22 (15) | 50 (18) |
| Leeds General Infirmary, Leeds | 69 | 29 (17) | 22 (13) | 51 (15) | 4 (8) | 27 (20) | 20 (13) | 47 (17) |
| Ninewells Hospital, Dundee | 71 | 11 (6) | 15 (9) | 26 (8) | 3 (6) | 9 (7) | 14 (9) | 23 (7) |
| Radcliffe Hospital, Oxford | 39 | 15 (9) | 11 (7) | 26 (8) | 7 (14) | 9 (7) | 10 (7) | 19 (7) |
| St Georges Hospital, London | 11 | 6 (4) | 5 (3) | 11 (3) | 4 (8) | 2 (2) | 5 (3) | 7 (3) |
| TOTAL | 518 | 168 | 167 | 335 | 51 | 133 | 151 | 284 |

**Table 2. MRE sequence protocol**

|  |  |
| --- | --- |
| **Minimum**  | **Optional** |
| Coronal steady state free precession gradient echo (SSFP GE) sequences without fat saturation | Axial steady state free precession gradient echo (SSFP GE) sequences without fat saturation |
| Hyoscine butylbromide 20mg IV | Axial fast spin echo (FSE) T2W sequence with fat saturation |
| Axial and coronal fast spin echo (FSE) T2W sequences without fat saturation | Axial contrast-enhanced coronal T1W sequences with fat saturation (60-70 sec post injection) |
| Coronal coronal fast spin echo (FSE) T2W sequence with fat saturation | Coronal steady state free precession gradient echo (SSFP GE) dynamic Motility sequences  |
| Axial diffusion weighted images (b values 50 and 600) |  |
| Non-enhanced coronal T1W sequence with fat saturation followed by contrast-enhanced coronal T1W sequences with fat saturation (60-70 sec post injection) |  |

**Table 3. Investigations and results available to the consensus panels**

|  |  |  |
| --- | --- | --- |
|  | New diagnosis [n (%)]N=133 | Relapse [n (%)]N=151 |
|  |  |  |
| MR enterography | 133 (100) | 151 (100) |
| US | 133 (100) | 151 (100) |
| Colonoscopy | 123 (92) | 66 (44) |
| Gastroscopy | 11 (8) | 6 (4) |
| Sigmoidoscopy | 5 (4) | 12 (8) |
| Capsule endoscopy | 10 (8) | 8 (5) |
| CT enterography | 4 (3) | 9 (6) |
| CT abdo pelvis | 21 (16) | 13 (9) |
| MR enteroclysis | 4 (3) | 6 (4) |
| MRI abdomen and/or pelvis | 5 (4) | 8 (5) |
| Barium FT | 8 (6) | 19 (13) |
| Barium enteroclysis | 3 (2) | 7 (5) |
| Hydrosonography | 28 (21) | 37 (25) |
| White cell scan | 0 (0) | 0 (0) |
| CRP (baseline) | 127 (95) | 145 (96) |
| HBI (baseline | 124 (93) | 142 (94) |
| Calprotectin (baseline) | 87 (65) | 89 (59) |
| CRP (10-20 weeks) | 108 (81) | 120 (79) |
| HBI (10-20 weeks) | 71 (53) | 77 (51) |
| Calprotectin (10-20 weeks) | 53 (40) | 65 (43) |
| Surgical resection specimen (post recruitment) | 1 (1) | 2 (1) |
| Other | 8 (6) | 20 (13) |

**Table 4. Per segment sensitivity and specificity for disease presence against the consensus reference standard.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Sensitivity % (CI 95%)** |  | **Specificity % (CI 95%)** |
|  | **Number of disease positive segmentsa** | **MRE** | **US** | **Difference****(P value)** | **Number of negative segmentsa** | **MRE** | **US** | **Difference****(P value)** |
| **Small bowel segments** |  |  |  |  |  |  |  |  |
| Duodenumb | 8 | 25(7 to 59) | 25(7 to 59) | 0(-13 to 13) p=1.000 | 276 | 100(99 to 100) | 99(97 to 100) | 1(0 to 3) p=250 |
| Jejunum | 13 | 71(38 to 91) | 63(32 to 86) | 8(-29 to 46) p=0.664 | 271 | 99(93 to 100) | 99(94 to 100) | 0 (-2 to 1)p=0.741 |
| Ileum | 38 | 84(67 to 93) | 56(38 to 73) | 28(8 to 49)p=0.008 | 246 | 93(87 to 97) | 93(87 to 96) | 0(-4 to 4)p=0.871 |
| Terminal ileum | 217 | 96(91 to 99) | 92(84 to 96) | 4(0 to 8)p=0.051 | 66 | 97(90 to 99) | 93(81 to 98) | 4(-2 to 10)p=0.197 |
| **Colonic segmentsc** |  |  |  |  |  |  |  |  |
| -Caecum | 78 | 46 (35 to 57) | 46 (35 to 57) | 0 (-12 to 12) p=1.000 | 147 | 96 (92 to 99) | 90 (85 to 94) | 6 ( 0 to 12) p=0.036 |
| -Ascending | 67 | 49 (38 to 61) | 49 (38 to 61) | 0 (-10 to 10) p=1.000 | 200 | 96 (93 to 98) | 92 (88 to 95) | 4 (0 to 8) p=0.058 |
| -Transverse | 61 | 46 (34 to 58) | 44 (32 to 57) | 2 (-12 to 15) p=0.809 | 218 | 97 (93 to 98) | 95 (91 to 97) | 2 (-1 to 5) p=0.130 |
| -Descending | 59 | 53 (40 to 65) | 41 (29 to 54) | 12 (-1 to 24) p=0.063 | 221 | 98 (95 to 99) | 95 (91 to 97) | 3 (0 to 6) p=0.033 |
| -Sigmoid | 76 | 46 (35 to 57) | 43 (33 to 55) | 3 (-11 to 16) p=0.695 | 203 | 96 (92 to 98) | 93 (89 to 96) | 3 (-1 to 7) p=0.179 |
| -Rectum | 54 | 44 (32 to 58) | 22 (13 to 35) | 22 (9 to 35) p=0.001 | 228 | 97 (94 to 99) | 93 (89 to 96) | 4 (0 to 7) p=0.072 |

aby consensus reference standard

b McNemar’s test due to small number of patients with disease

cAnalysis for individual colonic segments uses a population average approach

**Table 5. Per patient sensitivity and specificity for the presence of active disease against the consensus reference standard, according to patient cohort.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | New diagnosisN=133 |  | Suspected relapseN=151 |
|  |  | Sensitivity % (CI 95%) | Specificity % (CI 95%) |  | Sensitivity % (CI 95%) | Specificity % (CI 95%) |
|  | DA,DIa | MRE | US | Difference(P value) | MRE | US | Difference(P value) | DA,DIa | MRE | US | Difference(P value) | MRE | US | Difference(P value) |
| Active Small bowel diseaseb  | 104,29 | 96(90 to 99) | 90(79 to 96) | 6(0 to 13)p=0.056 | 90(68 to 98) | 83(56 to 95) | 7(-11 to 25)p=0.453 | 105,46 | 96(90 to 99) | 90(79 to 96) | 6(0 to 13) p=0.056 | 79(57 to 91) | 73(51 to 88) | 6(-14 to 25)p=0.584 |
| Active Colonic diseaseb | 76,57 | 48(30 to 66) | 55(36 to 72) | -7(-28 to 14)p=0.522 | 96(88 to 99) | 97(90 to 99) | -1(-5 to 4) p=0.720 | 50,101 | 83(63 to 93) | 81(59 to 92) | 2(-14 to 19)p=0.779 | 96(89 to 99) | 98(93 to 99) | -2(-5 to 2)p=0.309 |
| Active Small bowel and colonic diseasec | 130,3 | 55(47 to 64) | 58(50 to 67) | 3(-7 to 13)p=0.627 | 0(0 to 56) | 0(0 to 56) | 0(-33 to 33) p=1.000 | 121,30 | 64(56 to 72) | 77(69 to 83) | 12(3 to 22) p=0.014 | 40(25 to 58) | 40(25 to 58) | 0(-22 to 22) p=1.000 |

aDisease active (DA), disease inactive (DI) by consensus reference standard

bagreement with reference standard for disease activity

cagreement with reference standard for disease activity to identify patient as having disease in small bowel, colon or both

**Table 6. Sensitivity and specificity for terminal ileal and colonic disease presence and against a ileo colonoscopy reference. Both patient cohorts combined**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Sensitivity % (CI 95%) |  | Specificity % (CI 95%) |
|  | Number of disease positive patientsa | MRE | US | Difference(P value) | Number of disease negative patientsa | MRE | US | Difference(P value) |
| Terminal ileum | 105 | 97 (91 to 99) | 91 (79 to 97) | 6 (-1 to 12) p=0.091 | 81 | 41 (21 to 64) | 33 (15 to 57) | 8 (-14 to 30) p=0.474 |
| Colonic disease extentb | 109 | 3(1 to 11) | 2(0 to 8) | 1(-2 to 4)p=0.429 | 77 | 94(81 to 98) | 89(73 to 96) | 5(-3 to 14)p=0.240 |
| Colonic disease presence | 109 | 41(26 to 58) | 49(33 to 65) | -8(-26 to 9)p=0.368 | 77 | 95(85 to 98) | 90(76 to 96) | 5(-3 to 13)p=0.233 |
| Colonic segmentsc |  |  |  |  |  |  |  |  |
| -Caecum | 73 | 22 (14 to 33) | 25 (16 to 36) | -3 (-14 to 9) p=0.638 | 101 | 72 (63 to 80) | 65 (56 to 74) | 7 (0 to 13) p=0.043 |
| -Ascending | 62 | 26 (16 to 38) | 23 (14 to 35) | 3 (-6 to 12) p=0.479 | 121 | 88 (80 to 92) | 81 (73 to 87) | 7 (0 to 13) p=0.043 |
| -Transverse | 62 | 24 (15 to 37) | 24 (15 to 37) | 0 (-9 to 9) p=1.000 | 121 | 92 (86 to 96) | 90 (84 to 94) | 2 (-2 to 6) p=0.256 |
| -Descending | 58 | 27 (18 to 40) | 24 (15 to 37) | 3 (-6 to 13) p=0.479 | 128 | 95 (90 to 98) | 93 (87 to 96) | 2 (-1 to 6) p=0.178 |
| -Sigmoid | 74 | 24 (16 to 35) | 28 (19 to 40) | -4 (-17 to 9) p=0.532 | 111 | 94 (87 to 97) | 94 (87 to 97) | 0 (-6 to 6) p=1.000 |
| -Rectum | 61 | 26 (17 to 39) | 13 (7 to 24) | 13 (2 to 25) p=0.027 | 125 | 97 (92 to 99) | 94 (88 to 97) | 3 (-2 to 8) p=0.204 |

aby ileo colonoscopy reference standard

bagreement with reference standard for disease presence and segmental location

cAnalysis for individual colonic segments uses a population average approach to compare imaging accuracy for individual colon segments

**Author contributions**

Substantial contributions to the conception or design of the work -all authors ((SAT, SMa, GB, RBC, SB, AG, JH, AHa, AHi, IJ, SM, AM, SMo, CM, AP, RP, SP, LQ, MRJ, ZS, AS, DT, ST, AW, PW, IZ, SH)

Literature search-SAT, GB

Data collection-GB, RBC, SMo, AM, SAT, ZS

Data interpretation-SAT, LM, SMa, SH

Clinical studies and patient recruitment-SAT, GB, RBC, SB, AG, JH, AHa, AHi, SMo, CM, AP, RP,SP,MRJ,AS,DT,ST,AW,PW,IZ,SH

Statistical analysis-LQ, SMa

PPI representative-IJ

Initial manuscript draft-SAT, SM, SH

Drafting the work or revising it critically for important intellectual content-all authors (SAT, SMa, GB, RBC, SB, AG, JH, AHa, AHi, IJ, SM, AM, SMo, CM, AP, RP, SP, LQ, MRJ, ZS, AS, DT, ST, AW, PW, IZ, SH)

Final approval of the version to be published-all authors (SAT, SMa, GB, RBC, SB, AG, JH, AHa, AHi, IJ, SM, AM, SMo, CM, AP, RP, SP, LQ, MRJ, ZS, AS, DT, ST, AW, PW, IZ, SH)

Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved-all authors- (SAT, SMa, GB, RBC, SB, AG, JH, AHa, AHi, IJ, SM, AM, SMo, CM, AP, RP, SP, LQ, MRJ, AS, DT, ST, AW, PW, IZ, SH)

**Declarations of interest**

SAT is a remunerated research consultant to Robarts plc

XX

**Funding source acknowledgement**

This project was funded by the National Institute of Health Research health technology assessment NIHR HTA programme (project number 11/23/01) and will be published in full in Health Technology Assessment. The project is supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Department of Health disclaimer:

This report presents independent research commissioned by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC or the HTA programme or the Department of Health. The views and opinions expressed by the interviewees in this publication are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, MRC, CCF, NETSCC or the HTA programme or the Department of Health

The corresponding author had full access to all data and final responsibility for the decision to submit for publication