

BIROn - Birkbeck Institutional Research Online

Taylor, S. and Mallett, S. and Bhatnagar, G. and Baldwin-Cleland, R. and Bloom, S. and Gupta, A. and Hamlin, P. and Hart, A. and Higginson, A. and Jacobs, I. and McCartney, S. and Miles, Anne and Murray, C. and Plumb, A. and Pollok, R. and Punwani, S. and Quinn, L. and Rodriguez-Justo, M. and Shabir, Z. and Slater, A. and Tolan, D. and Travis, S. and Windsor, A. and Wylie, P. and Zealley, I. and Halligan, S. (2018) Diagnostic accuracy of magnetic resonance enterography and small bowel ultrasound for the extent and activity of newly diagnosed and relapsed Crohn's disease (METRIC): a multicentre trial. *Lancet Gastroenterology and Hepatology* 3 (8), pp. 548-558. ISSN 2468-1253.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/22280/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>

or alternatively

contact lib-eprints@bbk.ac.uk.

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 Trial

Professor Stuart A Taylor¹ FRCR, Susan Mallett DPhil², Gauraang Bhatnagar FRCR¹, Rachel Baldwin-Cleland MSc³, Stuart Bloom FRCP⁴, Arun Gupta FRCR³, Peter J Hamlin PhD⁵, Professor Ailsa L Hart FRCP⁶, Antony Higginson FRCR⁷, Ilan Jacobs⁸, Sara McCartney FRCP⁴, Anne Miles PhD⁹, Charles D Murray PhD¹⁰, Andrew A Plumb FRCR¹, Richard C Pollok FRCP¹¹, Shonit Punwani FRCR¹, Laura Quinn MSc², Manuel Rodriguez-Justo FRCPath¹², Zainib Shabir¹³, Andrew Slater FRCR¹⁴, Damian Tolan FRCR¹⁵, Professor Simon Travis DPhil¹⁶, Alastair Windsor FRCS¹⁷, Peter Wylie FRCR¹⁸, Ian Zealley FRCR¹⁹, Professor Steve Halligan FMedSci¹, 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 Research, NIHR Birmingham Biomedical Research Centre, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
3. Intestinal Imaging Centre, St Mark's Hospital, LNWUH NHS Trust, Harrow, UK
4. Department of Gastroenterology, University College Hospital, 235 Euston Road, London, NW1 2BU, UK
5. Department of Gastroenterology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Beckett Street, Leeds, LS9 7TF, UK
6. Inflammatory bowel disease Unit, St Mark's Hospital, LNWUH NHS Trust, Harrow, UK
7. Department of Radiology, Portsmouth Hospitals NHS Trust, Portsmouth, UK
8. c/o Centre for Medical Imaging, University College London, Charles Bell House, 43-45 Foley Street, W1W 7TS
9. Department of Psychological Sciences, Birkbeck University of London, Malet Street, London, WC1E 7HX 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, Oxford University Hospitals NHS Trust, Oxford UK
15. Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS 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

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. 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 enterography (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 in Jan 2018 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". The primary literature was retrieved but emphasis was placed on meta-analyses and systematic reviews using appropriate filters. We found a number of meta-analyses, which in general suggest MRE and US have comparable sensitivity for detection and activity assessment of small bowel Crohn's disease. However, the primary literature has limitations. Most studies are small single centre explanatory trials recruiting fewer 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. For example, in a recent meta-analysis, just one of 33 included studies compared MRE and US directly in the same patients. Many also rate poorly on the Quality Assessment of Diagnostic Accuracy Studies tool.

Added value of this study

This is the largest prospective multicenter trial 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

Both US and MRE achieve high diagnostic accuracy for the extent and activity of small bowel Crohn's disease in both newly diagnosed patients and those suffering relapse. Whilst both tests are valid first-line investigations, MRE is generally the preferred radiological investigation when available because its sensitivity and specificity exceed US significantly when tested in a prospective multi-center trial setting. Future research should investigate the role of cross sectional imaging in patients with nonspecific abdominal symptoms without an established diagnosis of Crohn's disease, and the complimentary role of MRE and US in targeted follow up of Crohn's disease patients with an established disease phenotype.

Abstract

Background

Magnetic resonance enterography (MRE) and ultrasound (US) are used to image Crohn's disease, but comparative accuracy for disease extent and activity is not known with certainty. We undertook a prospective multicentre cohort trial to address this

Methods

We 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. Consecutive patients underwent MRE and US in addition to standard 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. The trial is completed (ISRCTN03982913).

Findings

284 patients completed the trial (133 new diagnosis, 151 relapse). MRE sensitivity (n=233) 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; p=0.027), and 5% (1 to 9), difference respectively. MRE specificity for SB disease extent (95% [85 to 98]) was 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 95]), difference 6% (2 to 11). Overall, there were no significant accuracy differences for colonic disease presence. Accuracy in newly diagnosed and relapse patients was similar, although US had significantly greater sensitivity for colonic disease than MRE in newly diagnosed patients (67% [49 to 81] vs. 47% [31 to 64]), difference 20% (1 to 39). There were no serious adverse events.

Interpretation

MRE has higher diagnostic accuracy for the extent and activity of SB Crohn's disease than US when tested in a prospective multi centre cohort trial setting.

Funding

NIHR Health Technology Assessment

Introduction

Small bowel imaging is fundamental for comprehensive phenotyping of Crohn's disease and essential to direct therapeutic strategy.¹ Barium fluoroscopy has long been the bedrock of small bowel investigation, providing detailed mucosal evaluation.² However, in recent years enthusiasm has dwindled, and it is replaced increasingly by cross sectional imaging, namely computed tomography enterography (CTE), magnetic resonance imaging enterography (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 the various cross-sectional imaging technologies 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 irradiating generally young patients who require repeat imaging.⁴ Enteric US is longer established,⁵ 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 innovation,⁸ 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 have similar accuracy for diagnosing and staging Crohn's disease,^{6,9-20} the primary literature is of questionable quality. The 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 studies.²² For example in their recent meta-analysis, Greenup and al found just one of 33 included studies compared MRE and US directly in the same patients.¹⁵ 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 elusive.²³

In order to redress this, we conducted a prospective multicenter trial 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 patient 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

Trial Design and participants

The METRIC study is a multicentre, prospective cohort trial 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, and achieved ethics committee approval in September 2013 (13/SC/0394). The trial was supervised by University College London's Comprehensive Clinical Trials Unit and overseen by independent Data Monitoring and Trial Steering Committees. All patients recruited gave written informed consent. The full trial protocol has been published,²⁴ and can be found online (<http://www.ucl.ac.uk/cctu/research-areas/gastroenterology/metric>). The trial is registered with the International Standard Randomised Controlled Trial registry number ISRCTN03982913.

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 p2). All sites had an established inflammatory bowel disease service 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. Eligible patients for both arms were aged ≥ 16 . Patients were ineligible if pregnant or if they had contraindications 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 from consecutive unselected eligible patients. A screening log detailed all approached patients, and reasons for non-participation, if applicable. Patient demographics and clinical data were collated (for example age, sex, Montreal classification [relapse subgroup only], disease/symptom duration, medication and surgical history).

Procedures

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 p3). US was performed by radiologists or sonographers using standard platforms and both curvilinear and high-resolution probes, without oral or intravenous contrast agents (appendix p4).

Across all sites, 28 practitioners interpreted the MRE and US studies (27 radiologists and 1 sonographer). Eight radiologists interpreted MRE only, 3 performed and interpreted US only and 16 performed and interpreted US and interpreted MRE. 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 policies and was performing enteric US routinely and had 20 years of experience. Radiologists interpreting MRE had a median 10 years of experience (Interquartile Range [IQR] 6 to 11) and practitioners interpreting US had a median 8 years of experience (IQR 4 to 11). The median number examinations performed per month at each recruitment site during the conduct of the trial was 30 (IQR 20 to 45) for MRE and 25 (IQR 12 to 40) for US. Before trial 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 clinical requests). 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 criteria (appendix p5).^{6,16,25} The segmental location of any disease was also recorded, using standard definitions;²⁴ 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 clinical report was then generated as per usual clinical practice.

The 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 were 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.

Patients were asked if they found MRE and US acceptable and which test attribute they considered to be the most important.

We used the construct reference standard paradigm (panel diagnosis) incorporating the concept of clinical test validation, i.e. whether test results are meaningful in practice.²³ 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 if required and a member of the trial management group attended to ensure uniformity of process. For each 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. All panel decisions were recorded as present/absent, active/inactive with no option of an indeterminate outcome. 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 >250mcg/g and/or (iv) histopathological evidence of acute inflammation based on biopsy or surgery within 2 months of trial 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 (terminal ileum, ileum, jejunum or duodenum). 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.

All outcomes were pre-specified except accuracy for individual small bowel segments (duodenum, jejunum, ileum), accuracy for disease presence and extent in the colon, and per patient disease activity (small bowel and colonic disease combined), which were exploratory.

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 for small bowel disease extent between MRE (83% based on a sensitivity of 93% and 90% for disease presence and

location respectively) and US (73% based on a sensitivity of 88% and 83% for disease presence and location respectively), assuming 68% positivity for both tests and using methods for comparative studies.^{24,26} 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). The trial was not powered to detect differences between the cohorts, or between bowel segments.

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 (neo-terminal ileum was considered as the "terminal ileum").

Direct comparison of sensitivity and specificity differences between MRE and US were calculated using bivariate multilevel patient specific (conditional) random effects models, from paired data using meqrlogit in STATA 14.2 [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. Analysis by segment used a population averaged random effects model (using logit including robust standard errors). Statistical significance was based on 95% CI. There were no missing data for per patient diagnosis of disease presence or disease extent, for the reference standard, MRE or US.

Data sharing

Contact corresponding author

Role of Funding source

The primary funder (the National Institute for Health Research) stipulated a diagnostic accuracy trial using a cohort design but were not 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, median age 30 years [IQR 24 to 41]); 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) including 154 (54%) 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 (44%) had colonic disease, which was active in 126 (98%). Twenty-one had enteric fistulae, and 7 had intraabdominal abscess. No serious adverse events were reported.

In 53 patients (24 new diagnosis, 29 relapse) MRE and US 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 p6.

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; $p=0.027$), which was statistically significant (table 3, figure 2 appendix p7). 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), (table 3, 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 supplementary appendix p8. Although the trial 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). 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 tests, 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.

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 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 p9).

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 p10). Sensitivity for colonic disease presence was modest for both MRE and US (41% [26 to 58] and 49% (33 to 65)), and not statistically different.

Of responding patients, 128/145 (88%) and 144/146 (99%) rated MRE and US as acceptable, respectively. Diagnostic accuracy was rated as the most important test attribute.

Discussion

At the time of writing, METRIC is the largest prospective multicenter trial directly comparing diagnostic accuracy of MRE and US for the presence, extent and activity of Crohn's disease in the same patients. We found both MRE and US highly accurate for detecting small bowel Crohn's disease, achieving 97% 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 questioned.²⁷ Conversely, against a rigorous ileo-colonoscopy reference standard, we found MRE and US achieved 97% and 91% sensitivity for terminal ileal disease, strongly supporting their transition to first line investigations, 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 (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-analyses.^{6,9-20} 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 many 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.¹⁵ 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.²⁸ 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 elusive.²³ 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 that when our analysis was limited to an ileo-colonoscopy 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 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 stricturing and penetrating disease increases with time.²⁹ 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 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 (67% vs 47%). Optimised colonic evaluation using MRE requires purgation and fluid distension,³⁰ which are both omitted from routine MRE protocols, whereas, in general, US relies on evaluating the manually compressed uncleaned colon wall. Accuracy for both techniques in the colon still falls short of colonoscopy, and in the case of MRE is somewhat lower than previously reported.^{31,32} By way of explanation, ileo-colonoscopy and histopathology results were available to the consensus reference panel for most patients (particularly those newly diagnosed), and are exquisitely sensitive for early mucosal disease, beyond the resolution of cross sectional imaging. Our outcomes were dependent on disease presence regardless of severity. Data from existing single center explanatory studies either use cohorts enriched with more advanced colonic disease³², or report sensitivity for deep rather than superficial mucosal disease.³¹

Most patients found MRE and US acceptable, although more so for US than MRE. This is perhaps expected given the different attributes of the two tests. Patients however rated diagnostic accuracy as the most important test attribute, consistent with previous work,³³ suggesting patients will tolerate greater discomfort for improved test performance.

Our trial does have limitations. METRIC was conceived as a large pragmatic trial³⁴ 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. We do however acknowledge that specialist practitioners working in high volume practices may achieve sensitivities in excess of our findings. Imaging was interpreted according to local clinical practice in order to mirror “real world” procedures within the NHS, 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. We cannot however exclude occasional inadvertent unblinding of reporting practitioners. Recruited patients were representative of those undergoing MRE and US in daily practice, although we did exclude pregnant women, patients undergoing routine therapeutic response assessment, and patients with contraindications to MRI. Our results are therefore highly likely to extrapolate across the NHS and similar health care settings. The prevalence of active disease was predictably high given our recruited patient cohorts. The reported high specificity of MRE and US should therefore be viewed in this context.

We did not standardise the third small bowel investigation whenever discrepancy between MRE and US arose, and this was left to the discretion of the recruitment site. Direct mucosal visualisation is possible using push enteroscopy³⁵ or capsule endoscopy³⁶ but to insist on such investigations was not practicable in the setting of a pragmatic multicenter trial, given their cost, relatively limited availability and likely negative impact on patient compliance and safety. Push enteroscopy for example is a highly invasive and specialised investigation, and attracts a small but well documented risk of major complications such as perforation.³⁷ Similarly the risk of capsule retention is around 8% in those with known Crohn’s disease,³⁸ and specificity is questioned.³⁹ We also considered that the invasive nature of capsule/enteroscopy would result in considerable spectrum bias relating to differences between patients who would and would not agree to consent (even were they available and affordable).

To reduce incorporation bias from MRE or US, we required at least one independent biochemical, endoscopic or histological marker of disease activity before a patient could be “labeled” with active small bowel or colonic disease. Biochemical markers such as calprotectin and CRP provide evidence at the patient level, but the reference standard consensus panel also had access to a range of additional clinical material when making their decision, including endoscopy and a range of small bowel imaging investigations.

There is data suggesting that diagnostic accuracy of US can be improved with an oral contrast load (small intestine contrast enhanced US [SICUS]), particularly for luminal stenosis, and intravenous contrast enhanced US (CEUS) may have utility for assessing disease activity.⁴⁰ However, neither SICUS and CEUS have disseminated as first line investigations outside specialist units, and if used, are often employed as problem solving tools.^{40,41} Standard US is overwhelmingly the most commonly employed technique in routine clinical practice. Future prospective research could consider inclusion of both SICUS and CEUS in trial design.

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

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 and both tests are valid first-line investigations. In a national health service setting, MRE is generally the preferred radiological investigation when available because its sensitivity and specificity exceed US significantly.

Author contributions

All authors (SAT, SMa, GB, RBC, SB, AG, PJH, AHa, AHi, IJ, SM, AM, CM, AP, RP, SP, LQ, MRJ, ZS, AS, DT, ST, AW, PW, IZ, SH) made substantial contributions to the conception or design of the work, drafted the work or revised it critically for important intellectual content, agree to be accountable for all aspects of the work, and gave final approval of the version to be published

SAT and GB contributed to the literature search, data collection, clinical studies and patient recruitment. RBC, AM and ZS contributed to data collection. SAT, LM, SMa, SH contributed to data interpretation. RBC, SB, AG, PJH, AHa, AHi, CM, AP, RP, SP, MRJ, ZS, AS, DT, ST, AW, PW, IZ, and contributed to clinical studies and patient recruitment. LQ and SMa contributed to the statistical analysis. IJ acted as public/ patient representative. SAT, SMa, SH performed the initial manuscript draft. SAT is the study guarantor.

Declarations of interests

SAT, SH, ST, SMa, SB, DT, PJH, AG, AHa, AS, SM, IZ, AHi, AW, IJ, MRJ, SP, PW, CM report research grant support from the funder (NIHR HTA) to undertake the study. SAT reports personal fees from Robarts plc. ST reports directorships of charities IBD 2020 (UK 09762150), Crohn's Colitis Cure (ABN 85 154 588 717), Truelove Foundation (UK 11056711), Grants/Research Support from AbbVie, IOIBD, Lilly, UCB, Vifor, and Norman Collisson Foundation, Consulting Fees from AbbVie, Centocor, , Schering-Plough, Bristol-Myers Squibb, Chemocentryx, Cosmo, Elan Pharma Inc, Genentech, Giuliani SpA, Merck and co., Takeda, Otsuka Pharmaceuticals, PDL Biopharma, Pfizer, Shire, Glenmark Pharma, Synthon Pharma, Glenmark Pharma, NPS, Lilly, Warner Chilcott, Proximagen, VHSquared, Topivert, Ferring, Celgene, GlaxoSmithKline, Takeda, Amgen, Biogen, Enterome, Immunocore, Immunometabolism, Bioclinica, Boehringer Ingelheim, Gilead, Grunenthal, Janssen, Novartis, Celgene, Receptos, PharmOlam, Sigmoid Pharma, Theravance, expert testimony for santarus Inc, Cosmo Technologies and Tillotts Pharma, grants from Ferring, AbbVie, Schering-Plough, Merck Sharpe & Dhome (MSD), Proctet and Gamble, Warner Chilcott, International Organisation of IBD, Lilly,UCB,Vifor,Norman Collisson Foundation, lecture fees for AbbVie, Schering Ploufh, Centocor, Merck and co., Given Imaging, UCB Pharma, Ferring Pharmaceuticals, Tillotts Labs, Shire, Sanofi Aventis, Vifor, Takeda, Amgen, Biogen, payment for manuscript preparation for Ferring; Royalties for Wilry

Blackwell, Elsevier, Oxford university press, payment for development of educational presentations for Abbott Labs, Procter and Gamble and Warner Chilcott, travel/accommodation expenses from AbbVie, UEG, Shire, research nurse salary support from Schering Plough, Procter and Gamble, MSD, Vifor, and unrestricted education grants from Abbott laboratories, Procter and Gamble, PDL BioPharma, Takeda and ICHOM. AHa reports personal fees from AbbVie, Atlantic, Bristol-Myers Squibb, Celltrion, Falk, Ferring, Janssen, MSD, Napp Pharmaceuticals, Pfizer, Pharmacosmos, Shire and Takeda, and non-financial support from Genentech. AW- reports grants from Allergan, personal fees from Allergan, Cook, and grants and personal fees from BARD. AP reports grants from NIHR HTA, grants from NIHR Fellowships programme, and personal fees from Acelyty, Actavis, Dr Falk, Janssen-Cilag, and Takeda. IJ reports share ownership in General Electric and sells GE equipment. CM reports personal fees from Abbvie, MSD, and Janssen. AHi reports personal fees from Toshiba. SH reports non-financial support from iCAD. All other authors declared no conflicts of interest

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 and NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham.

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

Oversight committee Acknowledgments

The trial steering committee and independent data monitoring committee met at least annually and included

TSC: Vicky Goh (chair), Andrea Marshall (statistician), Ilan Jacobs (patient representative) and James Lindsay (subject expert)

IDMC: Tim Orchard (chair), Doh-Mu Koh (subject expert), Chris Rogers (statistician)

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**: 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**: 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**: 556-85.
4. Chatu S, Subramanian V, Pollok RC. Meta-analysis: diagnostic medical radiation exposure in inflammatory bowel disease. *Alimentary pharmacology & therapeutics* 2012; **35**: 529-39.
5. Sonnenberg A, Erckenbrecht J, Peter P, Niederau C. Detection of Crohn's disease by ultrasound. *Gastroenterology* 1982; **83**: 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**: 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**: 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**: 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**: 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**: 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**: 26-33.
13. Fraquelli M, Colli A, Casazza G, et al. Role of US in detection of Crohn disease: meta-analysis. *Radiology* 2005; **236**: 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**: 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**: 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**: 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**: 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**: 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**: 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**: 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**: 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* 2007; **11**: 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**: 367-84.
 26. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Statistics in medicine* 2002; **21**: 835-52.
 27. 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**: 954-64.
 28. 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**: 991-8.
 29. 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**: 777-82.
 30. 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**: 1759-68.
 31. Hyun SB, Kitazume Y, Nagahori M, et al. Magnetic resonance enterocolonography is useful for simultaneous evaluation of small and large intestinal lesions in Crohn's disease. *Inflammatory bowel diseases* 2011; **17**: 1063-72.

32. Oussalah A, Laurent V, Bruot O, et al. Diffusion-weighted magnetic resonance without bowel preparation for detecting colonic inflammation in inflammatory bowel disease. *Gut* 2010; **59**: 1056-65.
33. von Wagner C, Halligan S, Atkin WS, Lilford RJ, Morton D, Wardle J. Choosing between CT colonography and colonoscopy in the diagnostic context: a qualitative study of influences on patient preferences. *Health expectations : an international journal of public participation in health care and health policy* 2009; **12**: 18-26.
34. 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.
35. Takenaka K, Ohtsuka K, Kitazume Y, et al. Comparison of magnetic resonance and balloon enteroscopic examination of the small intestine in patients with Crohn's disease. *Gastroenterology* 2014; **147**: 334-42.e3.
36. Kopylov U, Klang E, Yablecovitch D, et al. Magnetic resonance enterography versus capsule endoscopy activity indices for quantification of small bowel inflammation in Crohn's disease. *Therapeutic advances in gastroenterology* 2016; **9**: 655-63.
37. Arulanandan A, Dulai PS, Singh S, Sandborn WJ, Kalmaz D. Systematic review: Safety of balloon assisted enteroscopy in Crohn's disease. *World journal of gastroenterology* 2016; **22**: 8999-9011.
38. Rezapour M, Amadi C, Gerson LB. Retention associated with video capsule endoscopy: systematic review and meta-analysis. *Gastrointestinal endoscopy* 2017; **85**: 1157-68.e2.
39. Solem CA, Loftus EV, Jr., Fletcher JG, et al. Small-bowel imaging in Crohn's disease: a prospective, blinded, 4-way comparison trial. *Gastrointestinal endoscopy* 2008; **68**: 255-66.
40. Maconi G, Nylund K, Ripolles T, et al. EFSUMB Recommendations and Clinical Guidelines for Intestinal Ultrasound (GIUS) in Inflammatory Bowel Diseases. *Ultraschall in der Medizin* **2018**; doi: 10.1055/s-0043-125329. [Epub ahead of print]
41. Taylor SA, Avni F, Cronin CG, et al. The first joint ESGAR/ ESPR consensus statement on the technical performance of cross-sectional small bowel and colonic imaging. *European radiology* 2017; **27**: 2570-82.

Tables

Table 1 Demographics of final trial 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 (36)	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 ^c	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)

^a Montreal classification not collected for new diagnosis patients

^b Surgical resection for inflammatory mass 1 year prior to Crohn's disease diagnosis

^c Patients could take more than one type of medication

Table 2 Patient characteristics: disease presence and activity-consensus reference standard

	New diagnosis (n=133) [n (%)]	Suspected relapse (n=151) [n (%)]	Full cohort (n=284) [n (%)]
Disease presence			
Small bowel disease present	111 (83)	122 (81)	233 (82)
Colonic disease present	77 (58)	52 (34)	129 (45)
Isolated small bowel disease present	56 (42)	85 (56)	141 (50)
Isolated colonic disease present	22 (17)	15 (10)	37 (13)
Both small bowel and colonic disease present	55 (41)	37 (25)	92 (32)
Total number patients with disease present	133 (100)	137 (91)	270 (95)
Median number of involved small bowel segments [median (IQR), max]	1 (1 to 1), 4	1 (1 to 1), 3	1 (1 to 1), 4
Median number of involved colonic segments [median (IQR), max]	1 (0 to 3), 6	0 (0 to 1), 6	0 (0 to 2), 6
Disease activity			
Small bowel disease active	104 (94)	105 (86)	209 (90)
Colonic disease active	76 (99)	50 (96)	126 (98)
Total number patients with disease active	130 (98)	121 (88)	251 (93)
Criteria for activity ^a			
<i>Ulceration at endoscopy</i>	71 (55)	26 (21)	97 (39)
<i>CRP >8 mg/l</i>	47 (36)	57 (47)	104 (41)
<i>Calprotectin >250 mcg/g</i>	41 (32)	43 (36)	84 (33)
<i>Histological evidence of activity</i>	100 (77)	36 (30)	136 (54)

^a Patients could meet more than one criteria for disease activity
IQR refers to interquartile range

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 ^a	MRE	US	Difference (P value)	Number disease negative ^a	MRE	US	Difference (P value)
Small bowel disease extent ^b	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 extent ^b	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 extent ^b	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 presence ^c	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

^a Patients by consensus reference standard

^b Agreement with reference standard for disease presence and segmental location

^c Agreement with reference standard for disease presence (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 diagnosis N=133							Suspected relapse N=151						
	DP, DN ^a	Sensitivity % (CI 95%)			Specificity % (CI 95%)			DP, DN ^a	Sensitivity % (CI 95%)			Specificity % (CI 95%)		
		MRE	US	Difference	MRE	US	Difference		MRE	US	Difference	MRE	US	Difference
Small bowel disease extent ^b	111,22	77 (66 to 86)	66 (54 to 77)	11 (-2 to 24)	98 (82 to 100)	88 (64 to 97)	10 (-5 to 24)	122,29	82 (72 to 89)	74 (62 to 83)	8 (-3 to 19)	92 (74 to 98)	75 (50 to 90)	17 (-3 to 37)
Small bowel disease presence	111,22	96 (89 to 99)	92 (82 to 96)	4 (-1 to 10)	99 (84 to 100)	91 (65 to 98)	8 (-5 to 21)	122,29	97 (91 to 99)	92 (82 to 96)	5 (0 to 11)	94 (76 to 99)	78 (50 to 92)	16 (-4 to 36)
Colonic disease extent ^b	77,56	17 (9 to 30)	9 (4 to 19)	8 (-2 to 19)	93 (82 to 98)	92 (80 to 97)	1 (-7 to 10)	52,99	31 (17 to 48)	33 (19 to 51)	-2 (-22 to 17)	93 (85 to 97)	94 (86 to 97)	-1 (-7 to 5)
Colonic disease presence	77,56	47 (31 to 64)	67 (49 to 81)	-20 (-39 to -1)	96 (86 to 99)	95 (84 to 98)	1 (-5 to 7)	52,99	84 (67 to 94)	80 (61 to 91)	4 (-11 to 20)	96 (88 to 98)	95 (89 to 99)	-1 (-5 to 4)
Small bowel and colonic disease extent ^b	133,0	33 (22 to 46)	20 (12 to 30)	13 (1 to 26)	N/A	N/A	N/A	137,14	56 (43 to 68)	40 (28 to 52)	16 (2 to 31)	80 (42 to 96)	61 (24 to 88)	19 (-20 to 59)
Small bowel and colonic disease presence ^c	133,0	65 (52 to 76)	66 (53 to 77)	-1 (-15 to 13)	N/A	N/A	N/A	137,14	88 (79 to 93)	76 (64 to 85)	12 (2 to 22)	80 (42 to 96)	61 (23 to 89)	19 (-20 to 59)

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

^b Agreement with reference standard for disease presence and segmental location

^c Agreement with reference standard for disease presence (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 active disease ^a	MRE	US	Difference (P value)	Number inactive disease ^a	MRE	US	Difference (P value)
Active small bowel disease ^b	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 ^b	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 ^c	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

^a Patients by consensus reference standard

^b Agreement with reference standard for disease active

^c Agreement with reference standard for active disease presence (patients with disease in the small bowel, colon or both)

Figures

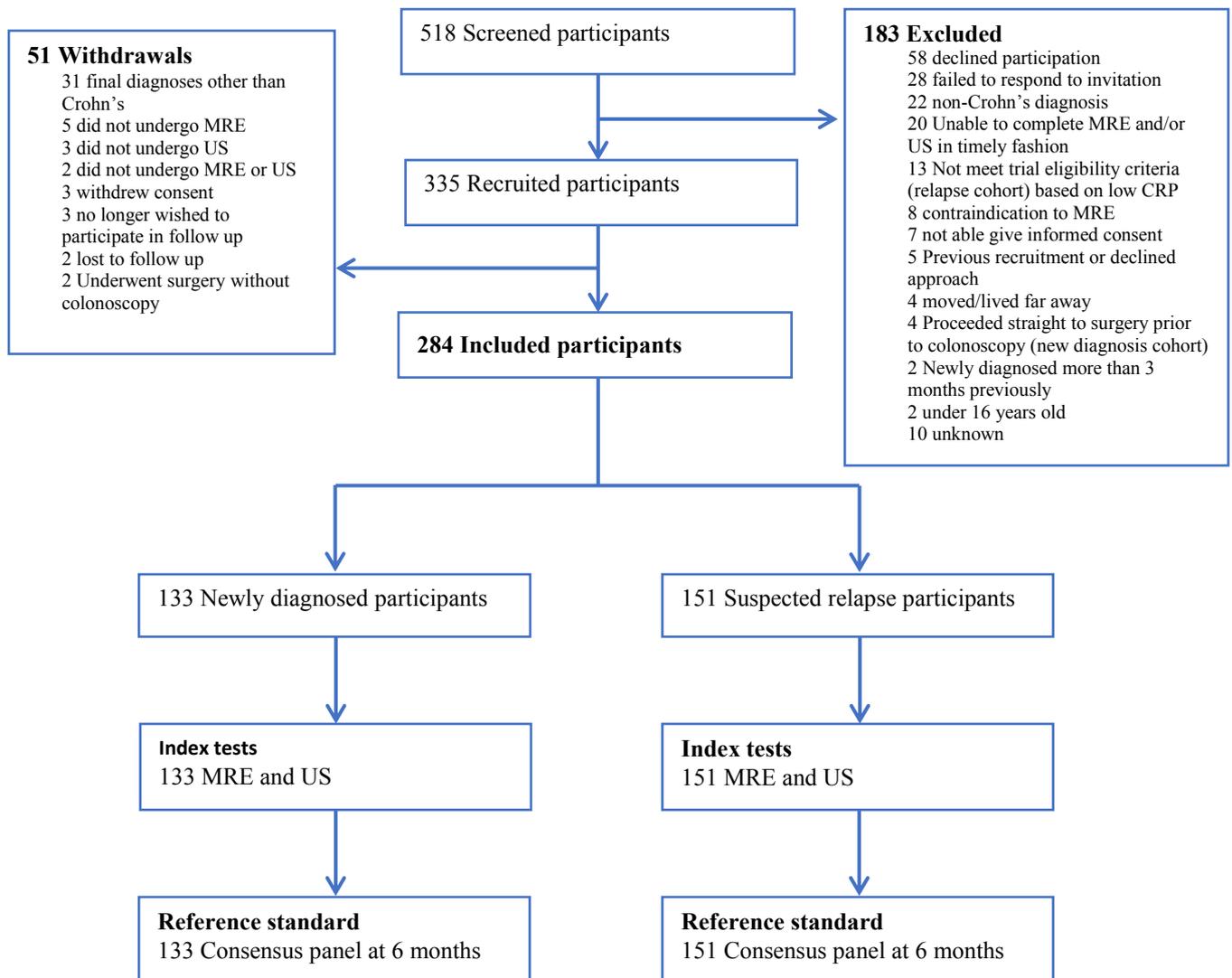


Figure 1. Patient flow diagram

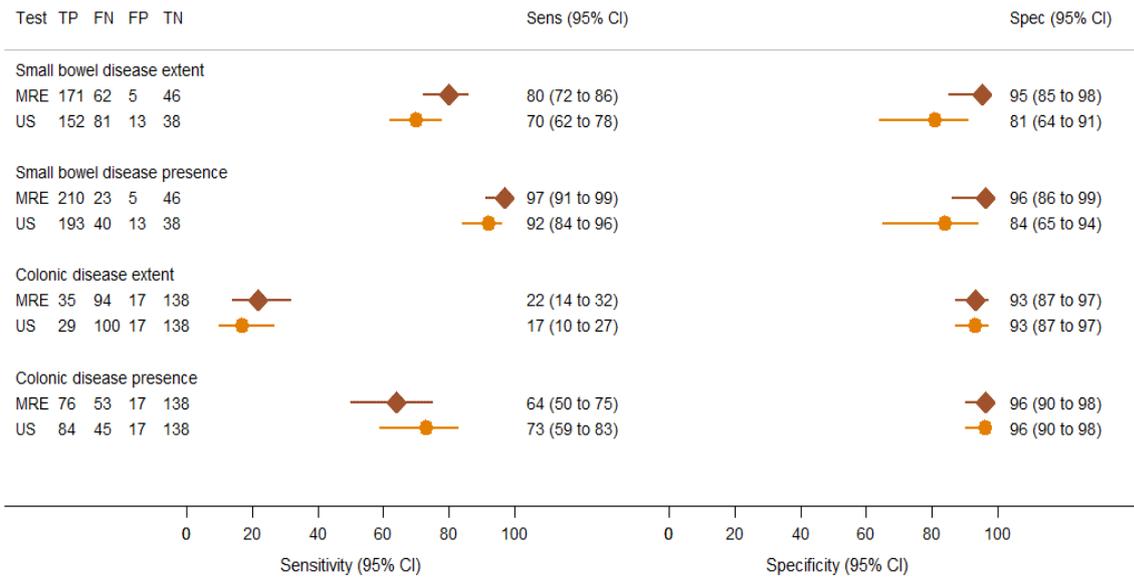


Figure 2 MRE and US sensitivity and specificity for small bowel and colonic disease extent and presence against the consensus reference standard