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

Stuart A Taylor, Susan Mallett, Gaurang Bhatnagar, Rachel Baldwin-Cleland, Stuart Bloom, Arun Gupta, Peter J Hamlin, Ailsa L Hart, Antony Higginson, Ilan Jacobs, Sara McCartney, Anne Miles, Charles D Murray, Andrew A Plumb, Richard C Pollok, Shonit Punwani, Laura Quinn, Manuel Rodriguez-Justo, Zainib Shabir, Andrew Slater, Damian Tolan, Simon Travis, Alastair Windsor, Peter Wylie, Ian Zealley, Steve Halligan, on behalf of the METRIC study investigators*

Summary

Background Magnetic resonance enterography (MRE) and ultrasound are used to image Crohn’s disease, but their comparative accuracy for assessing disease extent and activity is not known with certainty. Therefore, we did a multicentre trial to address this issue.

Methods We recruited patients from eight UK hospitals. Eligible patients were 16 years or older, with newly diagnosed Crohn’s disease or with established disease and suspected relapse. Consecutive patients had MRE and ultrasound in addition to standard investigations. Discrepancy between MRE and ultrasound for the presence of small bowel disease triggered an additional investigation, if not already available. The primary outcome was difference in per-patient sensitivity for small bowel disease extent (correct identification and segmental localisation) against a construct reference standard (panel diagnosis). This trial is registered with the International Standard Randomised Controlled Trial, number ISRCTN03982913, and has been completed.

Findings 284 patients completed the trial (133 in the newly diagnosed group, 151 in the relapse group). Based on the reference standard, 233 (82%) patients had small bowel Crohn’s disease. The sensitivity of MRE for small bowel disease extent (80% [95% CI 72–86]) and presence (97% [91–99]) were significantly greater than that of ultrasound (70% [62–78] for disease extent, 92% [84–96] for disease presence); a 10% (95% CI 1–18; p=0.027) difference for extent, and 5% (1–9; p=0.025) difference for presence. The specificity of MRE for small bowel disease extent (95% [85–98]) was significantly greater than that of ultrasound (81% [64–91]); a difference of 14% (1–27; p=0.039). The specificity for small bowel disease presence was 96% (95% CI 86–99) with MRE and 84% (65–94) with ultrasound (difference 12% [0–25]; p=0.054). There were no serious adverse events.

Interpretation Both MRE and ultrasound have high sensitivity for detecting small bowel disease presence and both are valid first-line investigations, and viable alternatives to ileocolonoscopy. However, in a national health service setting, MRE is generally the preferred radiological investigation when available because its sensitivity and specificity exceed ultrasound significantly.

Funding National Institute of Health and Research Health Technology Assessment.

Copyright © The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.
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. However, dissemination of cross-sectional imaging has occurred despite a paucity of supportive data from prospective multicentre studies recruiting consecutive and unselected patients. Emphasis is placed on magnetic resonance enterography (MRE) and enteric ultrasound because they avoid ionising radiation. Clinical uptake of ultrasound has been hampered by concerns over diagnostic accuracy and perceived high levels of interobserver variation. MRE is a newer innovation that necessitates access to comparatively restricted advanced technology imaging platforms. We searched PubMed and Embase in January, 2018, for articles published between Jan 1, 1990, and Jan 1, 2018, without language restriction. We used MeSH and full-text search for “Crohn’s disease”, “magnetic resonance imaging”, “ultrasound”, and “diagnostic accuracy”. We retrieved primary literature but we placed emphasis on meta-analyses and systematic reviews using appropriate filters. We found several meta-analyses, which generally suggest that MRE and ultrasound have similar 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 one meta-analysis, just one of 33 included studies compared MRE and ultrasound directly in the same patients. Many studies also rate poorly on the Quality Assessment of Diagnostic Accuracy Studies tool.

Evidence before this study

Cross-sectional imaging is fundamental for diagnosis and staging Crohn’s disease, the primary literature is of questionable quality. Most studies are small and done in a single centre, and few compare tests directly in the same patients, despite this being advocated as an optimal method for diagnostic accuracy studies. For example, in their meta-analysis, Greenup and colleagues found that just one of 33 included studies compared MRE and ultrasound directly in the same patients. Additionally, very few studies use a construct reference standard model (panel diagnosis), which incorporates concepts of diagnostic test validation based on patient outcomes and has distinct methodological advantages when a single reference standard is elusive.

To redress this, we did a multicentre trial to elucidate and then directly compare the diagnostic accuracy of MRE and ultrasound 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

Study design and participants

The METRIC study is a multicentre trial that compares the diagnostic accuracy of MRE and enteric ultrasound for the presence, extent, and activity of small bowel Crohn’s disease in newly diagnosed patients or patients with established disease and suspected relapse. We 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, and can be found online.

We recruited patients from eight UK National Health Service (NHS) teaching and general hospitals,
representative of institutions likely to implement MRE and ultrasound for patient management (appendix p 1). All sites had an established inflammatory bowel disease service and were already doing MRE and ultrasound as part of usual clinical practice.

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

Patients were eligible for the suspected luminal relapse group if they had established Crohn’s disease (>3 months) and there was a strong clinical suspicion of luminal relapse based on either objective markers of inflammatory activity (C-reactive protein [CRP] concentration >8 mg/L or faecal calprotectin concentration >100 μg/g), symptoms suggestive of luminal stenosis (including obstructive symptoms, such as colicky abdominal pain, vomiting), or abnormal endoscopy. Eligible patients for both groups were aged 16 years or older. Patients were ineligible if they were pregnant or if they had contraindications to MRI. Those with psychiatric or other disorders who were unable to give informed consent were also excluded, as were those with evidence of severe or uncontrolled systemic disease. Patients in the newly diagnosed group were excluded if they had surgical resection before colonoscopy.

Members of the local research team identified suitable patients from outpatient clinics, multidisciplinary team meetings, and inpatient wards, and they took informed consent from consecutive, unselected, eligible patients. A screening log detailed all approached patients and reasons for non-participation, if applicable. We collated patient demographics and clinical data (eg, age, sex, Montreal classification [relapse group only], disease or symptom duration, medication, and surgical history).

Procedures

Patients had MRE and ultrasound in addition to any other enteric imaging or endoscopic investigations done during their usual clinical care.

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

Across all sites, 28 practitioners interpreted the MRE and ultrasound studies (27 radiologists and 1 sonographer). Eight radiologists interpreted MRE only, three performed and interpreted ultrasound only, and 16 performed and interpreted ultrasound and interpreted MRE. All radiologists were affiliated with the British Society of Gastrointestinal and Abdominal Radiology, with declared subspecialty interest in gastrointestinal radiology, and had completed the Fellowship of the Royal College of Radiologists, with at least 1 year of subspecialty training in gastrointestinal radiology. The sonographer had received formal training according to their sites’ local policies, was doing enteric ultrasound routinely, and had 20 years of experience. Radiologists interpreting MRE had a median of 10 years (IQR 6–11) of experience, and practitioners interpreting ultrasound had a median of 8 years (4–11) of experience. The median number of examinations done per month at each recruitment site during the conduct of the trial was 30 examinations (20–45) for MRE and 25 examinations (12–40) for ultrasound. Before trial commencement, we held a 2-day hands-on workshop for investigators to standardise ultrasound technique and agree on description of enteric findings.

MRE and ultrasound were interpreted by two different practitioners, each masked to the findings of the other, and masked to all other imaging, endoscopic, and clinical data except the group to which the patient was recruited (ie, newly diagnosed 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 extraenteric complications, using established criteria (appendix p 4).\(^{24,25}\)

The segmental location of any disease was also recorded, using standard definitions;\(^{26}\) disease sites separated by more than 3 cm 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.

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

For cases in which MRE and ultrasound had a discrepancy for the presence or location of small bowel disease, we did an arbitral small bowel investigation if patients had not already had additional small bowel imaging as part of usual care. We defined discrepancy as terminal ileal disease reported on either MRE or ultrasound in the absence of endoscopic visualisation, or disease reported in the small bowel upstream of the terminal ileum on either MRE or ultrasound. The nature of the additional test was left to local discretion and could include, for example, barium follow through, CTE, or capsule endoscopy. We also permitted repeat, targeted, and unmasked MRE or ultrasound to resolve discrepancies.

Where possible, we collected CRP concentration, calprotectin concentration, and the Harvey Bradshaw
518 screened participants

183 excluded
- 58 declined participation
- 28 failed to respond to invitation
- 22 diagnosis other than Crohn’s disease
- 20 unable to complete MRE or ultrasound in timely fashion
- 13 did not meet trial eligibility criteria (relapse cohort) based on low CRP
- 8 contraindication to MRE
- 7 not able to give informed consent
- 5 previous recruitment or declined approach
- 4 moved or lived far away
- 4 proceeded straight to surgery before colonoscopy (new diagnosis cohort)
- 2 newly diagnosed >3 months previously
- 2 <16 years old
- 10 unknown

335 recruited participants

51 withdrawals
- 31 final diagnoses other than Crohn’s disease
- 5 did not have MRE
- 3 did not have ultrasound
- 2 did not have MRE or ultrasound
- 2 withdrew consent
- 3 no longer wished to participate in follow-up
- 2 lost to follow-up
- 2 had surgery without colonoscopy

518 screened participants

183 excluded
- 58 declined participation
- 28 failed to respond to invitation
- 22 diagnosis other than Crohn’s disease
- 20 unable to complete MRE or ultrasound in timely fashion
- 13 did not meet trial eligibility criteria (relapse cohort) based on low CRP
- 8 contraindication to MRE
- 7 not able to give informed consent
- 5 previous recruitment or declined approach
- 4 moved or lived far away
- 4 proceeded straight to surgery before colonoscopy (new diagnosis cohort)
- 2 newly diagnosed >3 months previously
- 2 <16 years old
- 10 unknown

335 recruited participants

284 included participants

133 newly diagnosed participants

151 suspected relapse participants

Index tests
- 133 MRE and ultrasound
- 151 MRE and ultrasound

Reference standard
- 133 consensus panel at 6 months
- 151 consensus panel at 6 months

Outcomes

The primary outcome was the per-patient difference in sensitivity between MRE and ultrasound for correct identification and localisation of small bowel Crohn’s disease, irrespective of activity—ie, the extent of small bowel disease. To be truly positive for disease extent, the index test had to correctly locate the presence and segmental location of disease (terminal ileum, ileum, jejunum, or duodenum). Secondary outcomes reported here were 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 and comparative patient experience. Secondary outcomes also included comparative impact of MRE and ultrasound on clinician diagnostic confidence for presence of Crohn’s disease and their influence on management, cost-effectiveness of MRE and ultrasound (compared to each other), diagnostic impact of novel MRE sequences (eg, diffusion-weighted imaging), influence of sequence selection on MRE diagnostic accuracy, diagnostic accuracy of small intestine contrast enhanced ultrasonography (SICUS) compared with standard ultrasound, influence of oral contrast agent and ingested volume on small bowel distension and patient experience during MRE, and interobserver variation, which will be reported elsewhere.

We reported most outcomes for the newly diagnosed and suspected luminal relapse groups individually, and for the terminal ileum and colon using colonoscopy as a standalone reference standard (when available) because of its robustness for identifying disease.

We prespecified all outcomes in the protocol except for the newly diagnosed and suspected luminal relapse groups individually, and for the terminal ileum and colon using colonoscopy as a standalone reference standard (when available) because of its robustness for identifying disease.

Articles

CRP=C-reactive protein, MRE=magnetic resonance enterography.
(small bowel and colonic disease combined), which were exploratory.

Safety reporting was limited to any suspected unexpected serious adverse reaction directly related to MRE, ultrasound, or any arbiter small bowel imaging test. Expected adverse reactions, such as contrast agent allergy, were not collated.

**Statistical analysis**

We estimated that a sample size of 210 patients with small bowel disease would give 90% power to detect a clinically significant (10%) sensitivity difference for small bowel disease extent between MRE (83%, based on a sensitivity of 93% for disease presence and 90% for disease location) and ultrasound (73%, based on a sensitivity of 88% for disease presence and 83% for disease location), assuming 68% positivity for both tests and using methods for comparative studies. A 10% difference in sensitivity was deemed to be clinically meaningful in routine practice by the study investigators at the time of trial design. We assumed a 70% prevalence of small bowel disease and 10% loss to follow-up or diagnosis with a disease other than Crohn’s disease, which gave a target sample size of 334 patients across both groups (167 in the newly diagnosed group and 167 in the relapse group). The trial was not powered to detect differences between the groups, or between bowel segments.

We treated disease reported as equivocal as positive in the analysis. We calculated the primary outcome per patient. We based the secondary outcomes for bowel segments on all segments, excluding those resected at baseline (neoterminal ileum was considered as the terminal ileum).

We calculated a direct comparison of sensitivity and specificity differences between MRE and ultrasound using bivariate, multilevel, patient-specific (conditional), random-effects models, from paired data using meqrlogit in STATA 14.2 (College Station, TX, USA). When models did not converge due to small numbers of patients, we used McNemar’s comparison of paired proportions to obtain univariable estimates and we calculated exact 95% CI. We did analysis by segment using a population-averaged, random-effects model (using logit, including robust standard errors). We based statistical significance on 95% CI.

This trial is registered with the International Standard Randomised Controlled Trial, number ISRCTN03982913.

**Role of the funding source**

The 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, or 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**

We commenced recruitment on Dec 4, 2013, and completed it on Sept 30, 2016. Overall, we assessed 518 patients for eligibility, of whom 183 were excluded (figure I). Of the 335 patients who entered the trial, 51 were subsequently excluded (20 men, median 30 years [IQR 24–41]); 31 did not have Crohn’s disease, two were lost to follow-up, ten did not have MRE or ultrasound or both, six withdrew consent or no longer wished to participate in follow-up, and two newly diagnosed patients had surgery without colonoscopy. With a final group...
The sensitivity of MRE (97% [95% CI 91–99]) for the presence of small bowel disease, regardless of location, was significantly greater than that of ultrasound (92% [84–96]), with a difference of 5% (1–9; table 3; figure 2).

The sensitivities of MRE and ultrasound for small bowel disease presence and extent in the newly diagnosed group and relapsed group were very similar to those estimated across all patients (table 4). However, ultrasound (67% [95% CI 49–81]) had significantly greater sensitivity for colonic disease presence than MRE (47% [31–64]) in the newly diagnosed patient group, with a difference of 20% (1–39). MRE and ultrasound had no significant difference in sensitivity or specificity for the extent and presence of colonic disease (table 3; figure 2). For both tests, sensitivity for colonic disease tended to be higher in the relapsed group than the newly diagnosed group (table 4), although the estimated sensitivity for colonic disease extent was poor for both groups.

The detection rate for individual small bowel and colonic segments is given in the appendix (p 7). Although the trial was not powered to detect differences on a segmental level, MRE was significantly more sensitive than ultrasound for ileal (84% [95% CI 67–93] vs 56% [38–73]) and rectal disease (44% [32–58] vs 22% [13–35]).

The sensitivity of MRE for active small bowel disease was 96% (95% CI 92–99) compared with 90% (82–95) for ultrasound, a significant difference of 6% (2–11; table 5).

The specificity for active small bowel disease and for active colonic disease were not significantly different between tests (table 5). The sensitivity and specificity for active disease split by patient group were very similar to those estimated across all patients (appendix p 8).

21 patients had enteric fistulae, and seven patients had intra-abdominal abscess. MRE detected five (71%) of seven abscesses, whereas ultrasound detected three (43%) of seven abscesses. MRE detected 18 (86%) of 21 patients with enteric fistulae compared with 11 (52%) of 21 patients for ultrasound.

Against a colonoscopic standard of reference (available in 186 patients), MRE had a sensitivity of 97% (95% CI 91–99) for terminal ileal disease presence compared with a sensitivity of 91% (79–97) for ultrasound, a difference of 6% (−1 to 12; appendix p 9). The sensitivity for colonic disease presence was modest for both MRE (41% [26–58]) and ultrasound (49% [33–65]) and was not statistically different.

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

No serious adverse events or any other adverse events were reported.

Discussion

In the METRIC trial, we found that both MRE and ultrasound were highly accurate for detecting small bowel Crohn’s disease, achieving 97% sensitivity for
MRE and 92% sensitivity for ultrasound. Barium fluoroscopy has long been advocated as a sensitive test for mucosal disease inaccessible to endoscopy, although its support is limited to a handful of small studies and its accuracy is increasingly questioned. Conversely, against a rigorous ileocolonoscopic reference standard, we found that MRE and ultrasound 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 ileocolonoscopy. Of the two, we found MRE had significantly higher sensitivity and specificity than ultrasound for small bowel extent, and higher sensitivity for disease presence. Overall, no significant difference was found in diagnostic accuracy for colonic disease, consistently lower than for small bowel disease, although ultrasound had greater sensitivity than MRE in newly diagnosed patients. To our knowledge, the METRIC trial is the largest prospective multicentre trial to date directly comparing diagnostic accuracy of MRE and ultrasound for the presence, extent, and activity of Crohn’s disease in the same patients.

Our primary outcome combined those aspects necessary to stage small bowel Crohn’s disease correctly—ie, 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 might tip the balance towards medical rather than surgical intervention in the face of otherwise isolated terminal ileal disease. As expected, sensitivity for disease extent was lower than that for disease detection alone. Our detection rates were at the upper end of estimates from previous meta-analyses. Dong and colleagues estimated ultrasound to have a sensitivity 88% and a specificity of 97%; Liu and colleagues reported corresponding figures of 86% sensitivity and 93% specificity for MRE. However, the primary literature is markedly heterogeneous, which affects the validity of point estimates. Most studies were single centre and typically recruited fewer than 50 patients, and many were methodologically poor. Direct comparison of diagnostic tests in

<table>
<thead>
<tr>
<th>Test</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small bowel disease extent</td>
<td>MRE 121</td>
<td>62</td>
<td>5</td>
<td>46</td>
<td>80% (72–86)</td>
<td>95% (85–98)</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>152</td>
<td>81</td>
<td>13</td>
<td>38</td>
<td>70% (62–78)</td>
<td>81% (64–91)</td>
</tr>
<tr>
<td>Small bowel disease presence</td>
<td>MRE 210</td>
<td>23</td>
<td>5</td>
<td>46</td>
<td>97% (91–99)</td>
<td>96% (86–99)</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>319</td>
<td>40</td>
<td>13</td>
<td>38</td>
<td>92% (84–96)</td>
<td>84% (65–94)</td>
</tr>
<tr>
<td>Colonic disease extent</td>
<td>MRE 35</td>
<td>94</td>
<td>17</td>
<td>138</td>
<td>22% (14–32)</td>
<td>93% (87–97)</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>95</td>
<td>29</td>
<td>100</td>
<td>17</td>
<td>17% (10–27)</td>
<td>93% (87–97)</td>
</tr>
<tr>
<td>Colonic disease presence</td>
<td>MRE 76</td>
<td>53</td>
<td>17</td>
<td>138</td>
<td>64% (50–75)</td>
<td>96% (90–98)</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>53</td>
<td>84</td>
<td>45</td>
<td>17</td>
<td>73% (59–83)</td>
<td>96% (90–98)</td>
</tr>
</tbody>
</table>

Data are n, % (95% CI), or p value. MRE=magnetic resonance enterography. *Patients by consensus reference standard. †Agreement with reference standard for disease presence and segmental location. ‡Agreement with reference standard for disease presence (patients with disease in the small bowel, colon, or both).

Table 3: Per-patient sensitivity and specificity for disease presence and extent against the consensus reference standard for patient groups combined

Figure 2: Sensitivity and specificity of MRE and ultrasound for the extent and presence of small bowel and colonic disease against the consensus reference standard

FN=false negative. FP=false positive. MRE=magnetic resonance enterography. TN=true negative. TP=true positive. Error bars represent 95% CI.
## Table 4:

### Sensitivity and specificity for disease presence and extent against the consensus reference standard (according to patient cohort)

<table>
<thead>
<tr>
<th>Disease present and extent†</th>
<th>Small bowel disease extent†</th>
<th>Colonic disease extent†</th>
<th>Small bowel and colonic disease extent†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suspected relapse group</strong></td>
<td>111.2 (66 to 86)</td>
<td>77% (95 to 99)</td>
<td>22% (74 to 89)</td>
</tr>
<tr>
<td><strong>New diagnosis group</strong></td>
<td>111.2 (66 to 86)</td>
<td>77% (95 to 99)</td>
<td>22% (74 to 89)</td>
</tr>
</tbody>
</table>

### Specificity

<table>
<thead>
<tr>
<th>Disease present and extent‡</th>
<th>Small bowel disease presence</th>
<th>Colonic disease presence</th>
<th>Small bowel and colonic disease presence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suspected relapse group</strong></td>
<td>77% (95 to 99)</td>
<td>84% (82 to 97)</td>
<td>52% (42 to 96)</td>
</tr>
<tr>
<td><strong>New diagnosis group</strong></td>
<td>77% (95 to 99)</td>
<td>84% (82 to 97)</td>
<td>52% (42 to 96)</td>
</tr>
</tbody>
</table>

### Notes

- Data are n or % (95% CI), unless otherwise specified. MRE = magnetic resonance enterography. NA = not applicable.
- *Disease positive and disease negative patients by consensus reference standard.
- †Agreement with reference standard for disease presence and segmental location.
- ‡Agreement with reference standard for disease presence (patients with disease in the small bowel, colon or both).

The same patients are advocated as the optimal method for diagnostic accuracy studies because differences are attributable directly to the tests and not to differences between participants or study methods. Such head-to-head comparisons are rare in the medical literature. Reference standards might also be applied inconsistently, with endoscopy, surgery, and imaging all variably employed. For example, in a comparative study with ultrasound, Castiglione and colleagues used MRE without any additional reference standard in many recruits, which introduces the potential for incorporation bias.

We used the construct reference standard model (panel diagnosis), which incorporates multiple data sources with clinical outcome. Although such an approach does have limitations, including potential panel bias, it is considered a very robust method for diagnostic accuracy studies in which a single external reference standard is elusive. To reduce incorporation bias, patients without supplementary small bowel imaging had a third small bowel investigation whenever discrepancy between MRE and ultrasound arose. Notably, when our analysis was limited to an ileocolonoscopic reference standard, any differences in accuracy between MRE and ultrasound closely mirrored those found using the consensus panel reference.

We recruited approximately equally from two patient groups: newly diagnosed Crohn’s disease and established disease with relapse. Both groups are clinically distinct and important, and might manifest with differing disease phenotypes; prevalence of stricturing and penetrating disease increases with time. Noting that the METRIC trial was not powered to detect differences between these two patient groups, we found that sensitivity for small bowel disease was similar, although specificity tended to be lower in patients in the relapse group. Conversely, sensitivity for colonic disease was higher in the relapse group, but was still poor for colonic disease extent (about 30%).

In newly diagnosed patients, ultrasound achieved significantly greater sensitivity for colonic disease than MRE (67% vs 47%). Optimised colonic assessment with MRE requires purgation and fluid distension, which are both omitted from routine MRE protocols; however, ultrasound generally relies on assessing the manually compressed uncleansed colon wall. Accuracy for both techniques in the colon still falls short of colonoscopy, and accuracy with MRE is somewhat lower than previously reported.

By way of explanation, ileocolonoscopy 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. Previous single-centre explanatory studies either use groups enriched with more advanced colonic disease, or report sensitivity for deep rather than superficial mucosal disease.

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

The METRIC trial does have some limitations. It was conceived as a large pragmatic trial since the medical 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. However, we acknowledge that specialist practitioners working in high volume practices might achieve sensitivities in excess of our findings. Imaging was interpreted according to local clinical practice to mirror real-world procedures within the NHS and enhance generalisability of our results. We acknowledge that masking 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 unmasking of reporting practitioners. Recruited patients were representative of those having MRE and ultrasound in daily practice, although we did exclude pregnant women, patients having routine therapeutic response assessment, and patients with contraindications to MRI. Our results are therefore highly likely to be extrapolable across the NHS and similar health-care settings. The prevalence of active disease was predictably greater for recruited patient groups. Therefore, the reported high specificity of MRE and ultrasound should be viewed in this context.

We did not standardise the third small bowel investigation whenever discrepancy between MRE and ultrasound arose, and this decision was left to the discretion of the recruitment site. Direct mucosal visualisation is possible with push enteroscopy or capsule endoscopy, but to insist on such investigations was not practicable in the setting of a pragmatic multicentre trial given their cost, relatively inadequate availability, and probable negative effect 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. Similarly the risk of capsule retention is around 8% in patients with known Crohn’s disease and specificity is questioned. We also considered that the invasive nature of capsule endoscopy or enteroscopy would result in considerable spectrum bias relating to differences between patients who would and would not agree to consent (even if they were available and affordable).

To reduce incorporation bias from MRE or ultrasound, we required at least one independent biochemical, endoscopic, or histological marker of disease activity before a patient could be diagnosed with active small bowel or colonic disease. Biochemical markers, such as calprotectin and CRP concentrations, 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.

Some data suggest that the diagnostic accuracy of ultrasound can be improved with an oral contrast load (SICUS), particularly for luminal stenosis, and intravenous contrast enhanced ultrasound (CEUS) might have utility for assessing disease activity. However, neither SICUS nor CEUS have disseminated as first-line investigations outside specialist units, and if used are often employed as problem solving tools. Standard ultrasound is overwhelmingly the most commonly used technique in routine clinical practice. Future prospective research could consider inclusion of 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 ultrasound and MRE achieve excellent diagnostic accuracy for the extent and

---

Table 5: Per-patient sensitivity and specificity for the presence of active disease versus the consensus reference standard for patient groups combined

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th></th>
<th>Specificity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients</td>
<td>MRE</td>
<td>Ultrasound</td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td>with active</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disease*</td>
<td>209</td>
<td>96% (92 to 99)</td>
<td>90% (82 to 95)</td>
<td>6% (2 to 11)</td>
</tr>
<tr>
<td></td>
<td>Active small bowel disease†</td>
<td>75</td>
<td>83% (68 to 92)</td>
<td>77% (60 to 88)</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>63% (48 to 76)</td>
<td>66% (51 to 79)</td>
<td>3% (-18 to 13)</td>
</tr>
<tr>
<td></td>
<td>Active colonic disease‡</td>
<td>158</td>
<td>97% (91 to 99)</td>
<td>98% (94 to 99)</td>
</tr>
<tr>
<td></td>
<td>251</td>
<td>77% (68 to 85)</td>
<td>66% (56 to 75)</td>
<td>11% (1 to 21)</td>
</tr>
<tr>
<td></td>
<td>Active small bowel and colonic disease‡</td>
<td>33</td>
<td>28% (10 to 56)</td>
<td>28% (10 to 56)</td>
</tr>
</tbody>
</table>

Data are n, % (95% CI), or p value. MRE=magnetic resonance enterography. †Patients by consensus reference standard. ‡Agreement with reference standard for disease active. *Agreement with reference standard for active disease presence (patients with disease in the small bowel, colon, or both).
activity of small bowel Crohn’s disease in newly diagnosed patients and those who have relapsed, and both tests are valid firstline investigations. In an NHS setting, MRE is generally the preferred radiological investigation when available because its sensitivity and specificity exceed ultrasound significantly.

Contributors
All authors 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. RB-C, AM, and ZS contributed to data collection. SAT, LQ, SMa, and SH contributed to data interpretation. RB-C, SB, SMc, AG, PJH, AH, ALH, CDM, AAP, RCP, SP, MR-J, JS, AS, DT, ST, AW, PW, and IZ contributed to clinical studies and patient recruitment. LQ and SMa contributed to the statistical analysis. IJ acted as a public and patient representative. SAT, SMa, and SH wrote the initial manuscript draft. SAT is the study guarantor.

Declaration of interests
SAT, SH, ST, SMA, SB, DT, PJH, AG, ALH, AS, SMc, IZ, AH, AW, IJ, MR-J, SP, PW, and CDM report research grant support from the funder (National Institute of Health and Research Health Technology Assessment [NIHR HTA]) to do the study. SAT reports personal fees from Robarts plc. ST reports directorships of charities. IBD 2020 (UK 07962150), Crohn’s Colitis Cure (ABN 85 154 588 777), Trueflow Foundation (UK 10506717); grants and research support from AbbVie, International Organization for the Study of Inflammatory Bowel Diseases, Lilly, Union Chimique Belge (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, Takeda, Otsuka Pharmaceuticals, PDL BioPharma, Pfizer, Shire, Glenmark Pharma, Synthon Pharma, Glenmark Pharma, NPS Pharmaceuticals, Lilly, Warner Chilcott, Proctognym, VHSquared, Topivert, Ferring, Celgene, GlaxoSmithKline, Takeda, Amgen, Biogen, Enterome, Immunocore, Immunometabolism, Bioclinica, Boehringer Ingelheim, Gilead, Grunenthal, Janssen, Novartis, Celgene, Receptos, PharmOlam, Sigmod Pharma, and Theravance; expert testimony for sanatarus Inc, Cosmo Technologies, and Titollos Pharma; grants from Ferring, AbbVie, Schering-Plough, Merck Sharpe and Dohme (MSD), Proctet and Gamble, Warner Chilcott, International Organisation of Inflammatory Bowel Disease, Lilly, UCB, Vifor, and Norman Collisson Foundation; lecture fees for AbbVie, Schering Plough, Centocor, Merck, Genentech, UCB Pharma, Ferring-Pharmaceuticals, Tillots Labs, Shire, Sanofi-Aventis, Vifor, Takeda, Amgen, and Biogen; payment for manuscript preparation for Ferring; royalties from Wiley Blackwell, Elsevier, Oxford University Press; payment for development of educational presentations for Abbott Labs, Procter and Gamble, and Warner Chilcott; travel or accommodation expenses from AbbVie, United European Gastroenterology, and Shire; research nurse salary support from Schering Plough, Procter and Gamble, MSD, and Vifor; and unrestricted education grants from Abbott laboratories, Procter and Gamble, PDL BioPharma, Takeda, and International Consortium for Health Outcomes Measurement. ALH reports personal fees from Proctetvie, 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, and Cook; and grants and personal fees from Bard. AAP reports grants from NIH HTA; grants from NIH. Fellowships program; and personal fees from Alexion, Actavis, Dr Falk, Janssen-Cilag, and Takeda. IJ reports shared ownership in General Electric and sells GE equipment. CDM reports personal fees from Abbvie, MSD, and Janssen. AH reports personal fees from Toshiba. SH reports non-financial support from iCAD. SAT and ST are NIH senior investigators. All other authors declare no competing interests. This project was funded by the 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 NIHR 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. This report represents independent research commissioned by 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, NIHR, NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC), HTA programme, or the Department of Health. The trial steering committee and independent data monitoring committee met at least annually and included Vicky Goh (chair), Andrea Marshall (statistician), Ilan Jacobs (patient representative), James Lindsay (subject expert), Tim Orchard (chair), Doh-Mu Koh (subject expert), and Chris Rogers (statistician).

References
24 Taylor S, Mallett S, Bhatnagar G, et al. METRIC (MREnterography or uTRasound 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 Gastroenterol 2014; 14: 122.