

BIROn - Birkbeck Institutional Research Online

Taylor, S.A. and Mallett, S. and Bhatnagar, G. and Morris, S. and Quinn, L. and Tomini, F. and Miles, Anne and Baldwin-Cleland, R. and Bloom, S. and Gupta, A. and Hamlin, P.J. and Hart, A.L. and Higginson, A. and Jacobs, I. and McCartney, S. and Murray, C.D. and Plumb, A.A. and Pollok, R.C. 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. (2019) Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study. *Health Technology Assessment* 23 (42), pp. 1-162. ISSN 1366-5278.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/28810/>

Usage Guidelines:

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

or alternatively

Health Technology Assessment

Volume 23 • Issue 42 • August 2019

ISSN 1366-5278

Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study

Stuart A Taylor, Sue Mallett, Gauraang Bhatnagar, Stephen Morris, Laura Quinn, Florian Tomini, Anne Miles, Rachel Baldwin-Cleland, Stuart Bloom, Arun Gupta, Peter John Hamlin, Ailsa L Hart, Antony Higginson, Ilan Jacobs, Sara McCartney, Charles D Murray, Andrew AO Plumb, Richard C Pollok, Manuel Rodriguez-Justo, Zainib Shabir, Andrew Slater, Damian Tolan, Simon Travis, Alastair Windsor, Peter Wylie, Ian Zealley and Steve Halligan



Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study

Stuart A Taylor,^{1*} Sue Mallett,² Gauraang Bhatnagar,¹ Stephen Morris,³ Laura Quinn,² Florian Tomini,³ Anne Miles,⁴ Rachel Baldwin-Cleland,⁵ Stuart Bloom,⁶ Arun Gupta,⁵ Peter John Hamlin,⁷ Ailsa L Hart,⁸ Antony Higginson,⁹ Ilan Jacobs,¹⁰ Sara McCartney,⁶ Charles D Murray,¹¹ Andrew AO Plumb,¹ Richard C Pollok,¹² Manuel Rodriguez-Justo,¹³ Zainib Shabir,¹⁴ Andrew Slater,¹⁵ Damian Tolan,¹⁶ Simon Travis,¹⁷ Alastair Windsor,¹⁸ Peter Wylie,¹⁹ Ian Zealley²⁰ and Steve Halligan¹

¹Centre for Medical Imaging, University College London, London, UK

²Institute of Applied Health Research, National Institute for Health Research Birmingham Biomedical Research Centre, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

³Applied Health Research, University College London, London, UK

⁴Department of Psychological Sciences, Birkbeck, University of London, London, UK

⁵Intestinal Imaging Centre, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK

⁶Department of Gastroenterology, University College Hospital, London, UK

⁷Department of Gastroenterology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

⁸Inflammatory Bowel Disease Unit, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK

⁹Department of Radiology, Portsmouth Hospitals NHS Trust, Portsmouth, UK

¹⁰Independent patient representative, c/o Centre for Medical Imaging, University College London, London, UK

¹¹Department of Gastroenterology and Endoscopy, Royal Free London NHS Foundation Trust, London, UK

¹²Department of Gastroenterology, St George's Hospital, London, UK

¹³Department of Histopathology, University College Hospital, London, UK

¹⁴Comprehensive Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

¹⁵Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹⁶Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

¹⁷Translational Gastroenterology Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹⁸Department of Surgery, University College Hospital, London, UK

¹⁹Department of Radiology, Royal Free London NHS Foundation Trust, London, UK

²⁰Department of Radiology, Ninewells Hospital, Dundee, UK

*Corresponding author

Declared competing interests of authors: Stuart A Taylor reports personal fees from Robarts Clinical Trials Inc. (London, ON, Canada) outside the submitted work. Simon Travis reports receiving fees for consultancy work and/or speaking engagements from the following: AbbVie Inc. (Chicago, IL, USA), Centocor Inc. (Horsham, PA, USA), Schering-Plough (Kenilworth, NJ, USA), Bristol-Myers Squibb (New York, NY, USA), Chemocentryx Inc. (Mountain View, CA, USA), Cosmo Pharmaceuticals (Dublin, Ireland), Elan Pharma Inc. (Dublin, Ireland), Genentech Inc. (San Francisco, CA, USA), Giuliani SpA (Milan, Italy), Merck & Co. Inc. (Kenilworth, NJ, USA), Takeda UK Ltd (Woodburn Green, Buckinghamshire, UK), Otsuka Pharmaceuticals (Tokyo, Japan), PDL BioPharma (Nevada, NV, USA), Pfizer Inc. (San Francisco, CA, USA), Shire Pharmaceuticals UK (St Helier, Jersey), Glenmark Pharmaceuticals (Maharashtra, India), Synthron Biopharmaceuticals (Nijmegen, the Netherlands), NPS Pharmaceuticals (Bedminster, NJ, USA), Eli Lilly and Company (Indiana, IN, USA), Warner Chilcott Ltd, Proximagen Group Ltd (London, UK), VHSquared Ltd (Cambridge, UK), Topivert Pharma Ltd (London, UK), Ferring Pharmaceuticals (Saint-Prex, Switzerland), Celgene Corporation (Summit, NJ, USA), GlaxoSmithKline plc (Brentford, UK), Amgen Inc. (Thousand Oaks, CA, USA), Biogen Inc. (Cambridge, MA, USA), Enterome SA (Paris, France), Immunocore Ltd (Oxford, UK), Immunometabolism/Third Rock Ventures (Boston, MA, USA), Bioclinica Inc. (Newtown, PA, USA), Boehringer Ingelheim GmbH (Ingelheim am Rhein, Germany), Gilead Sciences Inc. (Foster City, CA, USA), Grunenthal Ltd (Aachen, Germany), Janssen Pharmaceutica (Beerse, Belgium), Novartis AG (Basel, Switzerland), Receptos Inc. (San Diego, CA, USA), Pharm-Olam International UK Ltd (Bracknell, UK), Sigmoid Pharma (Dublin, Ireland), Theravance Biopharma Inc. (Dublin, Ireland), Given Imaging Ltd (Yokneam Illit, Israel), UCB Pharma SA (Brussels, Belgium), Tillotts Pharma AG (Rheinfelden, Switzerland), Sanofi Aventis SA (Paris, France), Vifor Pharma (St Gallen, Switzerland), Abbott Laboratories Ltd (Chicago, IL, USA) and Procter and Gamble Ltd (Cincinnati, OH, USA). Simon Travis reports directorships of charities IBD2020 (Barnet, UK; UK 09762150), Cure Crohn's Colitis (Sydney, Australia; ABN 85 154 588 717) and the Truelove Foundation (London, UK; UK 11056711). Simon Travis also reports receiving fees from the following for expert testimony work and/or royalties: Santarus Inc. (San Diego, CA, USA), Cosmo Technologies Ltd (Dublin, Ireland), Tillotts Pharma AG, Wiley-Blackwell Inc. (Hoboken, NJ, USA), Elsevier Ltd (Amsterdam, the Netherlands) and Oxford University Press (Oxford, UK). Simon Travis has received research grants from the following: AbbVie Inc., the International Organization for the Study of Inflammatory Bowel Disease, Eli Lilly and Company, UCB Inc. (Brussels, Belgium), Vifor Pharma, Norman Collisson Foundation (Bicester, UK), Ferring Pharmaceuticals, Schering-Plough, Merck Sharpe & Dohme Corp. (Kenilworth, NJ, USA), Procter and Gamble Ltd, Warner Chilcott Ltd, Abbott Laboratories Ltd, PDL BioPharma Inc. (Incline Village, NV, USA), Takeda UK Ltd and the International Consortium for Health Care Outcomes Measurement. Ailsa Hart reports personal fees from AbbVie Inc., Atlantic Healthcare Ltd (Saffron Walden, UK), Bristol-Myers Squibb, Celltrion Inc. (Incheon, South Korea), Dr Falk Pharma UK Ltd (Bourne End, UK), Ferring Pharmaceuticals, Janssen Pharmaceutica, Merck Sharpe & Dohme Corp., Napp Pharmaceuticals Ltd (Cambridge, UK), Pfizer Inc., Pharmacosmos A/S (Holbæk, Denmark), Shire Pharmaceuticals UK and Takeda UK Ltd, and non-financial support from Genentech Inc. Alastair Windsor reports personal fees from Takeda, grants from Allergan Inc. (Dublin, Ireland), personal fees from Allergan, personal fees from Cook Medical Inc. (Bloomington, IN, USA) and grants and personal fees from Bard Ltd (Crawley, UK) outside the submitted work. Andrew Plumb reports grants from the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme outside the submitted work, grants from the NIHR Fellowships programme during the conduct of the study and honoraria for educational lectures delivered at events

arranged by Acely Inc. (Crawley, UK), Actavis Pharma Inc. (Parsippany-Troy Hills, NJ, USA), Dr Falk Pharma UK Ltd, Janssen-Cilag Ltd (High Wycombe, UK) and Takeda UK Ltd on the subject of inflammatory bowel disease. Ilan Jacobs reports share ownership in General Electric Company (Boston, MA, USA), which manufactures and sells magnetic resonance imaging equipment. Charles D Murray reports personal fees from AbbVie Inc., Merck Sharpe & Dohme Corp. and Janssen Pharmaceutica outside the submitted work. Antony Higginson reports personal fees from Toshiba Corporation (Tokyo, Japan) outside the submitted work. Steve Halligan reports non-financial support from iCAD Inc. (Nashua, NH, USA) outside the submitted work, and sat on the HTA commissioning board (2008–14). Stephen Morris reports Health Services and Delivery Research (HSDR) Board membership (2014–18), HSDR Evidence Synthesis Sub Board membership (2016), HTA Commissioning Board membership (2009–13) and Public Health Research Board membership (2011–17).

Published August 2019

DOI: 10.3310/hta23420

This report should be referenced as follows:

Taylor SA, Mallett S, Bhatnagar G, Morris S, Quinn L, Tomini F, *et al*. Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study. *Health Technol Assess* 2019;**23**(42).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nhr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nhr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nhr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nhr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/23/01. The contractual start date was in September 2013. The draft report began editorial review in May 2018 and was accepted for publication in December 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded 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, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Taylor *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nhr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Honorary Professor, University of Manchester, and Senior Clinical Researcher and Associate Professor, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study

Stuart A Taylor,^{1*} Sue Mallett,² Gauraang Bhatnagar,¹ Stephen Morris,³ Laura Quinn,² Florian Tomini,³ Anne Miles,⁴ Rachel Baldwin-Cleland,⁵ Stuart Bloom,⁶ Arun Gupta,⁵ Peter John Hamlin,⁷ Ailsa L Hart,⁸ Antony Higginson,⁹ Ilan Jacobs,¹⁰ Sara McCartney,⁶ Charles D Murray,¹¹ Andrew AO Plumb,¹ Richard C Pollok,¹² Manuel Rodriguez-Justo,¹³ Zainib Shabir,¹⁴ Andrew Slater,¹⁵ Damian Tolan,¹⁶ Simon Travis,¹⁷ Alastair Windsor,¹⁸ Peter Wylie,¹⁹ Ian Zealley²⁰ and Steve Halligan¹

¹Centre for Medical Imaging, University College London, London, UK

²Institute of Applied Health Research, National Institute for Health Research Birmingham Biomedical Research Centre, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

³Applied Health Research, University College London, London, UK

⁴Department of Psychological Sciences, Birkbeck, University of London, London, UK

⁵Intestinal Imaging Centre, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK

⁶Department of Gastroenterology, University College Hospital, London, UK

⁷Department of Gastroenterology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

⁸Inflammatory Bowel Disease Unit, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK

⁹Department of Radiology, Portsmouth Hospitals NHS Trust, Portsmouth, UK

¹⁰Independent patient representative, c/o Centre for Medical Imaging, University College London, London, UK

¹¹Department of Gastroenterology and Endoscopy, Royal Free London NHS Foundation Trust, London, UK

¹²Department of Gastroenterology, St George's Hospital, London, UK

¹³Department of Histopathology, University College Hospital, London, UK

¹⁴Comprehensive Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

¹⁵Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹⁶Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

¹⁷Translational Gastroenterology Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹⁸Department of Surgery, University College Hospital, London, UK

¹⁹Department of Radiology, Royal Free London NHS Foundation Trust, London, UK

²⁰Department of Radiology, Ninewells Hospital, Dundee, UK

*Corresponding author stuart.taylor@ucl.ac.uk

Background: Magnetic resonance enterography and enteric ultrasonography are used to image Crohn's disease patients. Their diagnostic accuracy for presence, extent and activity of enteric Crohn's disease was compared.

Objective: To compare diagnostic accuracy, observer variability, acceptability, diagnostic impact and cost-effectiveness of magnetic resonance enterography and ultrasonography in newly diagnosed or relapsing Crohn's disease.

Design: Prospective multicentre cohort study.

Setting: Eight NHS hospitals.

Participants: Consecutive participants aged ≥ 16 years, newly diagnosed with Crohn's disease or with established Crohn's disease and suspected relapse.

Interventions: Magnetic resonance enterography and ultrasonography.

Main outcome measures: The primary outcome was per-participant sensitivity difference between magnetic resonance enterography and ultrasonography for small bowel Crohn's disease extent. Secondary outcomes included sensitivity and specificity for small bowel Crohn's disease and colonic Crohn's disease extent, and sensitivity and specificity for small bowel Crohn's disease and colonic Crohn's disease presence; identification of active disease; interobserver variation; participant acceptability; diagnostic impact; and cost-effectiveness.

Results: Out of the 518 participants assessed, 335 entered the trial, with 51 excluded, giving a final cohort of 284 (133 and 151 in new diagnosis and suspected relapse cohorts, respectively). Across the whole cohort, for small bowel Crohn's disease extent, magnetic resonance enterography sensitivity [80%, 95% confidence interval (CI) 72% to 86%] was significantly greater than ultrasonography sensitivity (70%, 95% CI 62% to 78%), with a 10% difference (95% CI 1% to 18%; $p = 0.027$). For small bowel Crohn's disease extent, magnetic resonance enterography specificity (95%, 95% CI 85% to 98%) was significantly greater than ultrasonography specificity (81%, 95% CI 64% to 91%), with a 14% difference (95% CI 1% to 27%). For small bowel Crohn's disease presence, magnetic resonance enterography sensitivity (97%, 95% CI 91% to 99%) was significantly greater than ultrasonography sensitivity (92%, 95% CI 84% to 96%), with a 5% difference (95% CI 1% to 9%). For small bowel Crohn's disease presence, magnetic resonance enterography specificity was 96% (95% CI 86% to 99%) and ultrasonography specificity was 84% (95% CI 65% to 94%), with a 12% difference (95% CI 0% to 25%). Test sensitivities for small bowel Crohn's disease presence and extent were similar in the two cohorts. For colonic Crohn's disease presence in newly diagnosed participants, ultrasonography sensitivity (67%, 95% CI 49% to 81%) was significantly greater than magnetic resonance enterography sensitivity (47%, 95% CI 31% to 64%), with a 20% difference (95% CI 1% to 39%). For active small bowel Crohn's disease, magnetic resonance enterography sensitivity (96%, 95% CI 92% to 99%) was significantly greater than ultrasonography sensitivity (90%, 95% CI 82% to 95%), with a 6% difference (95% CI 2% to 11%). There was some disagreement between readers for both tests. A total of 88% of participants rated magnetic resonance enterography as very or fairly acceptable, which is significantly lower than the percentage (99%) of participants who did so for ultrasonography. Therapeutic decisions based on magnetic resonance enterography alone and ultrasonography alone agreed with the final decision in 122 out of 158 (77%) cases and 124 out of 158 (78%) cases, respectively. There were no differences in costs or quality-adjusted life-years between tests.

Limitations: Magnetic resonance enterography and ultrasonography scans were interpreted by practitioners blinded to clinical data (but not participant cohort), which does not reflect use in clinical practice.

Conclusions: Magnetic resonance enterography has higher accuracy for detecting the presence, extent and activity of small bowel Crohn's disease than ultrasonography does. Both tests have variable interobserver agreement and are broadly acceptable to participants, although ultrasonography produces less participant burden. Diagnostic impact and cost-effectiveness are similar. Recommendations for future work include investigation of the comparative utility of magnetic resonance enterography and ultrasonography for treatment response assessment and investigation of non-specific abdominal symptoms to confirm or refute Crohn's disease.

Trial registration: Current Controlled Trials ISRCTN03982913.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 23, No. 42. See the NIHR Journals Library website for further project information.

Contents

List of tables	xvii
List of figures	xxi
List of supplementary material	xxiii
List of abbreviations	xxv
Plain English summary	xxvii
Scientific summary	xxix
Chapter 1 Introduction	1
Background	1
Radiological imaging of Crohn's disease	1
Existing literature on the diagnostic accuracy of magnetic resonance enterography and ultrasonography	2
Imaging of Crohn's disease in the NHS	2
Objectives of the METRIC trial	5
Chapter 2 Methods	7
Study design	7
Patient and public involvement	7
Recruitment sites	8
Participants	8
Inclusion criteria	9
<i>New diagnosis cohort</i>	9
<i>Suspected relapse cohort</i>	9
Exclusion criteria	9
Test methods	9
<i>Magnetic resonance enterography</i>	9
<i>Ultrasonography</i>	10
<i>Clinical investigations</i>	10
<i>Radiologist/sonographer competence and training</i>	10
<i>Study imaging interpretation and reporting</i>	11
<i>Generation of additional small bowel imaging for discrepant magnetic resonance enterography and ultrasonography</i>	12
Recruitment	12
Data collection and participant follow-up	13
Reference standard	13
Outcomes	14
Sample size	15
<i>Primary outcome</i>	15
<i>Secondary outcomes</i>	15
Analysis	15

Chapter 3 Results	17
Participants	17
Baseline participant characteristics	17
Consensus reference standard	20
Test results and outcomes	22
<i>Identification and localisation of small bowel and colonic Crohn's disease against the consensus reference standard</i>	22
<i>Identification and localisation of small bowel Crohn's disease and colonic Crohn's disease against an ileocolonoscopy reference</i>	27
<i>Extraenteric complications</i>	27
<i>Disease activity assessment against the consensus reference standard</i>	30
<i>Disease activity assessment against an ileocolonoscopy reference</i>	30
<i>Equivocal magnetic resonance enterography and ultrasonography findings</i>	30
<i>Equivocal magnetic resonance enterography and ultrasonography findings: sensitivity analysis for the primary outcome</i>	30
<i>Perceptual errors on magnetic resonance enterography and ultrasonography</i>	30
<i>Identification and localisation of small bowel and colonic Crohn's disease against the consensus reference standard according to recruitment site</i>	32
Chapter 4 Diagnostic benefit of oral contrast administration (small intestine contrast-enhanced ultrasonography)	35
Introduction	35
Methods	35
Outcomes	36
Results	36
Chapter 5 Interobserver variation in the interpretation of enteric ultrasonography and magnetic resonance enterography	39
Introduction	39
Methods	39
<i>Ultrasonography</i>	39
<i>Magnetic resonance enterography</i>	40
Analysis	40
<i>Ultrasonography</i>	40
<i>Magnetic resonance enterography</i>	41
Outcomes	41
Results	41
<i>Ultrasonography</i>	41
<i>Magnetic resonance enterography</i>	43
Chapter 6 Influence of sequence selection on magnetic resonance enterography diagnostic accuracy	45
Introduction	45
Methods	45
Analysis	46
Outcomes	46
Results	46

Chapter 7 Magnetic resonance enterography and enteric ultrasonography to diagnose Crohn's disease: participant acceptability, perceived burden and preferences	49
Introduction	49
Methods	49
<i>Participants</i>	49
<i>Questionnaire content</i>	49
<i>Scan preference</i>	50
<i>Importance of scan attributes</i>	50
<i>Statistical analysis</i>	50
Results	50
<i>Burden score reliability</i>	52
<i>Test/scan recovery time</i>	52
<i>Test/scan acceptability</i>	52
<i>Willingness to repeat investigations</i>	53
<i>Least acceptable part of the investigations</i>	53
<i>Burden of magnetic resonance enterography and ultrasonography</i>	54
Chapter 8 Influence of oral contrast agent and ingested volume on small bowel distension and participant experience during magnetic resonance enterography	57
Introduction	57
Methods	57
<i>Participant experience of oral contrast</i>	57
<i>Grading of bowel distension quality on magnetic resonance enterography</i>	57
Analysis	58
Outcomes	58
Results	58
Chapter 9 The impact of magnetic resonance enterography and ultrasonography on diagnostic confidence and patient management	63
Introduction	63
Methods	64
<i>Participant selection</i>	64
<i>Data presentation and case report form completion</i>	64
Analysis	65
Results	66
Chapter 10 Cost–utility analysis of magnetic resonance enterography versus ultrasonography for small bowel Crohn's disease	69
Introduction	69
Methods	69
<i>Overview of economic evaluation</i>	69
<i>Diagnostic and treatment stages</i>	69
<i>Diagnostic versus therapeutic impacts</i>	70
<i>Diagnostic outcomes</i>	70
<i>Measuring costs</i>	70
<i>Measuring quality-adjusted life-years</i>	72
<i>Dealing with missing data</i>	73
<i>Measuring cost-effectiveness</i>	73
<i>Sensitivity analyses</i>	74

Results	75
<i>Diagnostic outcomes</i>	75
<i>Resource use and costs at the treatment stage</i>	75
<i>Utility values</i>	78
<i>Cost–utility analysis</i>	78
<i>Sensitivity analyses</i>	79
Chapter 11 Discussion	81
Main trial	81
Diagnostic benefit of oral contrast administration: small intestine contrast-enhanced ultrasonography	85
Interobserver variation in the interpretation of enteric ultrasonography and magnetic resonance enterography	86
Influence of sequence selection on magnetic resonance enterography diagnostic accuracy	87
Magnetic resonance enterography and ultrasonography to diagnose Crohn’s disease: participant acceptability, perceived burden and preferences	88
Influence of oral contrast agent and ingested volume on small bowel distension and participant experience during magnetic resonance enterography	89
Diagnostic impact	90
Cost–utility analysis of magnetic resonance enterography versus ultrasonography for small bowel Crohn’s disease	91
Overall conclusions	92
Implications for practice	92
Recommendations for future research	92
Acknowledgements	95
References	99
Appendix 1 Magnetic resonance enterography sequence protocol	109
Appendix 2 Magnetic resonance enterography oral contrast agent by recruitment site	111
Appendix 3 Summary of study outcomes by disease site and cohort	113
Appendix 4 Definition of agreement with reference standard for disease extent	115
Appendix 5 Recruitment sites and recruitment numbers and withdrawals per site	117
Appendix 6 Presenting symptoms of the final study cohort	119
Appendix 7 Presence and activity of small bowel and colonic Crohn’s disease according to individual bowel segments	121
Appendix 8 Quality of segmental visualisation on magnetic resonance enterography and ultrasonography according to the participant cohort	123
Appendix 9 Imaging and endoscopic data available to the consensus panel in participants with discrepancy for small bowel disease presence or location between magnetic resonance enterography and ultrasonography	125
Appendix 10 Raw data for the primary outcome and selected secondary outcomes	127

Appendix 11 Potential impact of staging small bowel Crohn's disease presence with either magnetic resonance enterography and ultrasonography in a theoretical 1000-participant cohort	131
Appendix 12 Per-segment sensitivity and specificity for disease presence and extent against the consensus reference standard, according to participant cohort	133
Appendix 13 Sensitivity and specificity for terminal ileal and colonic Crohn's disease presence and extent (regardless of activity) versus ileocolonoscopy reference, according to participant cohort	135
Appendix 14 Per-participant sensitivity and specificity the presence of active disease against the consensus reference standard, according to participant cohort	137
Appendix 15 Sensitivity and specificity for presence of active terminal ileal and colonic Crohn's disease versus ileocolonoscopy reference, according to participant cohort	139
Appendix 16 Perceptual errors for disease detection (regardless of activity) on magnetic resonance enterography and ultrasonography against the consensus reference (both cohorts combined)	141
Appendix 17 Per-reader sensitivity and specificity for small bowel Crohn's disease extent (ultrasonography)	143
Appendix 18 Per-reader sensitivity and specificity for small bowel Crohn's disease extent (magnetic resonance enterography)	145
Appendix 19 Participant symptom record sheet following ingestion of oral contrast prior to magnetic resonance enterography	147
Appendix 20 Bar chart of participant symptoms according to the oral contrast agent	149
Appendix 21 Resource use by diagnostic outcome and unit costs	151
Appendix 22 Mean Crohn's disease management costs and utilities per participant: complete-case analysis with no imputed data	157
Appendix 23 Incremental cost-effectiveness of magnetic resonance enterography versus ultrasonography	159
Appendix 24 Mean Crohn's disease utilities per participant	161

List of tables

TABLE 1 Summary of recent meta-analyses and systematic reviews reporting the diagnostic accuracy of MRE and US for SBCD	3
TABLE 2 Reasons for exclusion of screened patients	17
TABLE 3 Reasons for participant withdrawal	18
TABLE 4 Demographics of final study cohort	19
TABLE 5 Range of investigations available to the consensus reference panel	20
TABLE 6 Per-participant disease presence and activity according to the consensus reference standard	22
TABLE 7 Per-participant sensitivity and specificity for disease presence and extent against the consensus reference standard (both cohorts combined)	23
TABLE 8 Per-segment sensitivity and specificity for disease presence against the consensus reference standard (both cohorts combined)	26
TABLE 9 Sensitivity and specificity for disease presence and extent against the consensus reference standard according to participant cohort	28
TABLE 10 Sensitivity and specificity for terminal ileal and colonic CD presence against an ileocolonoscopy reference (both cohorts combined)	29
TABLE 11 Per-participant sensitivity and specificity for the presence of active disease against the consensus reference standard (both cohorts combined)	31
TABLE 12 Sensitivity and specificity for presence of active terminal ileal and colonic CD vs. ileocolonoscopy reference (both cohorts combined)	31
TABLE 13 Number of equivocal MRE and US findings for per-participant disease presence (both cohorts combined)	32
TABLE 14 Number of equivocal MRE and US findings for per-participant disease activity (both cohorts combined)	32
TABLE 15 Per-participant sensitivity and specificity for SBCD extent against the consensus reference standard (both cohorts combined)	33
TABLE 16 Per-participant sensitivity and specificity for SBCD extent against the consensus reference standard according to recruitment site (both cohorts combined)	33
TABLE 17 Per-participant sensitivity and specificity of SICUS and conventional US for disease presence and extent against the consensus reference standard (both cohorts combined)	37

TABLE 18 Ultrasonography interobserver variability	42
TABLE 19 Magnetic resonance enterography interobserver variability	44
TABLE 20 Sensitivity and specificity for SBCD extent according to MRE sequence block	46
TABLE 21 Influence of diffusion-weighted imaging and contrast-enhanced imaging on radiologist diagnosis and confidence using MRE (both cohorts combined)	47
TABLE 22 Characteristics of participants in the questionnaire study	51
TABLE 23 Comparative scan experience of MRE and US	52
TABLE 24 Perceived participant scan burden for MRE and US	54
TABLE 25 Number of segments scored as excellent or good quality for distension according to the two observers (both oral contrast agent types combined)	58
TABLE 26 Number of segments achieving good (score of 3) or excellent (score of 4) distension according to oral contrast agent	59
TABLE 27 Participant symptoms according to oral contrast agent	59
TABLE 28 Participant symptoms according to volume of oral contrast agent ingested	60
TABLE 29 Grouping of therapeutic options	66
TABLE 30 Concordance of therapeutic decisions based on MRE or US compared with the reference standard decision based on all information	67
TABLE 31 Descriptive analysis of participant therapeutic decisions based on imaging for MRE against all tests, grouped by therapeutic options category	68
TABLE 32 Descriptive analysis of participant therapeutic decisions based on imaging for US against all tests, grouped by therapeutic options category	68
TABLE 33 Diagnostic accuracy matrices (new diagnosis cohort): MRE vs. consensus reference standard (133 participants)	75
TABLE 34 Diagnostic accuracy matrices (new diagnosis cohort): US vs. consensus reference standard (133 participants)	76
TABLE 35 Diagnostic accuracy matrices (suspected relapse cohort): MRE vs. consensus reference standard (151 participants)	76
TABLE 36 Diagnostic accuracy matrices (suspected relapse cohort): US vs. consensus reference standard (151 participants)	76
TABLE 37 Mean CD management costs and utilities per participant at the treatment stage	77

TABLE 38 Incremental cost-effectiveness of MRE vs. US	78
TABLE 39 Univariate sensitivity analyses: cost-effectiveness of MRE vs. US	80
TABLE 40 Magnetic resonance enterography sequence protocol	109
TABLE 41 Magnetic resonance enterography oral contrast agent by recruitment site	111
TABLE 42 Summary of study outcomes by disease site and cohort	113
TABLE 43 Definition of agreement with reference standard for disease extent	115
TABLE 44 Recruitment sites, recruitment numbers and withdrawals per site	118
TABLE 45 Presenting symptoms of the final study cohort	119
TABLE 46 Presence and activity of SBCD and colonic CD according to individual bowel segments	121
TABLE 47 Quality of segmental visualisation on MRE and US (new diagnosis cohort)	123
TABLE 48 Quality of segmental visualisation on MRE and US (suspected relapse cohort)	124
TABLE 49 Imaging and endoscopic data available to the consensus panel in participants with discrepancy for SBCD presence or location between MRE and US	125
TABLE 50 Raw data for the primary outcome and selected secondary outcomes	127
TABLE 51 Per-segment sensitivity and specificity for disease presence and extent against the consensus reference standard according to participant cohort	134
TABLE 52 Sensitivity and specificity for terminal ileal and colonic CD presence and extent (regardless of activity) vs. ileocolonoscopy reference, according to participant cohort	136
TABLE 53 Per-participant sensitivity and specificity for the presence of active disease against the consensus reference standard, according to participant cohort	138
TABLE 54 Sensitivity and specificity for the presence of active terminal ileal and colonic CD vs. ileocolonoscopy reference, according to participant cohort	140
TABLE 55 Perceptual errors for disease detection (regardless of activity) on MRE and US against the consensus reference (both cohorts combined)	141
TABLE 56 Per-reader sensitivity and specificity for SBCD extent (US)	143
TABLE 57 Per-reader sensitivity and specificity for SBCD extent (MRE)	145
TABLE 58 Participant symptom record sheet following ingestion of oral contrast prior to MRE	147

TABLE 59 Resource use by diagnostic outcome and unit costs	152
TABLE 60 Mean CD management costs and utilities per participant: complete-case analysis with no imputed data	158
TABLE 61 Incremental cost-effectiveness of MRE vs. US	159
TABLE 62 Mean CD utilities per participant: complete-case analysis with no imputed data using the EQ-5D-5L value set for England	161

List of figures

FIGURE 1 Participant flow through the main study	8
FIGURE 2 Participant flow diagram	18
FIGURE 3 Sensitivity and specificity of MRE and US for SBCD and colonic CD extent and presence against the consensus reference standard	24
FIGURE 4 Potential impact of staging SBCD extent with either MRE or US in a theoretical 1000-participant cohort	25
FIGURE 5 Flow chart showing the flow of participants in the study participating in the participant acceptability substudy	51
FIGURE 6 Least acceptable part of MRE	53
FIGURE 7 Least acceptable part of US	53
FIGURE 8 Perceived importance of different scan attributes	55
FIGURE 9 Diagnostic impact study flow	65
FIGURE 10 Cost-effectiveness acceptability curves showing the probability that MRE and US are cost-effective at different values of the cost-effectiveness threshold	79
FIGURE 11 Potential impact of staging SBCD presence with either MRE or US in a theoretical 1000-participant cohort	132
FIGURE 12 Bar chart of participant symptom tolerability by oral contrast type	150

List of supplementary material

Report Supplementary Material 1 Background to power calculation

Supplementary material can be found on the NIHR Journals Library report project page (www.journalslibrary.nihr.ac.uk/programmes/hta/112301/#/documentation).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

BaFT	barium small bowel follow-through	HRQoL	health-related quality of life
BSGAR	British Society of Gastrointestinal and Abdominal Radiology	HTA	Health Technology Assessment
CapE	capsule endoscopy	IBD	inflammatory bowel disease
CCTU	Comprehensive Clinical Trials Unit	INMB	incremental net monetary benefit
CD	Crohn's disease	IQR	interquartile range
CI	confidence interval	MDT	multidisciplinary team
CRF	case report form	METRIC	Magnetic Resonance Enterography or uTRAsound In Crohn's disease
CrI	credibility interval	MRE	magnetic resonance enterography
CRP	C-reactive protein	MRI	magnetic resonance imaging
CT	computed tomography	NIHR	National Institute for Health Research
CTE	computed tomography enterography	NMB	net monetary benefit
DCTI	diagnostic confidence and therapeutic impact	PABAK	prevalence-adjusted bias-adjusted kappa
ECCO	European Crohn's and Colitis Organisation	PACS	picture archiving and communications system
EQ-5D-5L	EuroQol-5 Dimensions, five-level version	PC	personal computer
ESGAR	European Society of Gastrointestinal and Abdominal Radiology	PSA	probabilistic sensitivity analysis
FC	faecal calprotectin	QALY	quality-adjusted life-year
FRCR	Fellowship of the Royal College of Radiologists	SBCD	small bowel Crohn's disease
GBP	Great British pounds	SICUS	small intestine contrast-enhanced ultrasonography
GHQ-12	General Health Questionnaire-12 items	TMG	Trial Management Group
GP	general practitioner	UCL	University College London
HBI	Harvey-Bradshaw Index	UCLH	University College Hospital
		US	ultrasonography

Plain English summary

Crohn's disease is a waxing and waning lifelong inflammatory condition that affects the colon (large bowel) and small bowel. Treatment relies on accurately determining disease extent and underlying inflammation. Colonoscopy is very good for examining the colon, but it is invasive and, at best, can only visualise a few centimetres of the small bowel, so radiological imaging is very important. Magnetic resonance enterography (a type of magnetic resonance imaging scan) and ultrasonography are both radiological tests commonly performed in the NHS, and it is unclear which method is better. We performed a study to compare the accuracy of magnetic resonance enterography and ultrasonography for determining the extent of Crohn's disease in the bowel of participants newly diagnosed and in those participants with established Crohn's disease but with suspected deterioration. We also investigated how often radiologists agree with each other during test interpretation, the participant experience of undergoing the tests and their cost-effectiveness. We compared the tests in 284 participants (133 newly diagnosed and 151 with suspected deterioration). We found that both tests were accurate for detecting the presence (97% for magnetic resonance enterography and 92% for ultrasonography) and location (80% for magnetic resonance enterography and 70% for ultrasonography) of disease in the small bowel, but magnetic resonance enterography was better than ultrasonography for both (correctly classifying disease extent in 107 more participants for every 1000 participants with Crohn's disease). Magnetic resonance enterography was similarly better than ultrasonography at determining if the bowel was inflamed. The results were similar in newly diagnosed participants and those participants with suspected deterioration. Agreement between radiologists interpreting the same images was, at best, moderate for both tests. A total of 88% of participants tolerated magnetic resonance enterography well or fairly well, which was less than the percentage (99%) of participants who tolerated ultrasonography well or fairly well. Both tests had a similar effect on the treatment decisions made by doctors. Both tests were also similar in their value for money for the NHS.

Scientific summary

Background

Crohn's disease is a chronic inflammatory bowel disease predominantly affecting the young and requiring lifelong therapy. The small bowel and/or colon are most commonly affected and diagnosis is based on a combination of clinical features, together with endoscopic, histopathological, biochemical and imaging findings. Management is contingent on disease extent, activity and the presence of extraluminal complications. Radiological imaging is fundamental to the comprehensive phenotyping of Crohn's disease, particularly in the small bowel, which is beyond the reach of colonoscopy. Several small bowel imaging techniques are currently utilised in the NHS, but ultrasonography and magnetic resonance enterography are favoured as they do not expose patients to ionising radiation. Small bowel ultrasonography has been established for many years and utilises standard technology widely available in the NHS, but uptake has been hampered by perceptions of reduced accuracy and concerns about operator dependence. Magnetic resonance enterography is a more recent innovation, requiring oral contrast and access to high-technology imaging platforms, which are comparatively restricted in many health-care settings. Although meta-analyses suggest that magnetic resonance enterography and ultrasonography have similar accuracy for diagnosing and staging Crohn's disease, the primary literature is of questionable quality and prospective multicentre comparative data are lacking. High-quality evidence is needed to guide implementation.

Objectives

The primary objective of the Magnetic Resonance Enterography or ulTRasound In Crohn's disease (METRIC) trial was to compare the diagnostic accuracy of magnetic resonance enterography and ultrasonography for the extent of small bowel Crohn's disease against a construct reference standard, incorporating 6 months of participant follow-up. We recruited from two cohorts of participants: those newly diagnosed and those with established Crohn's disease clinically suspected of luminal relapse. The secondary objectives were to investigate the accuracy in grading inflammatory activity of magnetic resonance enterography and ultrasonography, diagnostic accuracy in the colon of magnetic resonance enterography and ultrasonography, interobserver variability in interpretation of magnetic resonance enterography and ultrasonography, diagnostic impact of an oral contrast load prior to small intestine contrast-enhanced ultrasonography, effect of oral contrast agent type on magnetic resonance enterography distension quality and participant symptoms, impact of magnetic resonance enterography sequences on diagnostic accuracy and radiologist confidence, impact of ultrasonography and magnetic resonance enterography on clinician decision-making, and participant experience of magnetic resonance enterography and ultrasonography, and to carry out a cost-effectiveness analysis.

Methods

Study design

We performed a multicentre prospective cohort study. Ethics committee approval was granted in 2014. Participants were recruited from eight representative UK NHS teaching and general hospitals. Participants were eligible for the new diagnosis subgroup if they had been diagnosed with Crohn's disease in the 3 months preceding recruitment, or if Crohn's disease was strongly suspected based on imaging or endoscopic features but pending final diagnosis. Participants for whom the final diagnosis was not Crohn's disease were subsequently excluded. Participants eligible for the suspected relapse subgroup had established Crohn's disease and high clinical suspicion of relapse based on objective markers of inflammatory activity and/or symptoms suggestive of luminal stenosis and/or abnormal endoscopy. Participants were ineligible if they

were aged < 16 years, were pregnant or had magnetic resonance imaging contraindications. Consecutive (i.e. unselected) eligible participants underwent magnetic resonance enterography and ultrasonography in accordance with the standard protocols of their recruitment sites, in addition to any other enteric imaging or endoscopic investigations performed as part of usual clinical care. Magnetic resonance enterography and ultrasonography image outputs 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 participant was recruited. Practitioners recorded the presence and activity of Crohn's disease in the small bowel and colon, together with any extraenteric complications. In total, 28 suitably experienced practitioners (27 radiologists and one sonographer) interpreted the magnetic resonance enterography and ultrasonography. If there was discrepancy between magnetic resonance enterography and ultrasonography for the presence or location of small bowel Crohn's disease, an 'arbiter' small bowel investigation was performed if one had not already been performed as part of usual care. We used the construct reference standard paradigm (multidisciplinary panel diagnosis) incorporating the concept of clinical test validation. Participants' clinical course was followed for 6 months. For each participant, the panel considered the results of all small bowel investigations and all additional information including endoscopies, surgical findings, histopathology and biochemical and clinical course and recorded its opinion on whether or not small bowel Crohn's disease or colonic Crohn's disease was present, and, if so, whether or not disease was active. We estimated that a sample size of 210 participants with small bowel Crohn's disease would give 90% power to detect a significant (10%) sensitivity difference between magnetic resonance enterography (83%) and ultrasonography (73%) for disease extent. Direct comparison of sensitivity and specificity differences between magnetic resonance enterography and ultrasonography in the same participants were calculated from paired data using bivariate multilevel participant-specific (conditional) random-effects models.

Analysis by segment used a population-averaged random-effects model (using logit including robust standard errors). In subsets of participants, we also investigated interobserver variability in interpretation of magnetic resonance enterography and ultrasonography, the impact of an oral contrast load prior to small intestine contrast-enhanced ultrasonography (compared with conventional ultrasonography) on diagnostic accuracy, the effect of magnetic resonance enterography oral contrast agent type on small bowel distension quality and participant symptoms, the impact of magnetic resonance enterography sequences (T2-weighted and steady-state free precession gradient echo images alone; T2-weighted and steady-state free precession gradient echo images and diffusion-weighted images; and T2-weighted and steady-state free precession gradient echo images, diffusion-weighted images and contrast-enhanced images) on radiologist accuracy, and participant experience of small bowel imaging using questionnaires, modelled the diagnostic impact of magnetic resonance enterography and ultrasonography on clinician decision-making and carried out a cost-effectiveness analysis.

Results

Overall, 518 patients were assessed for eligibility, of whom 183 were excluded. In total, 335 participants entered the trial, of whom 51 were excluded (31 did not have Crohn's disease, two were lost to follow-up, 10 did not undergo magnetic resonance enterography and/or ultrasonography, six withdrew consent and two newly diagnosed participants underwent surgery without colonoscopy). The final cohort consisted of 284 participants (133 and 151 in new diagnosis and suspected relapse cohorts, respectively). There were no reported adverse events. Based on the reference standard, 233 (82%) participants had small bowel Crohn's disease, which was active in 209 (90%) participants, 129 (45%) participants had colonic Crohn's disease, which was active in 126 (98%) participants, 21 participants had enteric fistulae and seven had intra-abdominal abscess. For small bowel Crohn's disease extent, magnetic resonance enterography sensitivity [80%, 95% confidence interval (CI) 72% to 86%] was significantly greater than ultrasonography sensitivity (70%, 95% CI 62% to 78%), with a difference of 10% (95% CI 1% to 18%; $p = 0.027$). For small bowel Crohn's disease extent, magnetic resonance enterography specificity (95%, 95% CI 85% to 98%) was also significantly greater than ultrasonography specificity (81%, 95% CI 64% to 91%), with a difference of 14% (95% CI 1% to 27%). For small bowel Crohn's disease presence, magnetic resonance

enterography sensitivity (97%, 95% CI 91% to 99%) was significantly greater than ultrasonography sensitivity (92%, 95% CI 84% to 96%), with a difference of 5% (95% CI 1% to 9%). For small bowel Crohn's disease presence, magnetic resonance enterography and ultrasonography specificity was 96% (95% CI 86% to 99%) and 84% (95% CI 65% to 94%), respectively. There were no overall significant differences in sensitivity or specificity between magnetic resonance enterography and ultrasonography for colonic Crohn's disease extent or presence. Sensitivities of magnetic resonance enterography and ultrasonography for small bowel Crohn's disease presence and extent in the new diagnosis and suspected relapse cohorts were very similar to those sensitivities estimated across all participants, although for colonic Crohn's disease presence ultrasonography sensitivity (67%, 95% CI 49% to 81%) was significantly greater than magnetic resonance enterography sensitivity (47%, 95% CI 31% to 64%) in the new diagnosis cohort, with a difference of 20% (95% CI 1% to 39%). For active small bowel Crohn's disease, magnetic resonance enterography sensitivity (96%, 95% CI 92% to 99%) was significantly greater than ultrasonography sensitivity (90%, 95% CI 82% to 95%), with a difference of 6% (95% CI 2% to 11%). Magnetic resonance enterography detected five out of seven (71%) abscesses and 18 out of 21 (86%) enteric fistulae, whereas ultrasonography detected three out of seven (43%) abscesses and 11 out of 21 (52%) enteric fistulae. A total of 186 participants had a colonoscopic standard of reference against which magnetic resonance enterography had a sensitivity of 97% (95% CI 91% to 99%) and ultrasonography has a sensitivity of 91% (95% CI 79% to 97%) for terminal ileal disease presence, which was not statistically significant.

Small intestine contrast-enhanced ultrasonography

In total, 64 participants underwent small intestine contrast-enhanced ultrasonography. Compared with ultrasonography, sensitivity and specificity were identical for small bowel Crohn's disease extent (71%, 95% CI 58% to 81%, and 86%, 95% CI 49% to 97%, respectively).

Interobserver agreement

Ultrasonography

In total, 38 participants (11 and 27 in the new diagnosis and suspected relapse cohorts, respectively) underwent ultrasonography twice, performed by two practitioners from a pool of seven. For the presence of small bowel Crohn's disease, both reads agreed with the consensus reference standard for 9 out of 11 (82%) participants in the new diagnosis cohort (prevalence-adjusted bias-adjusted $\kappa = 0.64$) and 22 out of 27 (81%) participants in the suspected relapse cohort (prevalence-adjusted bias-adjusted $\kappa = 0.63$), suggesting substantial agreement for both cohorts. Agreement for small bowel Crohn's disease extent was at best only fair [64% agreement with the consensus reference standard in the new diagnosis cohort (prevalence-adjusted bias-adjusted $\kappa = 0.27$) and 56% agreement with the consensus reference standard in the suspected relapse cohort (prevalence-adjusted bias-adjusted $\kappa = 0.11$)]. Agreement for the presence of colonic Crohn's disease was inferior to that for small bowel Crohn's disease.

Magnetic resonance enterography

A total of 73 magnetic resonance enterography data sets (28 in the new diagnosis cohort and 45 in the suspected relapse cohort) were each read by three radiologists from a pool of 27. For the presence of small bowel Crohn's disease, in the new diagnosis cohort, on average, reads agreed with the consensus reference standard for 18 out of 26 (69%) disease-positive participants and one out of two (50%) disease-negative participants (prevalence-adjusted bias-adjusted $\kappa = 0.36$), indicating fair agreement. In the suspected relapse cohort, on average, reads agreed with the consensus reference standard for 25 out of 33 (76%) disease-positive participants and 9 out of 12 (75%) disease-negative participants (prevalence-adjusted bias-adjusted $\kappa = 0.51$), indicating moderate agreement. Agreement for small bowel Crohn's disease extent was at best only slight [53% agreement with the consensus reference standard (prevalence-adjusted bias-adjusted $\kappa = 0.07$) in the suspected relapse cohort and 43% agreement with the consensus reference standard (prevalence-adjusted bias-adjusted $\kappa = -0.14$) in the new diagnosis cohort]. Agreement for the presence of colonic Crohn's disease was inferior to that for small bowel Crohn's disease.

Participant experience

A total of 159 participants completed all or part of the experience questionnaire. Overall, 88% (128/145) of participants rated magnetic resonance enterography as very or fairly acceptable, which is significantly lower than the percentage (99%; 144/146) of participants who rated ultrasonography as very or fairly acceptable. Participants were slightly less willing to undergo magnetic resonance enterography again (91%; 127/140) than to undergo ultrasonography again (99%; 133/135). On a scale of 1 (least) to 7 (most), test burden scores were generally low, although magnetic resonance enterography generated greater burden than ultrasonography (mean 2.72 vs. mean 1.66, respectively; $z = 9.5$). Younger age and emotional distress were associated with greater magnetic resonance enterography and ultrasonography burden. In a direct test comparison, the majority (80%; 100/125) preferred ultrasonography to magnetic resonance enterography, but participants prioritised diagnostic accuracy over scan burden.

Effect of oral contrast type

Full data were available for 66 participants (47 ingesting mannitol-based contrast agents and 19 ingesting polyethylene glycol-based contrast agents). There was no difference in segmental distension quality between the two contrast types. Distension of the jejunum, ileum and terminal ileum were rated as excellent or good in 27% (13/47), 65% (31/47) and 43% (20/47), respectively, of the mannitol group compared with 16% (3/19), 63% (12/19) and 42% (8/19), respectively, of the polyethylene glycol group. The participants' symptom profiles were also very similar between the two agents.

Sequence selection

A total of 73 magnetic resonance enterography data sets (28 in the new diagnosis cohort and 45 in the suspected relapse cohort) were each read by three radiologists from a pool of 27. There was no increase in sensitivity or specificity for small bowel Crohn's disease extent with the addition of diffusion-weighted images to T2-weighted and steady-state free precession gradient echo images. However, compared with T2-weighted and steady-state free precession gradient echo images alone (sensitivity for small bowel Crohn's disease extent 63%, 95% CI 51% to 75%), the addition of contrast-enhanced images led to a significant drop in sensitivity to 56% (95% CI 42% to 68%), a difference of 7% (95% CI -1% to -14%).

Modelling of the diagnostic impact of magnetic resonance enterography and ultrasonography

Out of 158 included participants, 74 were new diagnosis participants and 84 were suspected relapse participants. Therapeutic decisions based on clinical information and magnetic resonance enterography alone agreed with the final reference standard treatment decision category (based on all available information) for 122 out of 158 (77%) participants and disagreed for 36 out of 158 (23%) participants. Therapeutic decisions based on clinical information and ultrasonography alone agreed with the reference standard treatment decision category for 124 out of 158 (78%) participants and disagreed for 34 out of 158 (22%) participants.

Cost-effectiveness

There were no differences in costs or quality-adjusted life-years between magnetic resonance enterography and ultrasonography. For the new diagnosis cohort, the net monetary benefits for magnetic resonance enterography and ultrasonography were not significantly different, at £7288 (95% credibility interval £4797 to £9111) and £7513 (95% credibility interval £4936 to £9392), respectively, giving an incremental net monetary benefit for magnetic resonance enterography versus ultrasonography that was not significantly

different from zero (–£225, 95% credibility interval –£1085 to £713). For the suspected relapse cohort also, the net monetary benefits for magnetic resonance enterography and ultrasonography were not significantly different [£4020 (95% credibility interval £2426 to £5488) and £4321 (95% credibility interval £2696 to £5787), respectively], giving a non-significant incremental net monetary benefit for magnetic resonance enterography versus ultrasonography of –£301 (95% credibility interval –£993 to £305). Sensitivity analyses showed little uncertainty in these findings.

Conclusions

We found that both ultrasonography and magnetic resonance enterography achieve excellent diagnostic accuracy for the extent and activity of small bowel Crohn's disease in both new diagnosis and suspected relapse participants, but that magnetic resonance enterography has superior sensitivity for small bowel Crohn's disease presence, extent and activity.

Implications for health care

In a NHS setting, magnetic resonance enterography has significantly higher sensitivity and specificity than ultrasonography for the presence and extent of small bowel Crohn's disease, although both tests are valid first-line investigations. Ultrasonography performs better than magnetic resonance enterography for detection of colonic Crohn's disease in those patients newly diagnosed but is inferior to colonoscopy. Both magnetic resonance enterography and ultrasonography are deemed acceptable by the majority of patients, although ultrasonography induces less patient burden and is generally preferred. However, patients rank diagnostic accuracy as more important than test burden. Interobserver agreement is variable between practitioners, particularly for disease extent, which should be considered. We found no evidence that one oral contrast agent is better than another in achieving good bowel distension during magnetic resonance enterography or reducing patient symptom load. We also found no evidence that small intestine contrast-enhanced ultrasonography increases diagnostic accuracy for small bowel Crohn's disease extent compared with conventional ultrasonography. Addition of diffusion-weighted imaging does not improve the accuracy of magnetic resonance enterography protocols based only on T2-weighted steady-state free precession gradient echo images, and post-contrast images may be detrimental to sensitivity. Modelled diagnostic impact on clinician therapeutic strategy was similar between magnetic resonance enterography and ultrasonography. There is no reason to prefer magnetic resonance enterography or ultrasonography on economic grounds.

Recommendations for future research

Future research should investigate:

- the role of ultrasonography in targeted follow-up of Crohn's disease patients with an established disease phenotype, and the utility of magnetic resonance enterography and ultrasonography in treatment response assessment
- the most clinically effective and cost-effective cross-sectional imaging investigation in patients with non-specific abdominal symptoms to confirm or refute a diagnosis of Crohn's disease
- the impact of dedicated training programmes and clinical case volumes on practitioner accuracy.

Trial registration

This trial is registered as ISRCTN03982913.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

Background

Crohn's disease (CD) is a chronic inflammatory bowel disease (IBD) predominantly affecting the young and requiring lifelong medical and surgical therapy. Most patients are diagnosed before the age of 35 years and CD has an incidence of between 1.9 and 11.0 per 1000 person-years in western Europe.¹ According to a recent UK audit,² IBD accounts for 0.3% of work absenteeism, costs £115M in lost productivity and accounts for 27,000 hospital admissions annually. NHS expenditure on IBD is estimated to be > £1B, with each CD patient costing, on average, £6156 per year.³

Crohn's disease most commonly affects the small bowel and/or colon, with a range of manifestations, from superficial bowel wall ulceration to deep penetrating disease characterised by strictures, fistulae and abscesses. A host of potentially toxic medical treatments, such as immune modulators, and targeted surgical interventions are currently used to manage patients. Diagnosis is based on a combination of clinical features, together with endoscopic, histopathological, biochemical and imaging findings. Thereafter, patient management is contingent on the extent of disease, the underlying biological activity and the presence of extraluminal complications.

Colonoscopy is fundamental to diagnosis and follow-up of CD, given its exquisite views of the bowel mucosa and ability to take biopsies for histological analysis.⁴ However, colonoscopy is invasive, with a small but defined risk of serious complications, such as perforation, visualises only the bowel lumen and, at best, can interrogate only the last few centimetres of the small bowel (terminal ileum). Radiological imaging is therefore complementary for the diagnosis, staging and monitoring of CD, and can define disease presence, extent, biological activity and complications, particularly in the small bowel.⁵

Radiological imaging of Crohn's disease

Several small-bowel imaging investigations are currently utilised within the NHS, including barium small bowel follow-through (BaFT), computed tomography (CT), computed tomography enterography (CTE), ultrasonography (US) and magnetic resonance enterography (MRE). The various tests differ in their attributes. BaFT, for example, interrogates the bowel mucosa, whereas CT, US and MRE (cross-sectional imaging techniques) evaluate both the bowel wall and extraenteric tissues. An important differentiating attribute is the use or otherwise of ionising radiation. Both BaFT and CT impart a significant radiation dose, which is of concern given that CD patients are generally young and need repeat imaging over several years. A recent meta-analysis found that 11% of CD patients are exposed to potentially harmful doses of ionising radiation.⁶ Furthermore, exposure to diagnostic ionising radiation is increasing, largely owing to CT,^{7,8} despite technological developments reducing dose exposure.⁹ Conversely, neither US nor MRE impart ionising radiation and are therefore intuitively attractive modalities for imaging patients with CD.¹⁰ Indeed, international consensus guideline committees recommend MRE and US as the imaging modalities of choice in CD.⁵

Small bowel US has been established for many years¹¹ and is potentially well suited to imaging CD; it is non-invasive, well tolerated, requires no specific patient preparation and uses standard technology widely available in the NHS. However, uptake has been somewhat hampered by perceptions of reduced accuracy, particularly in the proximal small bowel,¹² and concerns about operator dependence and high levels of observer variability.¹³ Furthermore, interrogation of the bowel and deeper tissues may be limited by patient obesity, obscuring bowel gas or deep pelvic location. Conversely, MRE [magnetic resonance imaging (MRI) of the abdomen and pelvis following ingestion of an oral contrast agent] is a more recent innovation¹⁴

and an increasingly supportive evidence base has emerged over the last 15 years.⁵ MRE requires patients to ingest large volumes of oral contrast agent to distend the bowel, potentially resulting in abdominal cramps and diarrhoea, and utilises MRI technology, access to which is comparatively limited in the NHS. However, visualisation of the whole bowel is assured assuming a technically complete examination, and newer generations of radiologists are increasingly familiar with abdominal and pelvic MRI.

Existing literature on the diagnostic accuracy of magnetic resonance enterography and ultrasonography

We searched PubMed and EMBASE for articles published between 1 January 1990 and 1 December 2017 without language restriction. We used MeSH (medical subject heading) and full-text searches for 'Crohn's disease', 'magnetic resonance imaging', 'ultrasound' and 'diagnostic accuracy'. Emphasis was placed on meta-analyses and systematic reviews using appropriate search limits. There have been several systematic reviews and meta-analyses reporting the accuracy of MRE and US for diagnosing and staging CD.^{12,15–26} Many have considered MRE or US in isolation,^{15,17–20,22,23} whereas others have attempted to compare the two modalities (along with others, such as CT).^{12,21,25,26} Some have primarily focused on assessment of disease activity,^{17,22,26} with most assessing diagnostic accuracy. A summary of the main meta-analyses and systematic reviews undertaken since 2010 is given in *Table 1*.

These meta-analysis and systematic review data suggest essentially equivalent diagnostic accuracy for MRE and US in detection and staging of CD, with sensitivity and specificity generally > 80%. However, there is marked heterogeneity in the primary literature, with most contributory studies being single centre and recruiting relatively small participant numbers, typically < 50. Many studies are also retrospective and quality is generally poor. For example, in their 2017 meta-analysis, Liu *et al.*²³ reported that just 38% of included studies were rated as 'good quality' using the quality assessment of diagnostic accuracy studies (QUADAS) tool.²⁷ Similarly, using the modified tool (QUADAS-2),²⁸ Puylaert *et al.*²⁶ found that in six of their 19 included studies blinding to the reference standard was either not done or not explicitly stated. Most studies do not compare MRE and US in the same participants, which is the most effective design for diagnostic accuracy studies because, for example, it reduces bias caused by difference between participants.²⁹ There is also much variation in the applied standard of reference between studies, with endoscopy, surgery and imaging itself all employed. By way of example, in the largest study of US to date (296 participants),³⁰ the standard of reference was simple barium fluoroscopic studies in > 70% of recruited participants. Similarly, in their single-centre comparative study of MRE and US in 249 participants with suspected CD (120 with a confirmed diagnosis), Castiglione *et al.*³¹ used MRE itself as the standard of reference for small bowel Crohn's disease (SBCD) extent in the majority of participants who did not have surgical reference standard, clearly risking incorporation bias. There is no single reference standard for defining the location, extent and activity of CD. In such circumstances, the National Institute for Health Research (NIHR) acknowledges the advantages of the construct reference standard paradigm (panel diagnosis) and incorporating the concept of clinical test validation, that is, whether or not the results of an index test are meaningful in practice.³² Very few studies have used such a consensus panel standard of reference, which considers all available clinical, endoscopic and imaging data as well as patient outcomes.

Imaging of Crohn's disease in the NHS

According to a 2010 UK survey,³³ 90% of NHS radiology departments routinely perform BaFT to investigate known or suspected CD in patients, 80% perform CT, 56% perform US and 38% perform MRE. The use of MRE has certainly increased since this survey was conducted.

TABLE 1 Summary of recent meta-analyses and systematic reviews reporting the diagnostic accuracy of MRE and US for SBCD

Author	Year	Modalities considered	Number of included studies (participants)	Number of participants in largest contributing study (MRE or US)	Main outcome	Main findings
Liu <i>et al.</i> ²³	2017	CTE, MRE	21 (913)	72 (4 sites)	Diagnostic accuracy	CTE: sensitivity 0.87 (95% CI 0.78 to 0.92); specificity 0.91 (95% CI 0.84 to 0.95) MRI: sensitivity 0.86 (95% CI 0.79 to 0.91); specificity 0.93 (95% CI 0.84 to 0.97)
Greenup <i>et al.</i> ²¹	2016	CTE, MRE, US	21 (1135)	249 (120 with CD) (1 site)	Diagnostic accuracy	CTE: sensitivity 67–95%; specificity 70–90% MRE: sensitivity 66–100%; specificity 72–100% US: sensitivity 86–97%; specificity 83–97%
Choi <i>et al.</i> ¹⁶	2017	CapE, CTE, MRE, BaFT	24 (781)	89 (4 sites)	Diagnostic accuracy	Suspected CD: <ul style="list-style-type: none"> ● CE: sensitivity 62–100%; specificity 50–100% ● MRE: sensitivity 77%; specificity 80% ● BaFT: sensitivity 27.6%; specificity 100% Established CD: <ul style="list-style-type: none"> ● CE: sensitivity 77–100%; specificity 53–100% ● CTE: sensitivity 76–82%; specificity 85–89% ● MRE: sensitivity 77–81%; specificity 80–90%
Qiu <i>et al.</i> ²⁴	2015	CTE, MRE	6 (290)	73 (4 sites)	Diagnostic accuracy (active disease)	CTE: sensitivity 85.8% (95% CI 79.2% to 90.9%); specificity 83.6% (95% CI 75.3% to 90.1%) MRE: sensitivity 87.9% (95% CI 81.8% to 92.5%); specificity 81.2% (95% CI 71.9% to 88.4%)
Giles <i>et al.</i> ²⁰	2013	MRE	11 (496) (paediatric participants only)	87 (1 site)	Diagnostic accuracy	MRE: sensitivity (terminal ileum) 0.84 (95% CI 0.77 to 0.90); specificity 0.97 (0.91 to 0.99)

continued

TABLE 1 Summary of recent meta-analyses and systematic reviews reporting the diagnostic accuracy of MRE and US for SBCD (*continued*)

Author	Year	Modalities considered	Number of included studies (participants)	Number of participants in largest contributing study (MRE or US)	Main outcome	Main findings
Dong <i>et al.</i> ¹⁸	2014	US	15 (1558)	249 (120 with CD) (1 site)	Diagnostic accuracy	Sensitivity 88% (95% CI 85% to 91%); specificity 97% (95% CI 96% to 98%)
Panés <i>et al.</i> ¹²	2011	CTE, MRE, US	Locating active CD: <ul style="list-style-type: none"> • CTE: 1 (44) • MRE: 5 (180) • US: 8 (939) Assessment of disease activity: <ul style="list-style-type: none"> • CTE: 9 (216) • MRE: 6 (217) • US: 6 (207) 	296 (1 site)	Diagnostic accuracy (active disease)	Locating active CD: <ul style="list-style-type: none"> • CTE: sensitivity 88%; specificity 88% • MRE: sensitivity 74% (95% CI 68% to 80%); specificity 91% (95% CI 86% to 95%) • US: sensitivity 86% (95% CI 83% to 88%); specificity 94% (95% CI 93% to 95%) Assessment of disease activity: <ul style="list-style-type: none"> • CTE: sensitivity 81% (95% CI 77% to 86%); specificity 88% (95% CI 82% to 91%) • MRE: sensitivity (per participant) 80% (95% CI 77% to 83%); specificity 82% (95% CI 78% to 85%) • US: sensitivity 85% (95% CI 79 to 89%); specificity 91% (95% CI 87% to 95%)
Ahmed <i>et al.</i> ¹⁵	2016	MRE	19 (102)	249 (120 with CD) (1 site)	Disease activity	Sensitivity: 88% (95% CI 86% to 91%); specificity: 88% (95% CI 84% to 91%)
Puylaert <i>et al.</i> ²⁶	2015	CTE, MRE, scintigraphy, US	19 (549)	76 (1 site)	Disease activity	Per-participant accuracy: <ul style="list-style-type: none"> • CTE: 86% (95% CI 75% to 93%) • MRI: 84% (95% CI 67% to 93%) • Scintigraphy: 40% (95% CI 16% to 70%) • US: 44% (95% CI 28% to 61%)

CapE, capsule endoscopy; CI, confidence interval.

Across the NHS there is ad hoc provision and utilisation of newer imaging technologies in CD, with little consistency between hospitals and no coherent implementation strategy. The choice of small bowel imaging investigation currently depends largely on non-evidence-based decision-making, such as clinician personal preference, perceived costs, available infrastructure and radiological expertise.

Ultimately, the optimal imaging strategy for CD remains uncertain and single-centre data are of limited utility. Unbiased, robust data to inform the implementation strategy for newer imaging technologies are currently unavailable.

Objectives of the METRIC trial

The primary aim of the Magnetic Resonance Enterography or ultrasound In Crohn's disease (METRIC) trial was to compare the diagnostic accuracy of MRE and US for extent of SBCD against a construct reference standard, incorporating 6 months of participant follow-up. We recruited from two cohorts of participants: newly diagnosed participants and participants with established CD clinically suspected of luminal relapse. Secondary objectives included comparative accuracy of MRE and US in grading of inflammatory activity and diagnostic accuracy in the colon, interobserver variability in interpretation of MRE and US and a cost-effectiveness analysis. The impact of an oral contrast load prior to small intestine contrast-enhanced ultrasonography (SICUS) on diagnostic accuracy compared with conventional US was investigated in a subset of participants. We also modelled the diagnostic impact of both tests on clinician decision-making, investigated the influence of MRE sequence selection on accuracy and assessed participant experience of small bowel imaging.

Chapter 2 Methods

Parts of this chapter have been reproduced or adapted from Taylor *et al.*³⁴ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

Study design

The METRIC trial was a multicentre, non-randomised, single-arm, prospective cohort study comparing the diagnostic accuracy of MRE and enteric US for the presence, extent and activity of SBCD in newly diagnosed participants or participants with established disease and suspected relapse. The full protocol has been published.³⁴ The trial achieved NHS Research Ethics Committee approval in September 2013 (reference 13/SC/0394) and was conducted in accordance with the principles of Good Clinical Practice. The trial was supervised by University College London (UCL)'s Clinical Trials Unit, an independent Data Monitoring Committee and a Trial Steering Committee. All recruited participants gave written informed consent.

Consecutive (i.e. unselected) eligible participants underwent MRE and US in addition to standard investigations performed as part of their usual care, and their clinical course was followed for a period of 6 months. A multidisciplinary consensus panel derived the reference standard for the presence, location and biological activity of SBCD and colonic CD using all available clinical, endoscopic, imaging, biochemical and histological data over the 6-month follow-up period. A summary of participant flow in the main trial is shown in *Figure 1*. Agreement between radiologists' interpretations of MRE and US was tested in a sample of participants, and the contribution of specific MRE sequences on radiologist accuracy was investigated. The influence of oral contrast type and volume on quality of bowel distension during MRE was evaluated and the effect of an oral contrast load prior to SICUS on diagnostic accuracy compared with conventional US was tested in a sample of participants. An exercise was undertaken by gastroenterologists to assess the diagnostic and therapeutic impact of MRE and US on clinical decision-making, and comparative participant experience of MRE and US was evaluated. The cost-effectiveness of MRE and US in both participant cohorts was assessed in an economic evaluation. The full study protocol can be accessed via the UCL project page [URL: www.ucl.ac.uk/cctu/research-areas/gastroenterology/metric (accessed 29 May 2019)].

Patient and public involvement

The METRIC trial was developed in collaboration with Crohn's and Colitis UK [URL: <https://crohnsandcolitis.org.uk> (accessed 29 May 2019)], which nominated a patient representative to join the trial team at the inception of the project. The patient representative helped refine the research questions, devise the protocol and successfully applied for the funding. By way of example, on their advice and guidance, we included a detailed assessment of patient priorities for imaging their disease to better understand and interpret the research findings. All patient-facing materials were designed with the patient representative (and in general were very well received by participants). The representative sat on the Trial Management Committee and Trial Steering Committee, providing guidance throughout the running of the trial and subsequent write-up, for example helping to refine recruitment strategies and advising on dissemination. This collaboration has been very productive and has now been expanded with the set-up of a patient forum to advise on current and future imaging research in CD. The forum ensures a wide representation of opinion, including age, sex and disease focus, and has been supported by Bowel & Cancer Research [URL: www.bowelcancerresearch.org (accessed 29 May 2019)] and Motilent (London, UK), which supported (in kind) meetings costs.

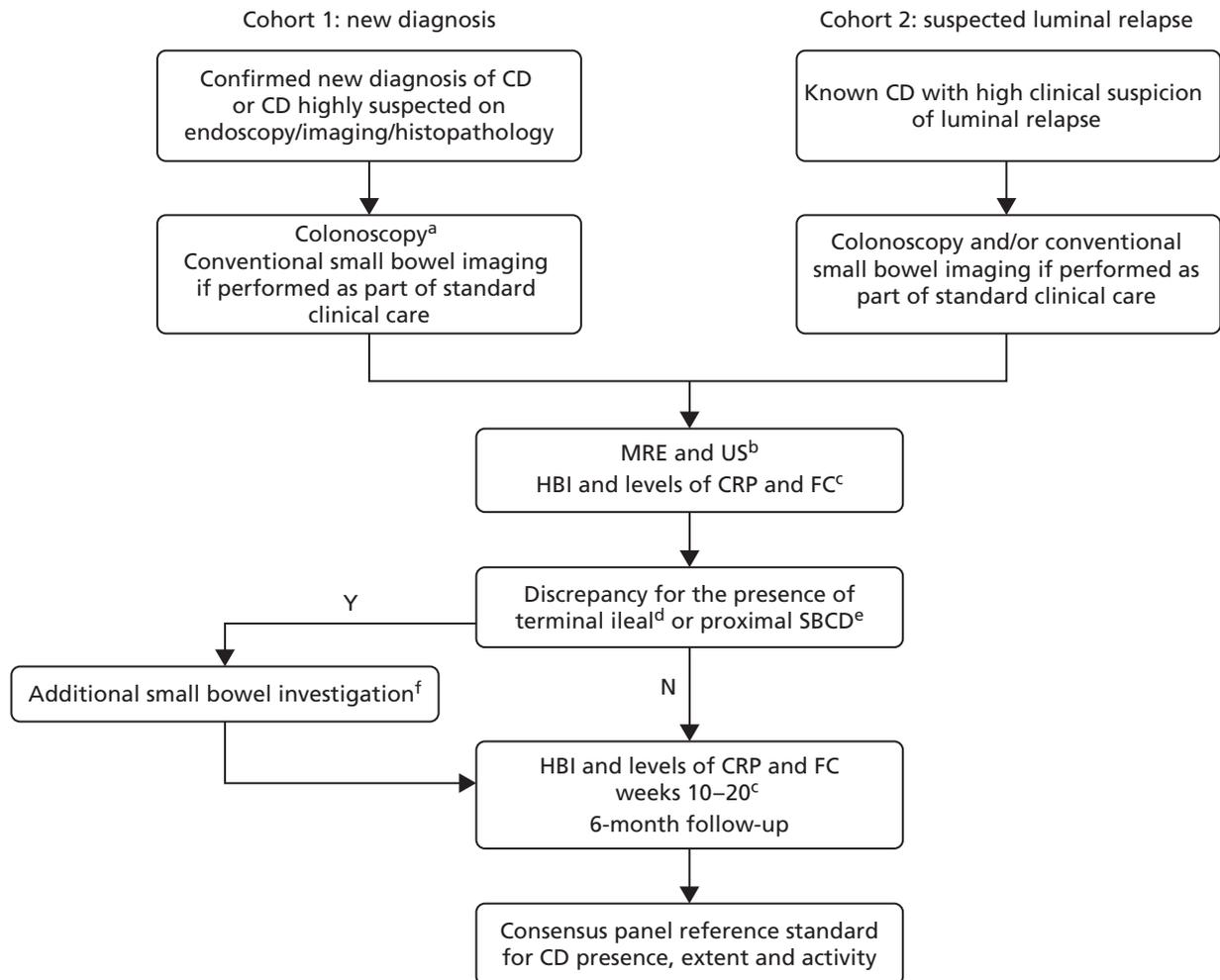


FIGURE 1 Participant flow through the main study. CRP, C-reactive protein; FC, faecal calprotectin; HBI, Harvey–Bradshaw Index. a, Already performed or pending at recruitment; b, interpreted by two independent blinded practitioners; c, required specific participant consent; d, in the absence of endoscopic visualisation; e, in the absence of existing small bowel imaging investigations or capsule endoscopy; f, for example capsule endoscopy, barium small-bowel follow-through and/or CTE, at discretion of recruitment site.

Recruitment sites

Participants were recruited from eight NHS hospitals in England and Scotland. A mixture of teaching and general hospitals was included. All sites were required to have an established IBD service seeing > 150 participants annually and lead radiologists had to be affiliated with the British Society of Gastrointestinal and Abdominal Radiology (BSGAR) to ensure expertise in CD imaging, including MRE and US (see *Radiologist/sonographer competence and training*). To meet the trial imaging blinding protocol requirements, each site had to nominate at least two radiologists/sonographers to participate in the study. Each site had a named research nurse/practitioner or researcher responsible for recruitment.

Participants

Participants were recruited to two defined cohorts:

1. those newly diagnosed with CD
2. those previously diagnosed with CD with a high suspicion of luminal relapse requiring radiological investigation.

Inclusion criteria

New diagnosis cohort

Participants were eligible for the new diagnosis cohort if either they had undergone colonoscopy or colonoscopy was planned, and they:

- had been recently diagnosed with CD (within 3 months of recruitment) based on endoscopic, histological and radiological findings or were highly suspected of CD based on characteristic endoscopic, imaging or histological features but pending final diagnosis
- were aged ≥ 16 years
- were able to give fully informed consent.

Suspected relapse cohort

Participants were eligible for the suspected relapse cohort if they:

- had a known diagnosis of CD with a high clinical suspicion of luminal relapse indicating radiological investigation
- were aged ≥ 16 years
- were able to give fully informed consent.

A high clinical suspicion of luminal relapse was defined as objective markers of inflammatory activity [a C-reactive protein (CRP) level of > 8 mg/l or a faecal calprotectin (FC) level of > 100 $\mu\text{g/g}$] or symptoms suggestive of luminal stenosis (including obstructive symptoms such as colicky abdominal pain and vomiting) or abnormal endoscopy suggesting relapse. A CRP level of > 8 mg/l (rather than > 5 mg/l) was selected to increase specificity for participants with true relapse.

Exclusion criteria

Participants were not eligible for recruitment to either study cohort if they:

- had any psychiatric or other disorder likely to affect informed consent
- had evidence of severe or uncontrolled systemic disease that at the principal investigator's discretion rendered the individual unsuitable for participation
- were pregnant
- had contraindications to MRE (e.g. allergy to all suitable contrast agents, cardiac pacemaker, severe claustrophobia, an inability to lie flat).

Participants were not eligible for recruitment to the new diagnosis cohort if they:

- had a final diagnosis other than CD
- underwent surgical resection prior to a pending colonoscopy.

Test methods

Recruited participants to both study cohorts underwent MRE and US.

Magnetic resonance enterography

Technical parameters

Magnetic resonance enterography was performed by the usual clinical radiographer team at each recruitment site. To improve generalisability of results, the MRI platform (i.e. manufacturer and tesla (T) strength) utilised

was allocated by the lead site radiologist according to availability and their usual practice. Exact imaging parameters varied according to MRI platform but a minimum data set of sequences was acquired, including T2-weighted images with and without fat saturation, steady-state free precession gradient echo images, diffusion-weighted images and T1-weighted images after intravenous gadolinium injection, as defined by the study investigators (see *Appendix 1, Table 40*). The choice of oral contrast prior to MRE was also at the discretion of the recruitment site and in accordance with their usual practice (see *Appendix 2, Table 41*). Where possible, radiographers recorded the volume of oral contrast ingested by participants and the time taken for ingestion prior to MRE.

In some participants, MRE had been performed as part of usual clinical care prior to recruitment. If it had been performed within the preceding 4 weeks and according to the minimum data set of sequences, it was deemed sufficient for the purposes of the study and not repeated (see *Trial blinding*).

Ultrasonography

Technical parameters

Ultrasonography was performed by local site radiologists or sonographers (see *Radiologist/sonographer competence and training*) using standard US platforms at recruitment sites. Participants were nil by mouth for 4 hours and no oral contrast was administered prior to US, although ingestion of two cups of water by participants just prior to the scan was permissible to improve visualisation of the duodenum. A sample of the participants were recruited to a substudy of SICUS and underwent an additional US after ingesting oral contrast medium (see *Chapter 4*). The colon and small bowel were systematically interrogated using both curvilinear and high-resolution linear probes (minimum 5 MHz frequency). Colour Doppler was routinely applied (with typical flow settings 6–9 m/second) but intravenous US contrast agents were not administered. US performed as part of usual clinical care prior to recruitment was repeated for the purposes of the study if appropriate blinding of the operator could not be assured (see *Trial blinding*).

Quality assurance

All sites sent in MRE and US images to the lead site [University College Hospital (UCLH)] for secure upload and storage. Compliance with the minimal protocol data set was confirmed but formal quality assurance was not undertaken given that all sites were experienced in both MRE and US techniques.

Clinical investigations

Colonoscopy

Participants recruited to the new diagnosis cohort either had undergone colonoscopy or had colonoscopy planned as part of usual clinical care at recruitment. Participants recruited to the suspected relapse cohort underwent colonoscopy only if deemed necessary as part of their routine clinical care, irrespective of study recruitment. Colonoscopy was performed and reported by local gastroenterologists as per usual clinical practice.

Additional small bowel imaging

All participants recruited to the trial underwent MRE and US. However, in addition, some participants underwent additional small bowel imaging, for example CTE or BaFT, either prior to or after recruitment as part of their usual clinical care. All additional small bowel imaging tests were performed and reported by local radiologists as per their standard clinical practice.

Radiologist/sonographer competence and training

Competence and training requirements for the study were defined a priori such that individuals interpreting trial imaging were representative of those individuals who report small bowel imaging in the NHS. We specifically avoided using a small number of highly experienced subspecialty practitioners who would not be representative of the NHS radiological workforce. Across all sites, 28 practitioners interpreted the MRE and US studies (27 radiologists and one sonographer). Practitioners were selected by the sites' lead radiologists

and all met the training and experience criteria detailed below. Eight radiologists interpreted MRE only, three performed and interpreted US only and 16 performed and interpreted US and MRE. All site lead radiologists were affiliated to the BSGAR to ensure local dissemination of best practice. All reporting radiologists were post Fellowship of the Royal College of Radiologists (FRCR) and either were at consultant level and/or had ≥ 1 year of subspecialty gastrointestinal experience. All had a declared subspecialty interest in gastrointestinal radiology with previous experience of MRE and US. Specifically, radiologists interpreting MRE had a median of 10 years of experience [interquartile range (IQR) 6–11 years] and practitioners interpreting US had a median of 8 years of experience (IQR 4–11 years). The participating sonographer had undergone formal training, was already performing enteric US in clinical practice with 20 years of experience and had been deemed competent by their lead radiologist. The median number of examinations performed per month at each recruitment site (including those patients not recruited to the METRIC trial) during the conduct of the trial was 30 (range 20–50, IQR 20–45) for MRE and 25 (range 4–80, IQR 12–40) for US. The monthly range of MRE examinations across sites was 20 (Queen Alexandra Hospital, Portsmouth, and St George's Hospital, London) to 50 (UCLH, London, and John Radcliffe Hospital, Oxford). The range of monthly US examinations across sites was three (John Radcliffe Hospital, Oxford) to 80 (Queen Alexandra Hospital, Portsmouth). A 2-day hands-on US training workshop was run in Portsmouth before trial commencement to standardise US technique and agree the description of enteric findings (see *Reporting of trial imaging*).

Study imaging interpretation and reporting

Blinding of interpreting radiologists/sonographers

Unbiased estimates of imaging test diagnostic accuracy can be achieved only if those individuals interpreting the tests are unaware of the findings of contemporaneous imaging and endoscopy. For example, a practitioner aware of endoscopically confirmed terminal ileal disease could not give an unbiased evaluation of subsequent US or MRE in the same patient. MRE and US for individual recruited participants were therefore interpreted by different practitioners blinded to all clinical information other than the cohort (suspected relapse or new diagnosis) and surgical history. Practitioners were blinded to all other current/past imaging investigations and endoscopies. Surgical history was disclosed so as not to disadvantage US (sites of surgical resection are usually indicated at the time of a clinical request for imaging). Practitioners performing US were explicitly instructed to not converse with participants about their medical history, and where possible the examination was witnessed by a research nurse. If blinding of the reporting practitioner could not be assured, for example in the case of MRE or US performed prior to study recruitment, MRE images were reanalysed by a blinded local radiologist, or a central radiologist at UCLH (if a suitable local radiologist was not available), and the US was repeated by an appropriately blinded individual (as US interpretation occurs in real time and cannot be reproduced by review of static images).

Reporting of trial imaging

The imaging appearances of CD on MRE and US are well described³⁵ and utilised for the purposes of the study. Guidance on the criteria for disease activity was provided to practitioners based on the literature at the time of study design.^{12,22} Specifically, signs of active disease on MRE included wall thickening, increased mural T2 signal, increased mesenteric T2 signal, increased enhancement (mucosal or layered), ulceration and abscess; signs of active disease on US included wall thickening, focal hyperechoic mesentery (with or without fat wrap), isolated thickened submucosal layer, poorly defined antimesenteric border, increased Doppler vascular pattern, ulceration and abscess.

Practitioners completed a case report form (CRF) recording their findings for MRE and US. Items recorded on the CRFs included the imaging platform used to acquire the images and confirmation of radiologist/sonographer blinding to other investigations. The quality of visualisation for 10 bowel segments (duodenum, jejunum, ileum, terminal ileum, caecum, ascending colon, transverse colon, descending colon, sigmoid colon and rectum) was graded as good, moderate or poor. The terminal ileum was defined as the last 10 cm of the small bowel and the jejunum was defined as the proximal bowel lying largely to the left of a diagonal line drawn from the right lower quadrant to the left lower quadrant, demonstrating a typical feathery fold pattern. Colonic segments were defined as previously described.³⁶ The presence of any small bowel and/or

colonic CD was then recorded using six confidence levels grouped into normal (levels 1 and 2), equivocal (levels 3 and 4) and abnormal (levels 5 and 6). If disease presence was recorded as equivocal or abnormal (i.e. confidence level 3 or higher), the level of disease activity was also scored from 1 (disease definitely not active) to 6 (disease definitely active). The same grading system for disease presence and activity was then applied to each of the 10 individual bowel segments. Terminal ileal disease extending contiguously for greater than 10 cm was considered terminal ileal disease only and was not considered to be affecting both the terminal ileum segment and ileal segments separately. If a segment contained more than one site of disease (defined as > 3 cm of normal-appearing bowel between disease sites), this was recorded as a separate disease site in that segment. Similarly, if there was > 3 cm of normal-appearing bowel between terminal ileal disease and ileal disease, this was considered to be affecting both segments. In participants with prior terminal ileal resection, the neo-terminal ileum was considered to be the terminal ileum for the purposes of the trial. The presence of extraenteric complications including abscesses and fistulae were also recorded. For MRE alone, the reporting radiologist recorded if their diagnostic confidence had been influenced by diffusion-weighted images and/or contrast-enhanced images and if their diagnosis for disease presence or activity had been changed after review of these sequences (see *Chapter 6*).

Following completion of the CRF, radiologists/sonographers produced an unblinded standard report for the MRE and US as per usual clinical practice.

As part of an interobserver variation study, a proportion of participants underwent a second US by a separate radiologist/sonographer who completed an identical CRF (see *Chapter 5*).

Generation of additional small bowel imaging for discrepant magnetic resonance enterography and ultrasonography

There is no single reference standard for the presence of CD in the proximal small bowel upstream of the terminal ileum (which is usually assessed during ileocolonoscopy). In participants who had not undergone an additional small bowel imaging test [e.g. CTE, BaFT or capsule endoscopy (CapE)] as part of their usual clinical care, the only available assessment of the proximal small bowel was MRE and US, risking incorporation bias during the derivation of the reference standard. In such participants, radiologist/sonographer interpretations of MRE and US were therefore reviewed at the time of reporting to ascertain if they were discrepant for the presence of SBCD. Discrepancy was defined as (1) disease reported in the terminal ileum on only MRE or US in the absence of endoscopic visualisation of the terminal ileum and/or (2) disease reported in the small bowel upstream of the terminal ileum on only MRE or US, including additional disease sites in those patients with multifocal involvement. To avoid unnecessary additional tests, in participants with a reported single site of SBCD that differed only in segmental localisation between MRE and US (e.g. ileum vs. jejunum), the local research team reviewed the imaging and opined if the tests were probably concordant (i.e. same abnormality detected) or truly discrepant and recorded this accordingly.

Participants with a true SBCD discrepancy on MRE and US according to the above definitions were invited to undergo an additional small bowel investigation within 8 weeks of the initial study imaging. The choice of investigation was at the discretion of the recruitment site and could include BaFT, CTE or CapE for example. Emphasis was placed on performing a new imaging modality but a repeat MRE or unblinded US was permitted if an alternative imaging modality was deemed inappropriate by the clinical care team or participant. The additional small bowel imaging test was interpreted by a site radiologist fully unblinded to all other investigations and formed part of the later consensus panel review process for the study reference standard.

Recruitment

Suitable participants were identified by members of the local research team, who established whether or not the individual met the study entry criteria, from outpatient clinics, multidisciplinary team (MDT) meetings, inpatient wards and lists of requests for small bowel imaging and endoscopy. A screening log

recorded the details of all individuals approached to take part in the study and reasons for non-participation if applicable. All individuals were handed or posted a participant information sheet detailing the study and contact details of the study team should they have any questions. The study purpose and requirements were also explained to participants face to face by an appropriately trained member of the research team. All participants gave written consent prior to participation in the main METRIC study. Participants gave additional written consent if they agreed to participate in the SICUS substudy (see *Chapter 4*), participate in the US interobserver substudy (see *Chapter 5*) or complete patient experience questionnaires (see *Chapter 7*). Participants retained a copy of their consent form and participant information sheet and were informed that they could withdraw from the study at any time.

Data collection and participant follow-up

Data collation was co-ordinated by the UCL Comprehensive Clinical Trials Unit (CCTU). Demographic and baseline clinical information was collected at recruitment using a specially designed CRF that recorded sex, age, family history of IBD, smoking status, current CD-related medication, previous history of bowel surgery, current symptoms and, for participants recruited to the suspected relapse cohort, time since CD diagnosis and current Montreal classification.³⁷ Participants were invited to supply a stool sample for measurement of FC, to supply a blood sample for CRP measurement and to complete a Harvey–Bradshaw Index (HBI) disease activity test³⁸ at recruitment and again at 10–20 weeks, if these had not already been carried out as part of usual clinical care. The findings of endoscopies and biopsies (if performed) were recorded on CRFs, as were the findings of all contemporaneous imaging investigations (including those findings of additional small bowel imaging triggered by discrepant MRE and US reports for the presence of SBCD). Complications related to MRE or US were recorded on a specific CRF. The clinical course of participants was followed for a period of 6 months after recruitment to inform the consensus panel review process and collate data for the health economic analysis (see *Chapter 10*). During this time, details of any CD-related surgical interventions (including histology) were recorded on CRFs, along with additional imaging/endoscopic investigations, outpatient visits, hospital day visits, inpatient stays and details of CD medication. All CRFs were collated by local site research nurses/practitioners and sent to the CCTU by post or fax. Forms were entered onto a bespoke study database and any missing fields or apparent data inaccuracies queried with the centre to optimise data collection.

Reference standard

The METRIC trial used the construct reference standard paradigm (panel diagnosis) incorporating the concept of clinical test validation (i.e. whether or not the results of an index test are meaningful in practice). Specifically, by following the participants' clinical course for 6 months after recruitment it was possible to assess the impact of clinical decision-making on participant outcomes based on the findings of MRE and US. For example, if biological therapy was commenced based on a MRE or US finding of active SBCD, the success or otherwise of this therapy could be ascertained using the 6-month participant outcome data, contributing to the panel decision as to the validity or otherwise of the imaging findings. Each recruitment site convened a series of consensus panels to derive the reference standard for disease presence, extent and activity at the time of consent for each participant recruited at their site. Typically, each consensus meeting considered around 10 participants in one 2- to 3-hour session and consisted of at least one gastroenterologist from the recruitment site and at least two radiologists (one internal to the site and one external, from another recruitment site), along with a member of the local research support team to aid the running of the meeting and CRF completion. A member of the Trial Management Group (TMG) attended each consensus meeting to ensure uniformity when defining disease presence, activity and extent and a histopathologist was available to the panel if required. The panel considered all available clinical information over the 6-month follow-up period, including the images and results of all small bowel investigations (including MRE, US and all generated third small bowel imaging tests), endoscopy (reports and images), surgical findings (if applicable), histopathology (surgical resection and biopsies), HBI, CRP level,

FC level (and changes thereof in response to therapy), follow-up imaging and clinical course. Panels had access to all completed follow-up CRFs as well as participant clinical results, records and letters via the participant notes and/or electronic participant record. Panels also had access to the hospital picture archiving and communications system (PACS) so they could review all small bowel imaging.

For each recruited participant, the panel completed a specific consensus reference standard CRF. The imaging, endoscopic and clinical data considered by the panel were recorded, along with their interpretation of the available data for presence and activity of SBCD and colonic CD. The panel specifically recorded the findings of endoscopic and biopsy data, given its robustness as a standard of reference for the presence of colonic and terminal ileal disease. Based on all information, the panel recorded if, in their opinion, there was any small bowel or colonic CD present, and if present whether or not the disease was active. Disease could only be categorised as active if there was at least one objective marker of activity [(1) ulceration as seen at endoscopy and/or (2) measured CRP level of > 8 mg/l and/or (3) measured FC level of > 250 µg/g and/or (4) histopathological evidence of acute inflammation based on biopsy or surgery within 2 months of study imaging]. Thereafter, the presence and activity of CD in each of 10 bowel segments (duodenum, jejunum, ileum, terminal ileum, caecum, ascending colon, transverse colon, descending colon, sigmoid colon and rectum) were recorded by the panel. The original blinded study interpretations of MRE and US were reviewed by the panel and any false-negative observations for disease presence (against the consensus reference standard) were classified by the panel as radiologist perceptual error if the abnormality was visible on reviewing the imaging in retrospect, or as a technical failure of the imaging modality if it was not. The panel also documented the presence or absence of extraluminal complications including abscesses and fistulae. If consensus panels could not reach agreement on any aspect of the reference standard CRF, they were able to refer the review to another recruitment site's consensus panel for their consideration.

Outcomes

A summary of the primary and secondary outcomes is shown in *Appendix 3, Table 42*. The primary outcome for the main study was the difference in per-participant sensitivity between MRE and US for the correct identification and localisation of SBCD, irrespective of activity, against the consensus reference standard. For location matching, the small bowel was divided into duodenum, jejunum, ileum and terminal ileum (see *Reporting of trial imaging*). Given the potential clinical impact of underdiagnosed SBCD (e.g. primary surgery for isolated terminal ileal disease vs. medical management for more diffuse disease), to be a true positive for the primary outcome the index test had to correctly locate both the presence and segmental location of the disease. A summary of the criteria for agreement with the reference standard for disease extent is given in *Appendix 4, Table 43*. Disease reported as equivocal was treated as positive for disease presence given the potential clinical implications of an equivocal result on patient management and the need for further investigations. A sensitivity analysis treating equivocal results as negative for disease presence was also performed.

Secondary outcome measures were the difference in per-participant specificity of MRE and US for the correct identification and localisation of SBCD irrespective of activity and the per-participant sensitivity and specificity for identification of colonic CD presence and extent. Additional secondary outcome measures were per-participant sensitivity and specificity for the presence of active SBCD and colonic CD against the consensus reference standard. The secondary analyses were repeated for the terminal ileum and colonic segments in participants with an available colonoscopic reference standard, given its robustness in CD identification and activity assessment.

The primary and secondary analyses were performed for both cohorts combined and then separately for the new diagnosis and suspected relapse cohorts.

Additional secondary outcomes pertaining to the comparative impact of MRE and US on clinician diagnostic confidence and management, the lifetime incremental cost and cost-effectiveness of assessment

using MRE or US, diagnostic accuracy of SICUS compared with conventional US, comparative participant experience of MRE and US, diagnostic impact of novel MRE sequences and interobserver variation in the evaluation of MRE and US data sets are described in the relevant chapters (see *Chapters 4–10*).

Sample size

Primary outcome

The sample size calculation was based on the primary outcome stipulated by the Health Technology Assessment (HTA) commissioning brief: diagnostic accuracy for SBCD extent. There are two aspects of correctly assigning disease extent: correctly detecting the presence of disease and correctly assigning its segmental location. Study power was thus based on a two-faceted compound accuracy measure (disease presence and disease location). Based on the available literature at the time of study design (see *Report Supplementary Material 1*), it was assumed that MRE had 93% sensitivity for disease presence and 90% sensitivity for disease location, resulting in an overall compound accuracy for disease extent of 83% (93% × 90%). The corresponding assumed values for US were 88%, 83% and 73%, resulting in a 10% difference in overall compound accuracy for disease extent between MRE and US (83% vs. 73%). Assuming moderate correlation between the two imaging tests (68% positive test result on both MRE and US), a total of 210 participants with SBCD were required to detect a 10% superiority of MRE over US at 90% power (type II error).³⁹ Assuming a SBCD prevalence of 70%, a cohort of 301 participants was required, and assuming 10% loss to follow-up/non-CD final diagnosis, a total cohort of 334 participants (167 new diagnosis participants and 167 suspected relapse participants) was required. A sample size of 156 participants was sufficient to detect a 13% difference in sensitivity for the primary outcome between MRE and US at 80% power.

Secondary outcomes

Study power for correct per-participant identification of disease activity assumed sensitivities of 88% and 78% for MRE and US, respectively (see *Report Supplementary Material 1*). A total of 204 participants with active disease gave 80% power to detect a 10% difference in correct per-participant disease activity classification. Assuming a SBCD prevalence of 70% and 10% loss to follow-up, a cohort of 324 participants was required. Sensitivity for correct per-participant identification of active disease in the terminal ileum against a colonoscopic reference standard assumed that 200 participants would have colonoscopic data available (all the new diagnosis cohort and one-third of the suspected relapse cohort). Assuming sensitivities for segmental disease activity of 75% and 60% for MRE and US, respectively, and 70% prevalence of SBCD, 195 participants was sufficient to detect a 15% difference in sensitivity.

Analysis

Disease reported as equivocal was treated as positive in the analysis. The primary outcome was calculated per participant. Secondary outcomes for bowel segments were based on all segments, excluding those segments resected at baseline (for terminal ileal resections, the neo-terminal ileum was considered to be the terminal ileum).

Direct comparison of sensitivity and specificity differences between MRE and US were calculated using bivariate multilevel participant-specific (conditional) random-effects models, from paired data using *meqrlogit* in *Stata*® 14.2 (StataCorp LP, College Station, TX, USA). When models did not converge owing to small numbers of participants, McNemar's comparison of paired proportions was used to obtain univariable estimates, and exact 95% confidence intervals (CIs) were calculated. Analysis by colonic segment used a population-averaged random-effects model (using *logit* including robust standard errors). Statistical significance was based on 95% CIs. There were no missing data for per-participant diagnosis of disease presence or disease extent.

Chapter 3 Results

This chapter contains material that is reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

Participants

Recruitment began in December 2013 and was completed in September 2016. Overall, 518 participants were assessed for eligibility, of whom 183 were excluded, predominantly because they did not meet the inclusion criteria or declined participation (*Table 2*).

Of the 335 participants who entered the trial, 51 were subsequently excluded (*Table 3*). The most frequent reason was an ultimate diagnosis other than CD (31 participants).

The final cohort was 284 participants (133 and 151 in new diagnosis and suspected relapse cohorts, respectively) (*Figure 2*). *Appendix 5, Table 44*, shows the recruitment and withdrawal numbers for each of the eight recruitment sites. *Figure 2* is the patient flow diagram.

Baseline participant characteristics

The demographic data of the final study cohort are shown in *Table 4*. There were marginally more men in the new diagnosis cohort and more women in the suspected relapse cohort. Overall, 154 (54%) participants were female. In the suspected relapse cohort, 101 (67%) participants had had CD for ≥ 6 years and 53 (35%) and 14 (9%) had a stricturing or penetrating disease phenotype, respectively.

TABLE 2 Reasons for exclusion of screened patients

Reason for exclusion	n (%)
Non-CD diagnosis	22 (12)
Not able to give informed consent	7 (4)
Declined participation	58 (32)
No response to invitation to participate	28 (15)
Contraindication to MRI	8 (4)
Newly diagnosed > 3 months previously	2 (1)
Unable to complete MRE and/or US in timely fashion	20 (11)
Aged < 16 years	2 (1)
Previous recruitment or declined approach	5 (3)
CRP level not raised (suspected relapse cohort)	13 (7)
Moved/lived far way	4 (2)
Proceeded straight to surgery prior to colonoscopy (new diagnosis cohort)	4 (2)
Unknown	10 (6)
Total	183

TABLE 3 Reasons for participant withdrawal

Reason for withdrawal	Cohort, n (%)		Total, n (%)
	New diagnosis	Suspected relapse	
Participant withdrew consent	0 (0)	3 (18)	3 (6)
Final diagnosis other than CD	25 (73)	6 (35)	31 (60)
Participant did not undergo MRE	0 (0)	5 (29)	5 (10)
Participant did not undergo US	2 (6)	1 (6)	3 (6)
Participant did not undergo MRE or US	2 (6)	0 (0)	2 (4)
Lost to follow-up	2 (6)	0 (0)	2 (4)
Underwent surgery without colonoscopy	2 (6)	0 (0)	2 (4)
No longer wished to participate at follow-up	1 (3)	2 (12)	3 (6)
Total	34	17	51

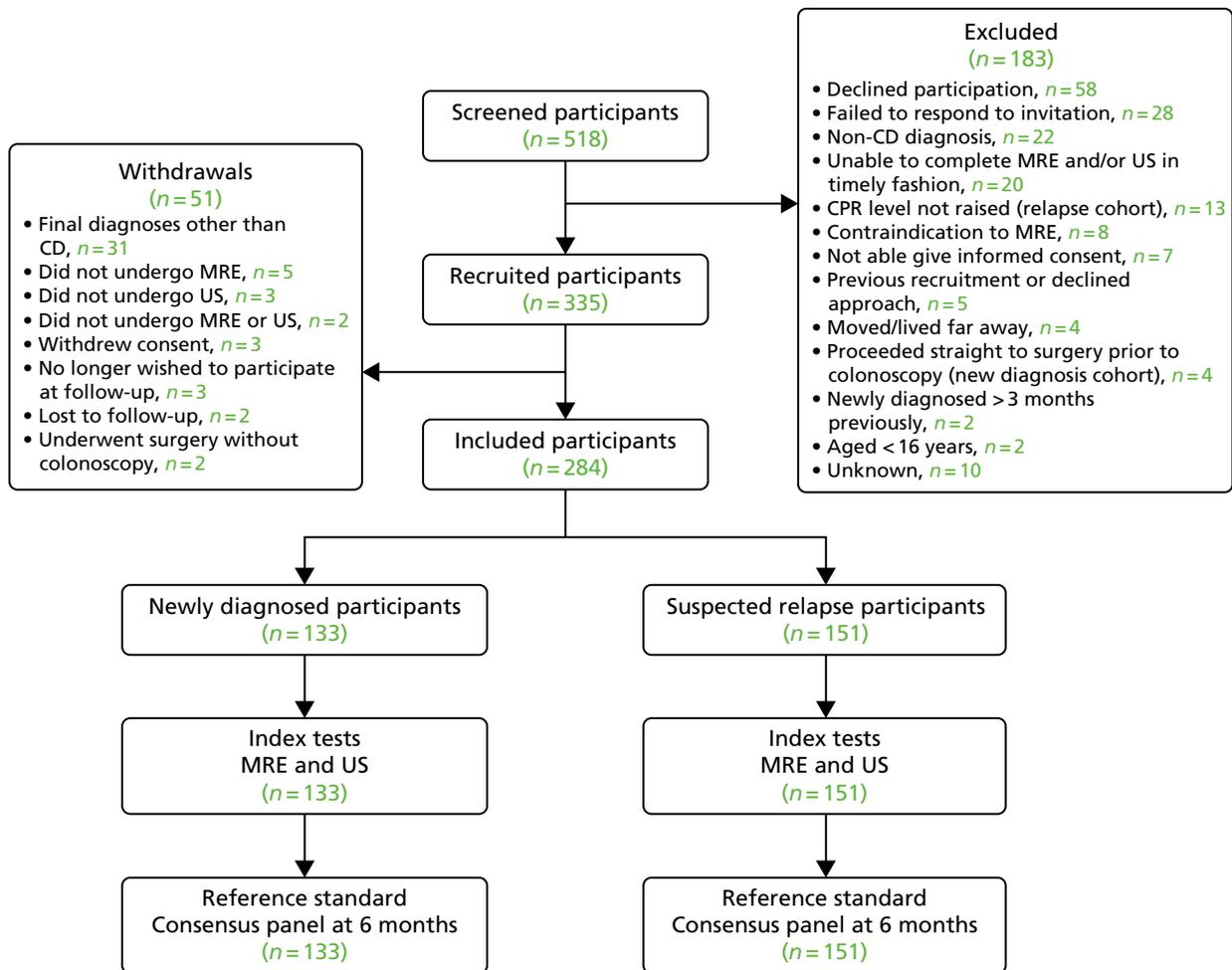
**FIGURE 2** Participant flow diagram. Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 4 Demographics of final study cohort

Variable	Cohort, <i>n</i> (%)	
	New diagnosis (<i>N</i> = 133)	Suspected relapse (<i>N</i> = 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	N/A	5 (3)
1–5 years	N/A	45 (30)
6–10 years	N/A	39 (26)
> 10 years	N/A	62 (41)
Disease location (Montreal classification) ^a		
L1	N/A	56 (37)
L2	N/A	17 (11)
L3	N/A	74 (49)
L4	N/A	4 (3)
Disease behaviour (Montreal classification) ^a		
B1	N/A	80 (53)
B1p	N/A	4 (3)
B2	N/A	52 (34)
B2p	N/A	1 (1)
B3	N/A	12 (8)
B3p	N/A	2 (1)
Medication ^b		
None	62 (47)	32 (21)
5-ASA	21 (16)	26 (17)
Steroids	48 (36)	28 (19)
Immunomodulators	16 (12)	75 (50)
Anti-TNF therapy	5 (4)	42 (28)
Previous enteric resection	1 (1) ^c	72 (48)

5-ASA, 5-aminosalicylic acid; N/A, not applicable; TNF, tumour necrosis factor.

a Montreal classification not collected for new diagnosis participants.

b Participants could take more than one medication type.

c Surgical resection for inflammatory mass 1 year prior to CD diagnosis.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Participants had a range of presenting symptoms, notably abdominal pain and diarrhoea (see *Appendix 6, Table 45*). In general, symptoms were similar between the two cohorts, although the proportion of participants reporting bloody diarrhoea was higher in those participants newly diagnosed, and a greater proportion of those participants with relapse reported obstructive symptoms. There were no reported major adverse events following MRE or US.

Consensus reference standard

The available small bowel imaging tests (including third tests generated by a discrepancy for SBCD presence or location between MRE and US), CRP and FC levels, HBI and surgical resection specimens available to the consensus panels are shown in *Table 5*. A total of 10 (8%) new diagnosis participants did not undergo full colonoscopy despite this being planned at recruitment. Of these, five underwent flexible sigmoidoscopy instead, and the remaining either declined or had a change in their clinical investigational plan. Colonoscopy data were available for 66 (44%) of the suspected relapse cohort participants. There was a range of small bowel imaging tests available over and above MRE and US, including CapE, BaFT and CTE.

TABLE 5 Range of investigations available to the consensus reference panel

Variable	Cohort, n (%)	
	New diagnosis (N = 133)	Suspected relapse (N = 151)
MRE	133 (100)	151 (100)
US	133 (100)	151 (100)
Colonoscopy	123 (92)	66 (44) ^a
Gastroscopy	11 (8)	6 (4)
Sigmoidoscopy	5 (4)	12 (8)
CapE	10 (8)	8 (5)
CTE	4 (3)	9 (6)
CT abdomen and/or pelvis	21 (16)	13 (9)
MR enteroclysis	4 (3)	6 (4)
MRI abdomen and/or pelvis	5 (4)	8 (5)
BaFT	8 (6)	19 (13)
Barium enteroclysis	3 (2)	7 (5)
Hydrosonography	28 (21)	37 (25)
White cell scan	0 (0)	0 (0)
CRP level		
Baseline	127 (95)	145 (96)
10–20 weeks	108 (81)	120 (79)
HBI		
Baseline	124 (93)	142 (94)
10–20 weeks	71 (53)	77 (51)
FC level		
Baseline	87 (65)	89 (59)
10–20 weeks	53 (40)	65 (43)

TABLE 5 Range of investigations available to the consensus reference panel (*continued*)

Variable	Cohort, n (%)	
	New diagnosis (N = 133)	Suspected relapse (N = 151)
Surgical resection		
Before recruitment	1 (1)	72 (48)
During trial follow-up	1 (1)	2 (1)
Other	8 (6)	20 (13)

a Three participants with follow-up colonoscopies; data not available for these participants. Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Based on the consensus reference standard, 233 (82%) participants had SBCD (meeting the requirements of the sample size calculation), which was active in 209 (89%) (*Table 6*). A total of 129 (45%) participants had colonic CD, which was active in 126 (98%). Participants often fulfilled more than one criterion for active disease (raised CRP/FC levels, ulceration at endoscopy, histopathological evidence of inflammation) (see *Table 6*). The prevalence of SBCD was similar between the new diagnosis and suspected relapse cohorts, although colonic CD tended to be more prevalent in the former. The prevalence of activity was high when disease was present and was also similar between the two cohorts.

The presence and activity of SBCD and colonic CD according to individual bowel segments is shown in *Appendix 7, Table 46*. The prevalence of individual small bowel segmental disease was similar between the new diagnosis and suspected relapse cohorts, although segmental colonic CD was more prevalent in the former.

A total of 21 participants had enteric fistulae and seven participants had an intra-abdominal abscess. Specifically, three (2%) and four (3%) participants had an abscess in the new diagnosis cohort and in the suspected relapse cohort, respectively. A total of 10 (8%) and 11 (7%) participants had a fistula in the new diagnosis cohort and in the suspected relapse cohort, respectively.

There were 61 bowel segments considered to have a stenosis causing obstruction: 18 in the new diagnosis cohort and 43 in the suspected relapse cohort.

The practitioners' opinion on the quality of segmental visualisation for both MRE and US according to the participant cohort is shown in *Appendix 8, Tables 47 and 48*. In general, visualisation of ileal and terminal ileal segments was rated as at least moderate in > 90% of participants on both MRE and US, with no major difference between cohorts. Visualisation of the duodenum was rated as poor in 15 out of 133 (11%) new diagnosis participants and 18 out of 151 (12%) suspected relapse participants on MRE, and in 23 out of 133 (17%) new diagnosis participants and 27 out of 151 (18%) suspected relapse participants for US. Jejunal visualisation tended to be better on US, for example rated as poor in 23 out of 151 (15%) suspected relapse participants on MRE compared with 8 out of 151 (5%) on US. Colonic visualisation was inferior to that of the small bowel on both MRE and US but better on US than MRE for five out of six segments. For example, visualisation of the descending colon was rated as good in 58 out of 133 (44%) new diagnosis participants using MRE compared with 92 out of 133 (69%) new diagnosis participants on US. The only exception was the rectum, where visualisation was rated as poor in 74 out of 133 (56%) new diagnosis participants on US compared with 40 out of 133 (30%) new diagnosis participants on MRE.

TABLE 6 Per-participant disease presence and activity according to the consensus reference standard

Variable	Cohort	
	New diagnosis (N = 133)	Suspected relapse (N = 151)
Disease presence		
SBCD present, n (%)	111 (83)	122 (81)
Colonic CD present, n (%)	77 (58)	52 (34)
Both SBCD and colonic CD present, n (%)	55 (41)	37 (25)
Total number of participants with disease present, n (%)	133 (100)	137 (91)
Average number of involved small bowel segments [median (IQR), maximum]	1 (1–1), 4	1 (1–1), 3
Average number of involved colonic segments [median (IQR), maximum]	1 (0–3), 6	0 (0–1), 6
Disease activity, n (%)		
Small bowel active disease	104 (94)	105 (86)
Colonic active disease	76 (99)	50 (96)
Total number of participants with active disease ^a	130 (98)	121 (88)
Criteria for activity, n (%)		
Ulceration at endoscopy	71 (55)	26 (21)
CRP level of > 8 mg/l	47 (36)	57 (47)
FC level of > 250 µg/g	41 (32)	43 (36)
Histological evidence of activity	100 (77)	36 (30)

^a Participants could meet more than one criterion for active disease.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Test results and outcomes

Identification and localisation of small bowel and colonic Crohn's disease against the consensus reference standard

In total, 53 participants (24 new diagnosis and 29 suspected relapse participants) were discrepant for SBCD presence or location between MRE and US, of whom 48 had an additional small bowel imaging test available to the consensus panel. Of these, 17 (71%) new diagnosis participants and 17 (59%) suspected relapse participants were discrepant for the presence of terminal ileal disease. The full range of imaging and endoscopic data available to the consensus panels for these participants is shown in *Appendix 9, Table 49*.

Appendix 10, Table 50, provides the raw data for the primary outcome and main secondary outcome.

Primary outcome

For SBCD extent, MRE sensitivity (i.e. presence and correct segmental location) was 80% (95% CI 72% to 86%), compared with 70% (95% CI 62% to 78%) for US, a difference of 10% (95% CI 1% to 18%), which was statistically significant ($p = 0.027$) (*Table 7* and *Figure 3*).

TABLE 7 Per-participant sensitivity and specificity for disease presence and extent against the consensus reference standard (both cohorts combined)

Variable	Disease-positive participants (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-negative participants (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
SBCD								
Extent ^b	233	80 (72 to 86)	70 (62 to 78)	10 (1 to 18; 0.027)	51	95 (85 to 98)	81 (64 to 91)	14 (1 to 27; 0.039)
Presence	233	97 (91 to 99)	92 (84 to 96)	5 (1 to 9; 0.025)	51	96 (86 to 99)	84 (65 to 94)	12 (0 to 25; 0.054)
Colonic CD								
Extent ^b	129	22 (14 to 32)	17 (10 to 27)	5 (-5 to 15; 0.332)	155	93 (87 to 97)	93 (87 to 97)	0 (-5 to 5; 1.000)
Presence	129	64 (50 to 75)	73 (59 to 83)	-9 (-23 to 5; 0.202)	155	96 (90 to 98)	96 (90 to 98)	0 (-3 to 3; 1.000)
SBCD and colonic CD								
Extent ^b	270	45 (36 to 54)	29 (21 to 38)	16 (6 to 25; 0.002)	14	80 (42 to 96)	61 (23 to 89)	19 (-20 to 59; 0.337)
Presence ^c	270	78 (70 to 85)	71 (62 to 79)	7 (-2 to 15; 0.117)	14	80 (42 to 96)	61 (23 to 89)	19 (-20 to 59; 0.335)

a By consensus reference standard.

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

c Agreement with reference standard for disease presence in participants with disease in the small bowel, colon or both.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

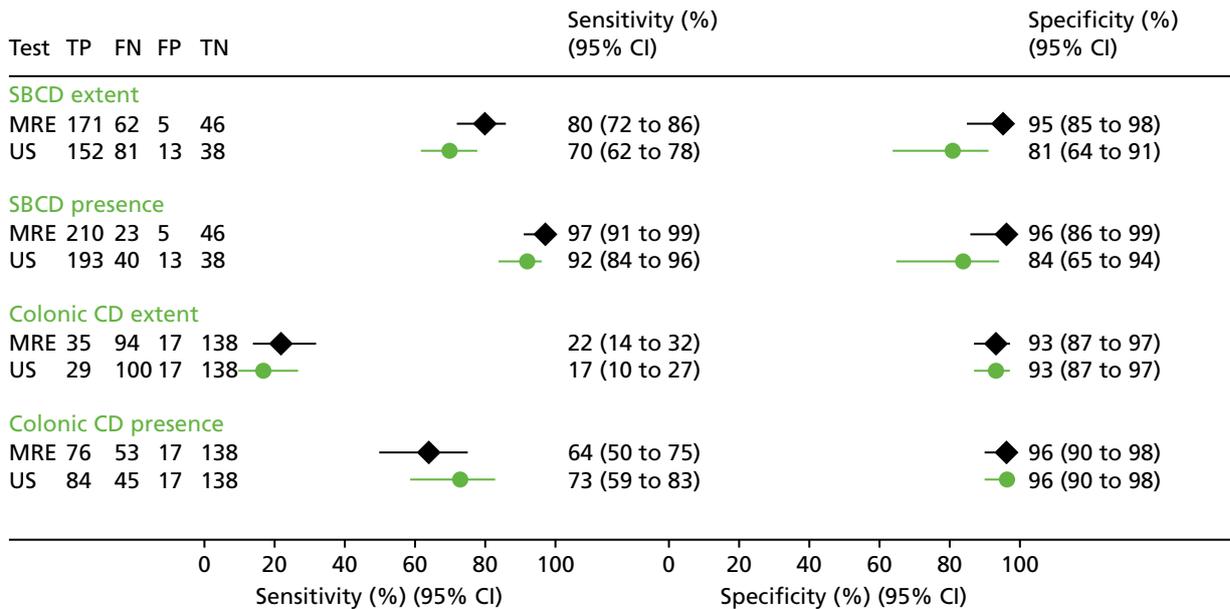


FIGURE 3 Sensitivity and specificity of MRE and US for SBCD and colonic CD extent and presence against the consensus reference standard. Sensitivity, specificity and 95% CIs are analysed to compare test accuracy within individual participants using bivariate multilevel participant-specific (conditional) random-effects modelling. FN, false negative; FP, false positive; TN, true negative; TP, true positive. Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Secondary outcomes

Magnetic resonance enterography specificity for SBCD extent was also significantly greater than that of US [95% (95% CI 85% to 98%) vs. 81% (95% CI 64% to 91%), respectively, a difference of 14% (95% CI 1% to 27%)].

The potential impact of staging SBCD extent with either MRE or US in a theoretical 1000-participant cohort is shown in *Figure 4*.

Regardless of location, sensitivity of MRE for SBCD presence was 97% (95% CI 91% to 99%), significantly greater than that of US [92% (95% CI 84% to 96%)] [a difference of 5% (95% CI 1% to 9%)]. MRE and US specificity for SBCD presence was 96% (95% CI 86% to 99%) and 84% (95% CI 65% to 94%), respectively: a difference of 12% (95% CI 0% to 25%). The potential impact of staging SBCD presence with either MRE or US in a theoretical 1000-participant cohort is shown in *Appendix 11, Figure 11*.

There were no significant differences in sensitivity or specificity between MRE and US for colonic CD extent or presence (see *Table 7*), although for both tests sensitivities were considerably less than for SBCD. MRE was 64% (95% CI 50% to 75%) sensitive for colonic CD presence but just 22% (14% to 32%) sensitive for extent (which required correct identification of involved colonic segments); the corresponding figures for US were 73% (59% to 83%) and 17% (10% to 27%).

The sensitivity and specificity for individual small bowel and colonic segments is given in *Table 8*. Although the study was not powered to detect differences on a segmental level, MRE was significantly more sensitive than US for ileal disease [84% (95% CI 67% to 93%) vs. 56% (95% CI 38 to 73), respectively]. Sensitivity for the eight diseased duodenal segments was low, at 25% (95% CI 7% to 59%), for both MRE and US. Sensitivity for jejunal disease was 71% (95% CI 38% to 91%) for MRE and 63% (95% CI 32% to 86%) for US. For five out of six colonic segments, sensitivity was \approx 40–50% for both tests. However, sensitivity for rectal disease was significantly lower for US [22% (95% CI 13% to 35%)] than for MRE [44% (95% CI 32% to 58%)], with a difference of 22% (95% CI 9% to 35%).

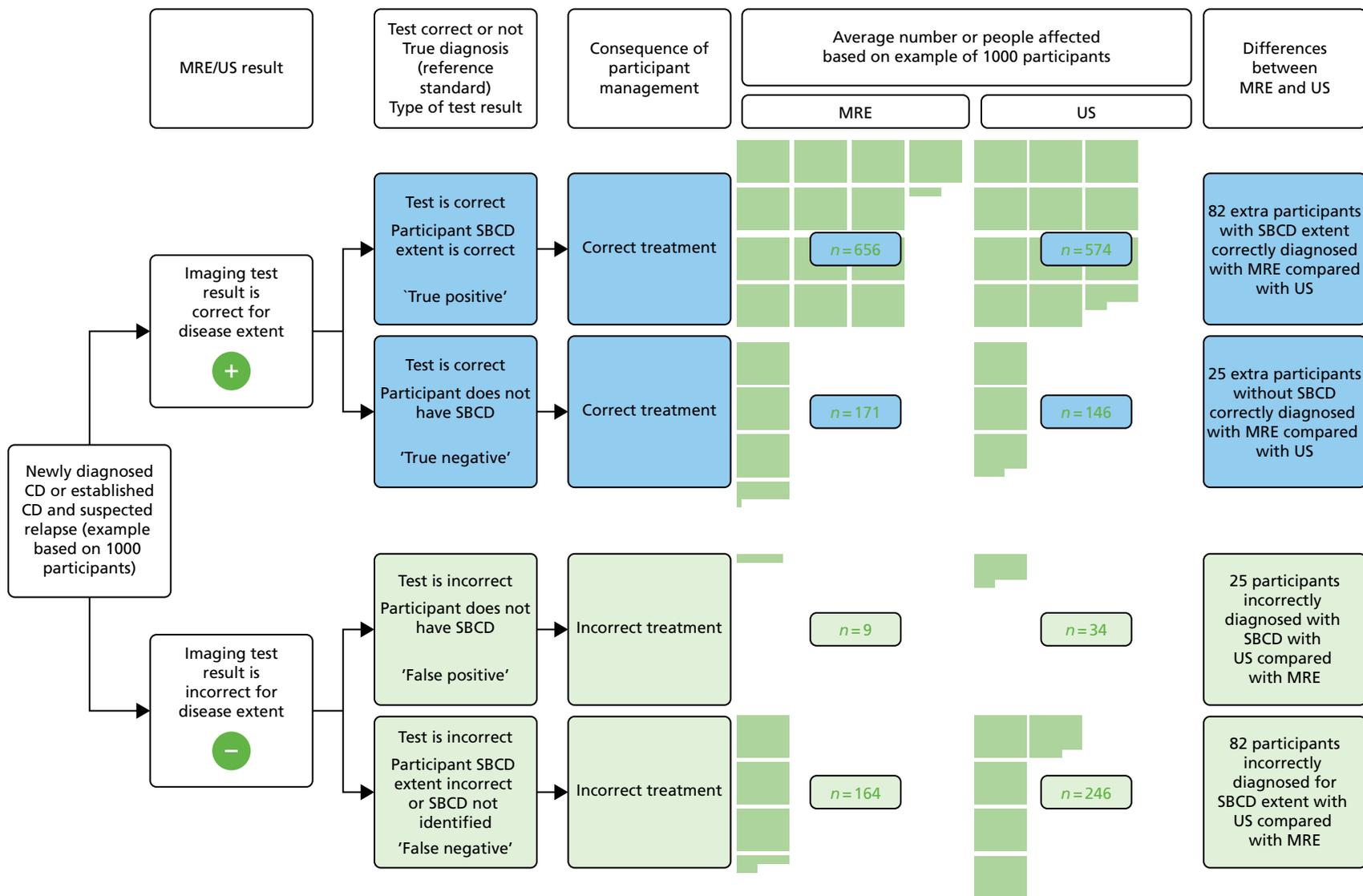


FIGURE 4 Potential impact of staging SBCD extent with either MRE or US in a theoretical 1000-participant cohort. Numbers of hypothetical participants are calculated from sensitivity and specificity results comparing test accuracy within individual participants.

TABLE 8 Per-segment sensitivity and specificity for disease presence against the consensus reference standard (both cohorts combined)

Variable	Disease-positive segments (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-negative segments (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
Small bowel segments								
Duodenum ^b	8	25 (7 to 59)	25 (7 to 59)	0 (-13 to 13; 1.000)	276	100 (99 to 100)	99 (97 to 100)	1 (0 to 3; 0.250)
Jejunum	13	71 (38 to 91)	63 (32 to 86)	8 (-29 to 46; 0.664)	271	99 (93 to 100)	99 (94 to 100)	0 (-2 to 1; 0.741)
Ileum	38	84 (67 to 93)	56 (38 to 73)	28 (8 to 49; 0.008)	246	93 (87 to 97)	93 (87 to 96)	0 (-4 to 4; 0.871)
Terminal ileum	217	96 (91 to 99)	92 (84 to 96)	4 (0 to 8; 0.051)	67	97 (90 to 99)	93 (81 to 98)	4 (-2 to 10; 0.197)
Colonic segments^c								
Caecum	78	46 (35 to 57)	46 (35 to 57)	0 (-12 to 12; 1.000)	147	96 (92 to 99)	90 (85 to 94)	6 (0 to 12; 0.036)
Ascending	67	49 (38 to 61)	49 (38 to 61)	0 (-10 to 10; 1.000)	200	96 (93 to 98)	92 (88 to 95)	4 (0 to 8; 0.058)
Transverse	61	46 (34 to 58)	44 (32 to 57)	2 (-12 to 15; 0.809)	218	97 (93 to 98)	95 (91 to 97)	2 (-1 to 5; 0.130)
Descending	59	53 (40 to 65)	41 (29 to 54)	12 (-1 to 24; 0.063)	221	98 (95 to 99)	95 (91 to 97)	3 (0 to 6; 0.033)
Sigmoid	76	46 (35 to 57)	43 (33 to 55)	3 (-11 to 16; 0.695)	203	96 (92 to 98)	93 (89 to 96)	3 (-1 to 7; 0.179)
Rectum	54	44 (32 to 58)	22 (13 to 35)	22 (9 to 35; 0.001)	228	97 (94 to 99)	93 (89 to 96)	4 (0 to 7; 0.072)

a By consensus reference standard.

b McNemar's test owing to the small number of participants with disease presence.

c Analysis for individual colonic segments uses a population-average approach.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

The sensitivity and specificity of MRE and US for SBCD and colonic CD presence and extent according to the participant cohort is shown in *Table 9*. Sensitivities of both tests for SBCD presence and extent in the new diagnosis and suspected relapse participant cohorts were very similar to those sensitivities estimated across all participants (see *Table 7*), although differences were not statistically significant (the study was not powered to detect differences in the cohorts).

However, US had significantly greater sensitivity for colonic CD presence than MRE in the new diagnosis cohort [67% (95% CI 49% to 81%) vs. 47% (95% CI 31% to 64%), respectively: a difference 20% (95% CI 1% to 39%)]. For both MRE and US, sensitivity for colonic CD presence was higher in the suspected relapse participant cohort, although the estimated sensitivity for colonic CD extent was poor for both: MRE had 17% (95% CI 9% to 30%) sensitivity for colonic CD extent in the new diagnosis cohort and 31% (95% CI 17% to 48%) sensitivity in the suspected relapse cohort. The corresponding figures for US were 9% (95% CI 4% to 19%) and 33% (95% CI 19% to 51%), respectively.

The sensitivity and specificity for individual small bowel and colonic segments according to participant cohort is given in *Appendix 12, Table 51*. In general, sensitivities of both tests for small bowel segmental disease presence in the new diagnosis and suspected relapse cohorts were very similar to those sensitivities estimated across all participants, although we note that the study was not powered to detect differences in the cohorts (see *Table 8*). MRE had 100% (95% CI 61% to 100%) sensitivity for jejunal disease in the suspected relapse cohort, compared with 43% (95% CI 16% to 75%) in the new diagnosis cohort, although there were only six and seven positive segments in each cohort, respectively. Sensitivity for colonic segmental disease was higher in the suspected relapse cohort than in the new diagnosis cohort for both MRE and US, consistent with the findings across all participants (see *Tables 7 and 8*).

Identification and localisation of small bowel Crohn's disease and colonic Crohn's disease against an ileocolonoscopy reference

Colonoscopy data were available for 186 participants (123 in the new diagnosis cohort and 63 in the suspected relapse cohort). The sensitivity and specificity of MRE and US for terminal ileal and colonic segmental disease against an ileocolonoscopy standard of reference is shown in *Table 10*.⁴⁰ MRE had a sensitivity of 97% (95% CI 91% to 99%) for terminal ileal disease presence, compared with a sensitivity of 91% (95% CI 79% to 97%) for US: a difference of 6% (95% CI -1% to 12%), which is similar to the 5% sensitivity difference between the tests for the presence of SBCD against the consensus reference standard (see *Table 7*). However, specificity was low, at 41% (95% CI 21% to 64%) for MRE and 33% (95% CI 15% to 57%) for US. Sensitivity for colonic CD presence was modest for both MRE and US [41% (95% CI 26% to 58%) and 49% (95% CI 33% to 65%)] and somewhat lower than the consensus reference standard, which included the 98 participants without ileocolonoscopy. The differences between MRE and US were not statistically significant (the study was not powered to detect differences based on a colonoscopic standard of reference alone).

The sensitivity and specificity of MRE and US for terminal ileal and colonic segmental disease against an ileocolonoscopy standard of reference according to the participant cohort is shown in *Appendix 13, Table 52*.

Extraenteric complications

Magnetic resonance enterography detected five out of seven (71%) abscesses and 18 out of 21 (86%) participants with enteric fistulae, compared with three out of seven (43%) and 11 out of 21 (52%) for US, respectively. Of the 61 participants with a stenosis considered to be causing obstruction by the consensus reference standard, MRE detected 33 (54%) and US detected 20 (33%) of these. There were 52 false-positive segments for stenosis on MRE and 45 for US.

TABLE 9 Sensitivity and specificity for disease presence and extent against the consensus reference standard according to participant cohort

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

N/A, not applicable.

a By consensus reference standard.

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

c Agreement with reference standard for disease presence in participants with disease in the small bowel, colon or both.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 10 Sensitivity and specificity for terminal ileal and colonic CD presence against an ileocolonoscopy reference (both cohorts combined)

Variable	Disease-positive participants (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-negative participants (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
Colonic CD								
Extent ^b	109	3 (1 to 11)	2 (0 to 8)	1 (-2 to 4; 0.429)	77	94 (81 to 98)	89 (73 to 96)	5 (-3 to 14; 0.240)
Presence	109	41 (26 to 58)	49 (33 to 65)	-8 (-26 to 9; 0.368)	77	95 (85 to 98)	90 (76 to 96)	5 (-3 to 13; 0.233)
	Disease-positive segments (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-negative segments (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
Small bowel segments								
Terminal ileum	105	97 (91 to 99)	91 (79 to 97)	6 (-1 to 12; 0.091)	81	41 (21 to 64)	33 (15 to 57)	8 (-14 to 30; 0.474)
Colonic segments^c								
Caecum	73	22 (14 to 33)	25 (16 to 36)	-3 (-14 to 9; 0.638)	101	72 (63 to 80)	65 (56 to 74)	7 (0 to 13; 0.043)
Ascending	62	26 (16 to 38)	23 (14 to 35)	3 (-6 to 12; 0.479)	121	88 (80 to 92)	81 (73 to 87)	7 (0 to 13; 0.043)
Transverse	54	24 (15 to 37)	24 (15 to 37)	0 (-9 to 9; 1.000)	132	92 (86 to 96)	90 (84 to 94)	2 (-2 to 6; 0.256)
Descending	58	27 (18 to 40)	24 (15 to 37)	3 (-6 to 13; 0.479)	128	95 (90 to 98)	93 (87 to 96)	2 (-1 to 6; 0.178)
Sigmoid	74	24 (16 to 35)	28 (19 to 40)	-4 (-17 to 9; 0.532)	111	94 (87 to 97)	94 (87 to 97)	0 (-6 to 6; 1.000)
Rectum	61	26 (17 to 39)	13 (7 to 24)	13 (2 to 25; 0.027)	125	97 (92 to 99)	94 (88 to 97)	3 (-2 to 8; 0.204)

a By ileocolonoscopy reference standard.

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

c Analysis for individual colonic segments uses a population-average approach.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Disease activity assessment against the consensus reference standard

Magnetic resonance enterography per-participant sensitivity for active SBCD was 96% (95% CI 92% to 99%), compared with 90% (95% CI 82% to 95%) for US: a difference of 6% (95% CI 2% to 11%), which was statistically significant (*Table 11*). Specificity for active SBCD and accuracy for active colonic CD were not significantly different between the two tests (see *Table 11*).

The sensitivity and specificity of MRE and US for detecting active SBCD and colonic CD according to the participant cohort is shown in *Appendix 14, Table 53*. Sensitivity and specificity for active SBCD according to the participant cohort were very similar to those estimated across all participants, although sensitivity for active colonic CD was generally higher in the suspected relapse cohort than in the new diagnosis cohort.

Disease activity assessment against an ileocolonoscopy reference

The sensitivity and specificity of MRE and US for active terminal ileal and colonic CD across all participants ($n = 186$) against an ileocolonoscopy standard of reference is shown in *Table 12* and, according to participant cohort, in *Appendix 15, Table 54*. Overall, MRE had significantly greater sensitivity for endoscopically diagnosed active terminal ileal disease than US [97% (95% CI 91% to 99%) vs. 86% (95% CI 70% to 95%), respectively]. There were no differences in sensitivity for active colonic CD. Sensitivity and specificity for active disease according to the participant cohort were very similar to those estimated across all participants.

Equivocal magnetic resonance enterography and ultrasonography findings

The numbers of MRE and US equivocal observations (i.e. marked with a confidence score of 3 or 4 by the practitioner) for disease presence are shown in *Table 13*. Overall, there were relatively few instances of equivocal scores for either SBCD presence [3% (9/284) for MRE; 6% (17/284) for US] or for colonic CD presence [4% (12/284) for MRE; 6% (18/284) for US]. Of the nine equivocal MRE scores for SBCD presence, eight were disease positive by the consensus reference standard and one was disease negative. Conversely, of the 12 equivocal MRE scores for colon disease presence, five were disease positive by the consensus reference standard and seven were disease negative.

Of the 17 equivocal US scores for SBCD presence, 12 were disease positive by the consensus reference standard and five were disease negative. Of the 18 equivocal US scores for colon disease presence, 13 were disease positive by the consensus reference standard and five were disease negative.

Overall, of the 21 MRE scores for small bowel or colonic CD presence, 12 were in the new diagnosis cohort and nine were in the suspected relapse cohort. Of the 35 equivocal US scores, 20 were in the new diagnosis cohort and 15 were in the suspected relapse cohort.

The number of equivocal scores for disease activity (*Table 14*) was higher than for disease presence. Specifically, for small bowel activity, 7% (19/284) of scores were equivocal for MRE and 11% (31/284) of scores were equivocal for US; for colonic CD activity, 6% (17/284) of scores were equivocal for MRE and 7% (19/284) of scores were equivocal for US.

Equivocal magnetic resonance enterography and ultrasonography findings: sensitivity analysis for the primary outcome

When equivocal scores were treated as disease negative, MRE sensitivity for SBCD extent (i.e. presence and correct segmental location) was 75% (95% CI 67% to 82%), compared with 60% (95% CI 51% to 68%) for US: a difference of 15% (95% CI 6% to 25%), which was statistically significant (*Table 15*). However, specificity was not significantly different, increasing to 90% (95% CI 77% to 96%) for US compared with 96% (95% CI 86% to 99%) for MRE.

Perceptual errors on magnetic resonance enterography and ultrasonography

The number of perceptual errors for MRE and US based on retrospective consensus panel review is shown in *Appendix 16, Table 55*. Detecting perceptual errors is intended to improve understanding of the maximum theoretical accuracy of the technology after correcting for any reader errors.

TABLE 11 Per-participant sensitivity and specificity for the presence of active disease against the consensus reference standard (both cohorts combined)

Variable	Participants with active disease (n) ^a	Sensitivity, % (95% CI; p-value)			Participants with inactive disease (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
Active SBCD	209	96 (92 to 99)	90 (82 to 95)	6 (2 to 11; 0.010)	75	83 (68 to 92)	77 (60 to 88)	6 (-8 to 20; 0.376)
Active colonic CD	126	63 (48 to 76)	66 (51 to 79)	-3 (-18 to 13); 0.735)	158	97 (91 to 99)	98 (94 to 99)	-1 (-4 to 1; 0.304)
Active SBCD and colonic CD ^b	251	77 (68 to 85)	66 (56 to 75)	11 (1 to 21; 0.024)	33	28 (10 to 56)	28 (10 to 56)	0 (-26 to 26; 1.000)

a By consensus reference standard.

b Agreement with reference standard for the presence of active disease in participants with disease in the small bowel, colon or both.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 12 Sensitivity and specificity for presence of active terminal ileal and colonic CD vs. ileocolonoscopy reference (both cohorts combined)

Variable	Disease-active participants (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-inactive participants (n) ^a	Specificity, % (95% CI; p-value)		
		MRE	US	Difference		MRE	US	Difference
Active terminal ileum disease	100	97 (90 to 99)	86 (70 to 95)	11 (1 to 22; 0.031)	86	45 (25 to 67)	38 (19 to 60)	7 (-14 to 29; 0.497)
Active colonic CD	90	44 (25 to 63)	44 (25 to 63)	0 (-20 to 20; 1.000)	96	93 (83 to 98)	94 (84 to 98)	-1 (-6 to 5; 0.808)

a By ileocolonoscopy reference standard.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 13 Number of equivocal MRE and US findings for per-participant disease presence (both cohorts combined)

Variable	Participants (n)	Equivocal scores for disease presence (n)					
		MRE			US		
		3	4	3 or 4	3	4	3 or 4
Small bowel							
Disease positive	233	2	6	8	1	11	12
Disease negative	51	0	1	1	3	2	5
Total	284	2	7	9	4	13	17
Colon							
Disease positive	129	1	4	5	3	10	13
Disease negative	155	0	7	7	3	2	5
Total	284	1	11	12	6	12	18

TABLE 14 Number of equivocal MRE and US findings for per-participant disease activity (both cohorts combined)

Variable	Participants (n)	Equivocal scores for disease presence (n)					
		MRE			US		
		3	4	3 or 4	3	4	3 or 4
Small bowel							
Disease active	209	8	7	15	7	17	24
Disease inactive	75	2	2	4	3	4	7
Total	284	10	9	19	10	21	31
Colon							
Disease active	125	1	7	8	5	12	17
Disease inactive	158	3	6	9	1	1	2
Total	284	4	13	17	6	13	19

Overall, the rates of perceptual error were relatively low, and similar between MRE and US. Specifically, across the 284-participant cohort, 22 (8%) participants had perceptual errors on MRE, compared with 33 (12%) on US. Similarly, in the colon, 25 (9%) participants had perceptual errors on MRE, compared with 23 (8%) on US. For both modalities, all perceptual errors were false negative, with no false-positive errors.

Identification and localisation of small bowel and colonic Crohn's disease against the consensus reference standard according to recruitment site

The sensitivity and specificity for SBCD extent according to the recruitment site is shown in *Table 16*. Data are combined for the lower-recruiting sites.

TABLE 15 Per-participant sensitivity and specificity for SBCD extent against the consensus reference standard (both cohorts combined). Equivocal MRE and US results considered disease negative

Variable	Disease-positive participants (n) ^a	Sensitivity, % (95% CI; <i>p</i> -value)			Disease-positive participants (n) ^a	Specificity, % (95% CI; <i>p</i> -value)		
		MRE	US	Difference		MRE	US	Difference
SBCD extent ^b	233	75 (67 to 82)	60 (51 to 68)	15 (6 to 25); 0.001	51	96 (86 to 99)	90 (77 to 96)	6 (-3 to 15; 0.209)

a By consensus reference standard.
b Agreement with reference standard for disease presence and segmental location. Equivocal MRE and US results considered disease negative.

TABLE 16 Per-participant sensitivity and specificity for SBCD extent against the consensus reference standard according to recruitment site (both cohorts combined)

Variable	Disease-positive participants (n) ^a	Sensitivity, % (95% CI; <i>p</i> -value)			Disease-negative participants (n) ^a	Specificity, % (95% CI; <i>p</i> -value)		
		MRE	US	Difference		MRE	US	Difference
UCLH, London	93	87 (74 to 94)	65 (48 to 78)	22 (7 to 37; 0.003)	23	83 (63 to 93)	70 (49 to 84)	13 (-13 to 39)
St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds	30	92 (74 to 98)	79 (53 to 93)	13 (-6 to 32; 0.189)	17	100 (82 to 100)	76 (53 to 90)	24 (-3 to 50; 0.125)
Queen Alexandra Hospital, Portsmouth	48	70 (47 to 86)	97 (87 to 99)	-27 (-46 to -8; 0.006)	2	100 (34 to 100)	0 (0 to 66)	100 (50 to 150; 0.500)
Other hospitals ^b	62	83 (66 to 93)	55 (35 to 74)	28 (8 to 48; 0.006)	9	89 (57 to 98)	100 (70 to 100)	-11 (-43 to 21; 1.000)

a By consensus reference standard.
b Includes St Mark's Hospital, Harrow; Royal Free Hospital, London; Ninewells Hospital, Dundee; John Radcliffe Hospital, Oxford; and St George's Hospital, London. Analysis per recruitment site uses McNemar's test for specificity owing to small number of participants with no disease. Agreement with reference standard for disease presence and segmental location.

Chapter 4 Diagnostic benefit of oral contrast administration (small intestine contrast-enhanced ultrasonography)

Introduction

Conventional enteric US relies on graded compression of the unprepared bowel wall. However, there are theoretical advantages to distending the bowel lumen with an oral contrast agent prior to the examination (akin to distending the bowel prior to MRE), referred to as SICUS or hydrosonography. Distension of the bowel lumen could facilitate better visualisation of the bowel wall, aiding identification and characterisation of CD. Indeed, there are data suggesting that SICUS may have higher accuracy than conventional US. In a small study of 28 participants, Calabrese *et al.*⁴¹ reported that SICUS had greater accuracy than conventional US for SBCD detection, particularly in the proximal bowel (100% vs. 96%, respectively), and for stricture detection (94% vs. 67%, respectively). Similarly, in a study of 57 participants with CD, sensitivity for SBCD against a BaFT reference was 98% for SICUS compared with 87% for conventional US.⁴² Using a reference standard of barium enteroclysis and endoscopy in 102 participants with CD, Parente *et al.*⁴³ found a small difference in sensitivity for disease detection between SICUS and conventional US (96.1% vs. 91.4%, respectively), which was more striking for stricture detection (89% vs. 74%, respectively). High sensitivity of SICUS for stricture detection (97.5%) was also reported by Pallotta *et al.*⁴⁴ in a selected cohort of 49 participants undergoing surgery for CD, and in a retrospective review of 67 participants undergoing MRE and/or SICUS prior to subsequent surgery⁴⁵ there was high agreement between the two tests for strictures ($\kappa = 0.84$) and fistulae ($\kappa = 0.61$).

The METRIC trial afforded an opportunity to prospectively compare the diagnostic accuracy of SICUS and conventional US for SBCD extent in a subset of participants against the construct reference standard.

Methods

A subset of participants was recruited from the main METRIC study cohort (see *Chapter 2*). Inclusion/exclusion criteria, recruitment pathways and participant follow-up are detailed in *Chapter 2*. Four out of the eight recruitment sites expressed a willingness to contribute participants to the substudy. The overall recruitment target was 75, which was a pragmatic figure based on the number of participating sites and expected rate of participant consent. At participating sites, participants recruited to the main METRIC study were invited to take part in the optional SICUS substudy after appropriate written and verbal explanation. Participants signed an additional consent form if they agreed to take part.

All participants underwent conventional US as part of the main METRIC study, as described in *Chapter 2*. Thereafter, participants also underwent an additional SICUS performed by the same practitioner who performed their conventional US (to remove the effects of interobserver variation and directly isolate the diagnostic impact of SICUS). The practitioners were therefore by definition unblinded to their own findings on the conventional US, but all other blinding safeguards were in place as for the main METRIC study.

Sites were encouraged to perform SICUS on the same day as the conventional US, 50–60 minutes after participants had ingested > 1000 ml of oral contrast agent, although an interval of up to 2 weeks between US and SICUS was permissible. The choice of oral contrast agent was at the discretion of the recruitment site. Because of the detrimental effects of oral contrast agents (such as diarrhoea and abdominal pain), it was permitted to perform SICUS immediately after MRE to utilise the bowel distension afforded by oral

contrast already ingested prior to MRE. Practitioners detailed their findings on SICUS using a CRF identical to that used for the conventional US, indicating if the examination had been performed immediately following MRE.

Outcomes

The primary outcome was the difference in per-participant sensitivity between US and SICUS for the correct identification and localisation of SBCD, irrespective of activity, against the consensus reference standard (see *Chapter 2*). Disease reported as equivocal was treated as positive for disease presence. Secondary outcome measures were the difference in per-participant specificity of SICUS and US for the correct identification and localisation of SBCD, irrespective of activity, and the per-participant sensitivity and specificity for identification of colonic CD. Sensitivity for the detection of small bowel stenosis was also examined.

Direct comparison of sensitivity and specificity differences between SICUS and US were calculated using bivariate multilevel participant-specific (conditional) random-effects models from paired data using meqrlogit in Stata 14.2. Statistical significance was based on 95% CIs. There were no missing data for per-participant diagnosis of disease extent in the small bowel or colon.

Results

A total of 75 participants were recruited to the SICUS substudy, of whom 11 were among the 51 subsequently excluded from the main study cohort (see *Chapter 3*), leaving a total of 64 participants. Of these, 26 had a standalone SICUS study following dedicated oral contrast preparation (three with mannitol and 23 with oral gastrograffin solution) and 38 had SICUS performed immediately after MRE (37 with oral mannitol and one with dilute oral barium solution).

The comparative sensitivity and specificity of SICUS and conventional US for SBCD and colonic CD presence is shown in *Table 17*.

The sensitivity [71% (95% CI 58% to 81%)] and specificity [86% (95% CI 49% to 97%)] for SBCD extent were identical between SICUS and conventional US. The specificity for colonic CD was non-significantly higher for SICUS than for conventional US [92% (95% CI 80% to 97%) and 82% (95% CI 68% to 91%), respectively].

Overall, there were 14 segments deemed to be stenosed and causing obstruction by the consensus reference standard. Both US and SICUS detected 9 out of 14 (64%) of these segments. There were 626 small bowel segments without stenosis by the final consensus reference. There were 11 false-positive diagnoses of stenosis on US and seven on SICUS.

TABLE 17 Per-participant sensitivity and specificity of SICUS and conventional US for disease presence and extent against the consensus reference standard (both cohorts combined)

Variable	Disease-positive participants (n) ^a	Sensitivity, % (95% CI; p-value)			Disease-negative participants (n) ^a	Specificity, % (95% CI; p-value)		
		SICUS	US	Difference		SICUS	US	Difference
SBCD extent ^b	57	71 (58 to 81)	71 (58 to 81)	0 (-2 to 2; 1.000)	7	86 (49 to 97)	86 (49 to 97)	0 (-54 to 54; 1.000)
Colonic CD extent ^b	24	17 (7 to 36)	13 (4 to 31)	4 (-8 to 16; 1.000)	40	92 (80 to 97)	82 (68 to 91)	10 (-2 to 22; 0.125)

^a By consensus reference standard.
^b Agreement with reference standard for disease presence and segmental location.

Chapter 5 Interobserver variation in the interpretation of enteric ultrasonography and magnetic resonance enterography

Introduction

Quantifying interobserver variability is an important part of the evaluation of any medical imaging technology. Interpretation of both MRE and US is potentially complex. Both rely on practitioners interpreting various signs that can indicate CD, including bowel wall thickening, changes in MR signal/ultrasonic echogenicity, enteric vascularity (either via Doppler US or contrast enhancement) and observations in the adjacent soft tissues such as the appearance of mesenteric fat. Findings can be subtle, particularly in early disease. The ease or otherwise of the observations is also influenced by the technical quality of the examination. For example, collapsed bowel in MRE can give the impression of bowel wall thickening and therefore mimic abnormal bowel.

To date there has been relatively little research into observer variability in either MRE or US. Although several studies have investigated interobserver variability for the imaging signs of CD, such as bowel wall thickness and length of disease involvement,^{13,46–49} few have focused on primary diagnostic accuracy. In a study involving 50 participants and four radiologists, Jensen *et al.*⁴⁷ reported only moderate agreement for the presence or absence of SBCD using MRE ($\kappa = 0.48$; 54% agreement between all four readers). Conversely, in a study of 103 participants with CD, Parente *et al.*⁴³ report very good agreement for the location of SBCD between two experienced practitioners ($\kappa = 0.91$) using US.

The METRIC trial afforded an opportunity to prospectively compare interobserver variability in MRE and US interpretation, incorporating performance against the construct reference standard.

Methods

Ultrasonography

A subset of participants was recruited from the main METRIC study cohort (see *Chapter 2*). Inclusion/exclusion criteria, recruitment pathways and participant follow-up are detailed in *Chapter 2*. Two out of the eight recruitment sites expressed a willingness to contribute participants to the substudy and seven practitioners participated. The overall recruitment target was 40, which was a pragmatic figure based on the number of participating sites and expected rate of participant consent. At participating sites, participants recruited to the main METRIC study were invited to take part in the optional interobserver substudy after appropriate written and verbal explanation. Participants signed an additional consent form if they agreed to take part.

All participants underwent conventional US as part of the main METRIC study (see *Chapter 2*). Thereafter, participants also underwent additional US performed by a practitioner who was different from the practitioner who performed the first (trial) US. The experience and training credentials matched those of the main METRIC study and the second practitioner was also participating in the main METRIC trial. The practitioners were fully blinded to each other's interpretation. All other blinding safeguards were in place as for the main METRIC trial.

Sites were encouraged to perform the interobserver US on the same day as the main trial US, although an interval of up to 2 weeks was permissible. Radiologist/sonographers detailed their findings using the same CRF used for the conventional US interpretation.

Magnetic resonance enterography

All participants recruited to the main METRIC study gave permission for their imaging data sets to be used for the interobserver variation substudy as part of the study consent process. Overall, 27 radiologists from all eight recruitment sites participated. Their experience and training credentials matched those of the main METRIC study (see *Chapter 2*), and for the most part they were also interpreting MRE as part of the main METRIC study.

Magnetic resonance enterography examinations of participants recruited to the main METRIC study were sent by recruitment sites to a central facility at UCLH via anonymised compact discs. They were then uploaded onto an online viewing platform (3Dnet®; Biotronics 3D Ltd, London, UK). The platform has all the image viewing functionality of a standard PACS but can be accessed from any personal computer (PC) via the internet with appropriate password credentials. The system allowed participating radiologists to access the MRE examinations from any geographical location without the need to receive copies on compact disc or hard drive, thus facilitating study efficiency.

Consecutive cases were selected from the database of uploaded examinations on the 3Dnet platform and the study statistician ensured that no radiologist was given an examination they had already interpreted as part of the main METRIC trial. Each selected case was read three times in total (once as part of the main METRIC trial and twice more by two different observers as part of the interobserver variability substudy). Cases were allocated to readers randomly for each read, with randomisations ensuring that readers had not previously read the same case in the main trial or previous reads.

Once radiologists were allocated their cases, in their own time (but within 6 weeks) they interpreted the examinations, noting their findings on a CRF based on that used for the main METRIC study. As part of a separate study investigating the contribution of MRE sequences to diagnostic accuracy (see *Chapter 6*), radiologists first interpreted T2-weighted and steady-state free precession gradient echo images alone, then T2-weighted and steady-state free precession gradient echo images and diffusion-weighted images and, finally, T2-weighted, steady-state free precession gradient echo, diffusion-weighted and contrast-enhanced sequences together, recording their findings on separate CRFs after each sequence block. For the interobserver study, only the final read based on all sequences was used for analysis.

Analysis

Ultrasonography

Equivocal confidence scores (3 and 4) were treated as positive for disease. Data were analysed for the new diagnosis and suspected relapse cohorts separately.

Analysis relates to comparisons of the two US reads in a participant. The outcomes refer to the agreement between two reads for the correct identification and localisation of SBCD and colonic CD, and the presence of SBCD and colonic CD regardless of disease activity.

Interobserver variability was measured by agreement between two radiologists, grouping results by the consensus reference as positive, negative and across all participants, with 95% CIs based on paired proportions. Prevalence-adjusted bias-adjusted kappas (PABAKs) were also reported.⁵⁰ Agreement between the reads regardless of the reference standard were reported. Data were analysed for the new diagnosis and suspected relapse cohorts separately.

Kappa statistics were interpreted as follows: 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–0.99, almost perfect agreement.⁵¹

Sensitivity and specificity were calculated by study reader. Statistical analysis was performed using Stata 14.2.

Magnetic resonance enterography

Analysis relates to comparisons of the three MRE reads in a participant. The outcomes refer to the average paired agreement between three reads for the correct identification and localisation of SBCD and colonic CD, and the presence of SBCD and colonic CD regardless of disease activity.

Interobserver variability was measured by average paired agreement between radiologists, grouping results by the consensus reference as positive, negative and across all participants, with 95% CIs based on paired proportions. PABAKs were also reported.⁵⁰ Data were analysed for the new diagnosis and suspected relapse cohorts separately.

Kappa statistics were interpreted as for US based on published interpretation thresholds.⁵¹

Sensitivity and specificity were calculated by study reader.

Statistical analysis was performed using Stata 14.2.

Outcomes

The primary outcome was variation between readers in per-participant sensitivity for the correct identification and localisation of SBCD, irrespective of activity, against the consensus reference standard (see *Chapter 2*). Secondary outcome measures were the difference in per-participant specificity for the correct identification and localisation of SBCD irrespective of activity, and the per-participant sensitivity and specificity for identification of colonic CD.

Results

Ultrasonography

In total, seven practitioners participated in the study examining 43 participants. Of these participants, five were subsequently withdrawn as they did not have a final diagnosis of CD, leaving a final study cohort of 38 participants (11 new diagnosis and 27 suspected relapse). Their contributory read number and individual sensitivity and specificity for SBCD extent against the consensus reference standard is given in *Appendix 17, Table 56*. The estimates of reader sensitivity and specificity should be treated with caution, because for most readers these are based on very small numbers of participants and will be influenced by the variability in interpretative difficulty between participants. In total, two readers from one recruitment site (UCLH) performed 59 out of 78 (78%) of the paired reads, reflecting difficulties with running the substudy across multiple sites. One reader had 15 years of enteric US experience and the other had 4 years of enteric US experience.

The level of agreement between the practitioners according to the participant cohort is shown in *Table 18*.

Agreement for SBCD extent was at best only fair: 64% agreement with the consensus reference standard (PABAK 0.27) and 56% agreement with the consensus reference standard (PABAK 0.11) for new diagnosis and suspected relapse participants, respectively. For the presence of SBCD, both reads agreed with the consensus reference standard in 9 out of 11 (82%) new diagnosis participants (PABAK 0.64) and 22 out of 27 (81%) suspected relapse participants (PABAK 0.63), suggesting substantial agreement for both cohorts. Agreement tended to be higher in participants with SBCD than in those without.

Reads disagreed with each other for SBCD presence or absence in 1 out of 11 and 4 out of 27 new diagnosis and suspected relapse participants, respectively (not considering the final consensus reference standard diagnosis). Overall, both reads agreed on the presence or absence of disease in 138 out of 152 (91%) small bowel segments. Of 152 small bowel segments, both reads disagreed with the reference standard in five segments, and one read disagreed with the reference standard in a further 14 segments.

TABLE 18 Ultrasonography interobserver variability

Variable	New diagnosis participants (N = 11)						Suspected relapse participants (N = 27)					
	SBCD						SBCD					
	Positive ^a (n = 11)		Negative ^a (n = 0)		Overall agree (%) ^b	PABAK	Positive ^a (n = 19)		Negative ^a (n = 8)		Overall agree (%) ^b	PABAK
R1	R2	Positive agree, % (95% CI)	Negative agree, % (95% CI)	R1			R2	Positive agree, % (95% CI)	Negative agree, % (95% CI)			
SBCD												
Presence	9	10	82 (52 to 95)		82	0.64	18	18	95 (75 to 99)	50 (22 to 78)	81	0.63
Extent	9	10	64 (35 to 85)		64	0.27	18	18	58 (36 to 77)	50 (22 to 78)	56	0.11
Variable	Colonic CD						Colonic CD					
	Positive ^a (n = 8)		Negative ^a (n = 3)		Overall agree (%) ^b	PABAK	Positive ^a (n = 15)		Negative ^a (n = 12)		Overall agree (%) ^b	PABAK
	R1	R2	Positive agree, % (95% CI)	Negative agree, % (95% CI)			R1	R2	Positive agree, % (95% CI)	Negative agree, % (95% CI)		
Colonic CD												
Presence	5	4	50 (22 to 78)	100 (44 to 100)	64	0.27	14	12	80 (55 to 93)	75 (47 to 91)	78	0.56
Extent	5	4	13 (2 to 47)	100 (44 to 100)	36	-0.27	14	12	13 (4 to 38)	75 (47 to 91)	41	-0.19

R1, first read; R2, second read.
a By consensus reference standard.
b Both readers agree with consensus reference standard.

Agreement for the presence of colonic CD was inferior to that for SBCD, particularly for those participants newly diagnosed. Both reads agreed with the consensus reference standard in 7 out of 11 (64%) new diagnosis participants (PABAK 0.27) and in 21 out of 27 (78%) suspected relapse participants (PABAK 0.56), suggesting fair and moderate agreement, respectively. Agreement for colonic CD extent was fair for those participants newly diagnosed and fair for participants with suspected relapse (see *Table 18*).

Magnetic resonance enterography

In total, 27 practitioners contributed to the study, reading MRE data sets from 73 participants (28 new diagnosis and 45 suspected relapse). Their contributory read number and individual sensitivity and specificity against the consensus reference standard are given in *Appendix 18, Table 57*. As with US, the estimates of reader sensitivity and specificity should be treated with caution where these are based on very small numbers of participants given the variability in the interpretative difficulty.

The level of agreement between the practitioners according to the participant cohort is shown in *Table 19*.

Agreement for SBCD extent was at best only slight [53% average agreement between pairs of readers with the consensus reference standard (PABAK 0.07) for suspected relapse participants]. Agreement tended to be higher in participants without SBCD than in those participants with SBCD. On average, for the presence of SBCD, pairs of readers agreed with the consensus reference standard in 18 out of 26 (69%) disease-positive new diagnosis participants and one out of two (50%) disease-negative new diagnosis participants (PABAK 0.36), indicating fair agreement. In the suspected relapse cohort, on average, for the presence of SBCD, pairs of readers agreed with the consensus reference standard in 25 out of 33 (76%) disease-positive participants and 9 out of 12 (75%) disease-negative participants (PABAK 0.51), indicating moderate agreement.

Agreement for the presence of colonic CD was a little inferior to that for SBCD. On average, pairs of readers agreed with the consensus reference standard in 6 out of 14 (43%) disease-positive new diagnosis participants and 12 out of 14 (68%) disease-negative new diagnosis participants (PABAK 0.21), indicating fair agreement. In the suspected relapse participants, on average, pairs of readers agreed with the consensus reference standard in 10 out of 17 (59%) disease-positive participants and 17 out of 28 (61%) disease-negative participants (PABAK 0.21), indicating fair agreement. Agreement for colonic CD extent was at best fair (for participants with suspected relapse).

TABLE 19 Magnetic resonance enterography interobserver variability

Variable	New diagnosis participants (N = 28)								Suspected relapse participants (N = 45)						
	SBCD								SBCD						
	Positive ^a (n = 26)				Negative ^a (n = 2)				Positive ^a (n = 33)				Negative ^a (n = 12)		
	R1	R2	R3	Average positive agree, % (95% CI)	Average negative agree, % (95% CI)	Overall agree (%) ^b	κ	R1	R2	R3	Average positive agree, % (95% CI)	Average negative agree, % (95% CI)	Overall agree (%) ^b	PABAK	
SBCD															
Presence	22	23	19	69 (50 to 83)	NR ^c	68	0.36	29	25	26	76 (59 to 87)	75 (47 to 91)	76	0.51	
Extent	19	17	14	42 (26 to 61)	NR ^c	43	-0.14	24	16	17	45 (30 to 62)	75 (41 to 91)	53	0.07	
Variable	Colonic CD								Colonic CD						
	Positive ^a (n = 14)								Positive ^a (n = 17)						
	Positive ^a (n = 14)				Negative ^a (n = 14)				Positive ^a (n = 17)				Negative ^a (n = 28)		
	R1	R2	R3	Average positive agree, % (95% CI)	Average negative agree, % (95% CI)	Overall agree (%) ^b	κ	R1	R2	R3	Average positive agree, % (95% CI)	Average negative agree, % (95% CI)	Overall agree (%) ^b	PABAK	
Colonic CD															
Presence	8	9	8	43 (21 to 67)	79 (52 to 92)	61	0.21	14	11	11	59 (36 to 78)	61 (42 to 76)	60	0.20	
Extent	7	5	4	14 (4 to 40)	79 (52 to 92)	46	-0.07	11	7	7	61 (42 to 76)	61 (42 to 76)	49	-0.02	

NR, not reported; R1, first (trial) read; R2, second (interobserver) read; R3, third (interobserver) read.

a By consensus reference standard.

b Average agreement between pairs of reads with consensus reference standard.

c Owing to small sample size, only two participants without disease (all reads correctly identified; no disease in one participant and, for the other participant, read reported a false-positive equivocal finding).

Chapter 6 Influence of sequence selection on magnetic resonance enterography diagnostic accuracy

Introduction

International consensus statements recommend a minimum data set of MRI sequences as part of standard MRE protocols, T2-weighted images (with and without fat saturation), steady-state free precession gradient echo images and T1-weighted images after intravenous gadolinium injection.⁵² T2-weighted images and steady-state free precession gradient echo images mainly provide information on bowel anatomy and structure based on the magnetic properties of the native tissue, fibrosis and oedema having differing signal characteristics for example. T1-weighted images after intravenous gadolinium injection reflect the vascularity of the bowel, with evidence linking increased perfusion with disease activity.^{53,54} A typical acquisition protocol takes around 30 minutes and the reporting radiologist must interpret around six individual image blocks per participant.

In recent years there have been increasing concerns about routine use of intravenous gadolinium injections and potentially detrimental long-term retention in the brain.⁵⁵ This is of particular concern in the CD population who often undergo many MRE examinations over the course of their disease. The use of diffusion-weighted imaging, which does not require intravenous contrast injection, is attracting increasing attention.⁵² Diffusion-weighted imaging reflects the changes in water motility caused by interactions with cell membranes, macromolecules and tissue structures that modify the Brownian motion of fluid. The histopathological changes of active CD, such as cellular infiltration and mural oedema, influence the signal on diffusion-weighted imaging and there are increasing data supporting its role in detection of CD and in classifying disease activity,⁵⁶ with reported sensitivity and specificity of 92.9% and 91.0%, respectively.⁵⁷ For example, in a crossover-design non-inferiority study of 44 participants, Seo *et al.*⁵⁸ found no significant difference in accuracy for active disease between T2-weighted images supplemented by diffusion-weighted imaging compared with conventional T2-weighted images supplemented by T1-weighted post-gadolinium sequences, reproducing the earlier findings of Neubauer *et al.*⁵⁹ in a paediatric cohort.

Reducing the number of sequences required for standard MRE protocols would also have efficiency advantages, reducing the overall scan acquisition time and potentially radiologist reporting time. Removing the need for intravenous contrast injection would also spare participants an injection, which is likely to improve participant experience.

Methods

The study was an additional component of the interobserver variation study, and methods are described in *Chapter 5*. In brief, for their allocated data sets, participating radiologists first interpreted T2-weighted and steady-state free precession gradient echo images alone, then T2-weighted and steady-state free precession gradient echo images and diffusion-weighted images, and then, finally, T2-weighted and steady-state free precession gradient echo images, diffusion-weighted images and contrast-enhanced images together, recording their findings on separate CRFs after each sequence block.

In addition, as part of their MRE read for the main METRIC study, radiologists noted if and how diffusion-weighted imaging and/or contrast-enhanced imaging had added to their diagnostic confidence or actual diagnosis on the MRE reporting CRF.

Analysis

The primary outcome was the difference in per-participant sensitivity based on each MRE sequence block for the correct identification and localisation of SBCD, irrespective of activity, against the consensus reference standard. Secondary outcome measures were the difference in per-participant specificity between MRE sequence blocks for the correct identification and localisation of SBCD, irrespective of activity.

Sensitivity and specificity values were calculated by bootstrapping with replacement by participant 1999 times and taking an average value over the bootstrapped data sets. This was completed using bootstrap and centile in Stata 14.2. Statistical significance was based on 95% CIs based on empirical values from 5% and 95% centile bootstrap estimates.

Influences of MRE sequences on radiologist confidence and diagnosis were summarised descriptively.

Outcomes

The primary outcome was the difference in per-participant sensitivity across the three image sequence blocks for the correct identification and localisation of SBCD, irrespective of activity, against the consensus reference standard as detailed in *Chapter 2*. Disease reported as equivocal was treated as positive for disease presence.

Results

The sensitivity and specificity for SBCD extent according to MRE sequence block is shown in *Table 20*.

Overall, there was no increase in sensitivity or specificity with the addition of diffusion-weighted images to T2-weighted and steady-state free precession gradient echo images. Conversely, compared with T2-weighted and steady-state free precession gradient echo images alone, the addition of diffusion-weighted images and contrast-enhanced images together led to a significant -7% (95% CI -1% to -14%) drop in sensitivity for SBCD extent. In the absence of gains in diagnostic accuracy, the additional reading times for different image sequences were not calculated.

Radiologists' opinions on the effect of diffusion-weighted and contrast-enhanced images on their final diagnosis, and diagnostic confidence during their interpretation of MRE, as part of the main METRIC study are shown in *Table 21*.

TABLE 20 Sensitivity and specificity for SBCD extent according to MRE sequence block

Outcome	Percentage (95% CI)			Difference, percentage (95% CI)		
	Block 1: T2 and SSFPGE images	Block 2: T2, SSFPGE and DWI images	Block 3: T2, SSFPGE, DWI and CE images	Block 2 minus block 1	Block 3 minus block 2	Block 3 minus block 1
Sensitivity	63 (51 to 75)	61 (47 to 73)	56 (42 to 68)	-2 (-8 to 4)	-5 (-13 to 2)	-7 (-14 to -1)
Specificity	79 (57 to 100)	79 (57 to 100)	71 (50 to 93)	0 (0 to 0)	-7 (-22 to 0)	-7 (-22 to 0)

CE, contrast-enhanced; DWI, diffusion-weighted; SSFPGE, steady-state free precession gradient echo; T2, T2-weighted.

TABLE 21 Influence of diffusion-weighted imaging and contrast-enhanced imaging on radiologist diagnosis and confidence using MRE (both cohorts combined)

Radiologist assessment	Imaging, n (%)	
	Diffusion weighted ^a	Contrast enhanced ^{a,b}
a. Not helpful	80 (30)	68 (25)
b. Diagnosis unchanged but increased confidence	168 (64)	189 (70)
c. Diagnosis changed: additional disease site detected	10 (4)	10 (4)
d. Diagnosis changed: disease site discounted	0 (0)	0 (0)
e. Diagnosis changed: disease reclassified as active	4 (1)	3 (1)
f. Diagnosis changed: disease reclassified as inactive	1 (0)	0 (0)
g. Other	0 (0)	1 (0)
Missing	22	16
Total diagnosis changed [c, d, e, f (g)]	15 (6)	14 (5)
Total diagnosis unchanged [a, b (g)]	248 (94)	257 (95)
Total assessments	263	271

a More than one option indicted by the radiologist for one participant.
b One participant missing contrast-enhanced imaging.

Radiologists stated that their diagnosis was changed by diffusion-weighted imaging in 15 out of 263 (6%) MRE reads, although diagnostic confidence was increased in 168 out of 263 (64%) reads (without a change in diagnosis).

Contrast-enhanced imaging led to a change in diagnosis in 14 out of 271 (5%) reads and an increase in diagnostic confidence in 189 out of 271 (70%) reads (without a change in diagnosis).

Chapter 7 Magnetic resonance enterography and enteric ultrasonography to diagnose Crohn's disease: participant acceptability, perceived burden and preferences

Parts of this chapter have been reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

Introduction

Participant experience and acceptability influence test utility. Although MRE and US avoid ionising radiation, they have their own specific attributes that may have an impact on tolerance. For example, participants must ingest large volumes of oral contrast prior to MRE, and US requires graded abdominal compression.⁶⁰ Participants' perceptions of test 'burden' (levels of physical and psychological discomfort) may influence their compliance with the test, even if it is considered to have higher diagnostic accuracy than other tests. For example, uptake of bowel cancer screening colonoscopy is low, in part owing to participant fears of the test itself,⁶¹ which has an impact on the utility of the test.⁶² Participants may delay seeking medical attention, fearing the discomfort associated with test procedures.⁶³ To date, few data report imaging test preferences among participants with CD.

We compared the perceived burden and acceptability of MRE and US in participants recruited to the METRIC trial to identify predictors of scan preference and to examine the perceived importance of different scan attributes.

Methods

Participants

Participants recruited to the METRIC trial were given the option to complete a questionnaire documenting their experience of MRE and US and provided with paper copies of the questionnaire at the time of consent or, if this was not possible, these were posted. Participants were asked to complete the questionnaire only after all of their investigations were completed. Of the 335 participants initially consented prior to exclusions (see *Figure 2*), 324 (96.7%) agreed to take part in the substudy, of whom 159 completed the questionnaire.

Questionnaire content

Demographics

Data were collected on participant age and sex. Missing demographic data on age and sex were supplied via the central trial database. Emotional distress was assessed using the General Health Questionnaire-12 items (GHQ-12).⁶⁴

Scan recovery and overall acceptability

The questionnaire was divided into sections concerning MRE and US.

Using a nine-point scale, for each investigation participants graded their recovery time from 'immediate' to 'a week'. For analysis, data were collapsed into six categories: 'immediate', ' ≤ 30 minutes', ' ≤ 6 hours', ' ≤ 1 day', ' ≤ 3 days' and ' ≤ 1 week'.

Participants also rated how acceptable they found the two investigations on a four-point scale with the response options 'not at all acceptable', 'slightly acceptable', 'fairly acceptable' and 'very acceptable'. Participants were also asked to select the least acceptable (or worst) part of the investigation from a range of attributes provided, specific to the particular investigation. Participants were also asked if they would repeat the investigation.

Scan burden for magnetic resonance imaging and ultrasonography

A questionnaire adapted from that used to assess colonoscopy and whole-body MRI was used to quantify scan burden for MRE and US.^{65,66} In addition, abdominal bloating, diarrhoea, nausea, vomiting and sleep difficulties were added as they are of direct relevance to small bowel investigations. The questionnaire combines a series of individual items into three main domains: satisfaction, worry and discomfort. The MRE questionnaire included 31 items (7, 6 and 18 in satisfaction, worry and discomfort domains, respectively) and the US questionnaire included 28 items (7, 6 and 15 in satisfaction, worry and discomfort domains, respectively),⁶⁰ additionally including an item relating to the abdominal pressure of the probe.

Participants rated their experiences by ticking agreement on a 7-point Likert scale in which 1 and 7 were anchored to bipolar statements related to the scan, for example from 'pressure of the probe was unbearable' (1) to 'the pressure of the probe was fine' (7). Scores for each item were reverse scored, totalled and averaged so that higher scores equated to higher burden.⁶⁰

Scan preference

Participants were asked to indicate whether, if they had to undergo just one test, they would prefer MRE or US. No specific information about the scans was given in addition to that provided as part of the trial consent procedure.

Importance of scan attributes

The questionnaire also provided a list of 25 possible attributes related to the two scans that participants were asked to rate in terms of importance: 'not at all important', 'a little bit important', 'moderately important', 'very important' and 'extremely important'. The attributes were related to the scan experience itself, such as drinking oral contrast, but also to scan diagnostic accuracy, feedback of results, etc.

Statistical analysis

Independent *t*-tests and chi-squared tests were used to assess differences between (1) questionnaire responders and non-responders and (2) new diagnosis and suspected relapse cohorts, for continuous or categorical data, respectively. Differences between scan recovery time and scan acceptability were tested using related samples Wilcoxon signed-rank tests, and differences in willingness to have the different investigations again were tested using McNemar's tests. Bonferroni corrections were applied to McNemar's tests, giving a threshold for statistical significance of $p < 0.01$. Scan burden related to MRE and US was compared between different participant subgroups using Mann-Whitney *U*-tests or Kruskal-Wallis tests as appropriate. Post hoc comparisons using a Bonferroni correction were used to assess the effect of age on scan burden (threshold for statistical significance of $p < 0.01$). Analysis was performed using IBM SPSS Statistics version 24 (IBM Corporation, Armonk, NY, USA).

Results

Characteristics of participants recruited to the questionnaire study are shown in *Table 22*, and *Figure 5* is the study flow chart.

TABLE 22 Characteristics of participants in the questionnaire study

Characteristic	All participants, mean (SD) (n = 159)	Cohort, mean (SD)		Group differences
		New diagnosis (n = 84)	Suspected relapse (n = 75)	
Age (years)	38.2 (16.4)	37.3 (17.3) ^a	39.1 (15.4) ^a	$t < 1$; df = 157; $p = 0.484$
Female	94 (59.1)	47 (56.0) ^a	47 (62.7) ^a	$\chi^2 = 0.739$; df = 1; $p = 0.390$
Educational qualifications ^b				
None	8 (5.1)	4 (4.9) ^b	4 (5.3) ^a	$\chi^2 = 0.641$; df = 2; $p = 0.726$
Some	91 (58.3)	45 (55.6)	46 (61.3)	
Degree level or higher	57 (36.5)	32 (39.5)	25 (33.3)	
Ethnicity (white)	127 (92.0)	71 (93.4) ^c	56 (90.3) ^c	$\chi^2 = 0.447$; df = 1; $p = 0.504$
Newly diagnosed	84 (52.8) ^a	–	–	–
Comorbidities (at least one comorbid illness)	65 (40.9)	35 (41.7) ^a	30 (40.0) ^a	$\chi^2 = 0.046$; df = 1; $p = 0.831$
Presence of high distress (GHQ-12)	73 (48.3)	43 (51.2) ^a	30 (44.8) ^c	$\chi^2 = 0.614$; df = 1; $p = 0.433$

df, degrees of freedom; SD, standard deviation.

a No missing data.

b Missing data < 5%.

c Missing data > 5%.

Where there is missing data, % is valid %.

Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

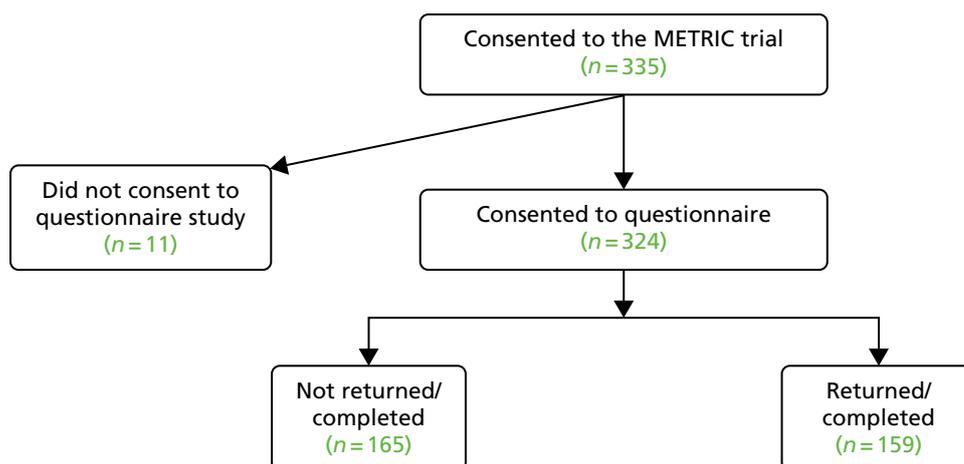


FIGURE 5 Flow chart showing the flow of participants in the study participating in the participant acceptability substudy. Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

There were no sex differences between participants completing the questionnaire ($\chi^2 = 1.606$; $df = 1$; $p = 0.205$) but responders were significantly older than non-responders, with a mean age of 38.2 years [standard deviation (SD) 16.4 years] and 33.8 years (SD 14.0 years), respectively ($t = 2.603$; $df = 322$; $p = 0.010$).

Burden score reliability

Internal reliability of subscales was good (MRI satisfaction subscale, $\alpha = 0.813$; MRI discomfort subscale, $\alpha = 0.872$; MRI worry subscale, $\alpha = 0.786$; US discomfort subscale, $\alpha = 0.868$; US satisfaction subscale, $\alpha = 0.849$; US worry subscale, $\alpha = 0.808$).

Test/scan recovery time

Magnetic resonance enterography recovery time was significantly longer than US recovery time ($p < 0.001$), with 10% (15/146) reporting immediate recovery following MRE compared with 69% (102/147) reporting immediate recovery following US (Table 23). Just 2% (3/147) of participants took > 1 day to recover after US, compared with 18% (26/146) after MRE.

Test/scan acceptability

Overall, 98.6% (144/147) of participants rated US as very or fairly acceptable. This was significantly higher than for MRE [88.3% (128/145); $p < 0.001$] (see Table 23).

TABLE 23 Comparative scan experience of MRE and US

Variable	Imaging technique, <i>n</i> (%)	
	MRE	US
Recovery time	N = 146	N = 147
Immediate	15 (10.3)	102 (69.4) ^a
≤ 30 minutes	17 (11.6)	16 (10.9)
≤ 6 hours	45 (30.8)	21 (14.3)
≤ 1 day	43 (29.5)	5 (3.4)
≤ 3 days	17 (11.6)	3 (2.0)
≤ 1 week	9 (6.2)	0 (0)
Acceptability	N = 145	N = 146
Very	66 (45.5)	126 (86.3) ^a
Fairly	62 (42.8)	18 (12.3)
Slightly	12 (8.3)	0 (0)
Not at all	5 (3.4)	2 (1.4)
Willingness to have again	N = 140	N = 131
Yes	127 (90.7)	133 (98.5)
Not sure	12 (8.6)	0 (0)
No	1 (0.7)	2 (1.5)

a Significantly different from MRE ($p < 0.001$).

Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Willingness to repeat investigations

The majority of participants were willing to repeat MRE (91%, 127/140), but the proportion was lower than for US (99%, 133/135) ($p = 0.012$) (approaching statistical significance at the $p < 0.01$ threshold).

Least acceptable part of the investigations

Figures 6 and 7 illustrate the attributes of MRE and US selected as the least acceptable. Drinking oral contrast (37.0%, 54/146) and repeated breath holding (14%, 20/146) were most commonly identified as the least acceptable aspects of MRE.

Ultrasonography was reported as being 'fine', with no least acceptable part, by 49% (73/148). Abdominal compression was reported as the least acceptable part by 30% (44/148).

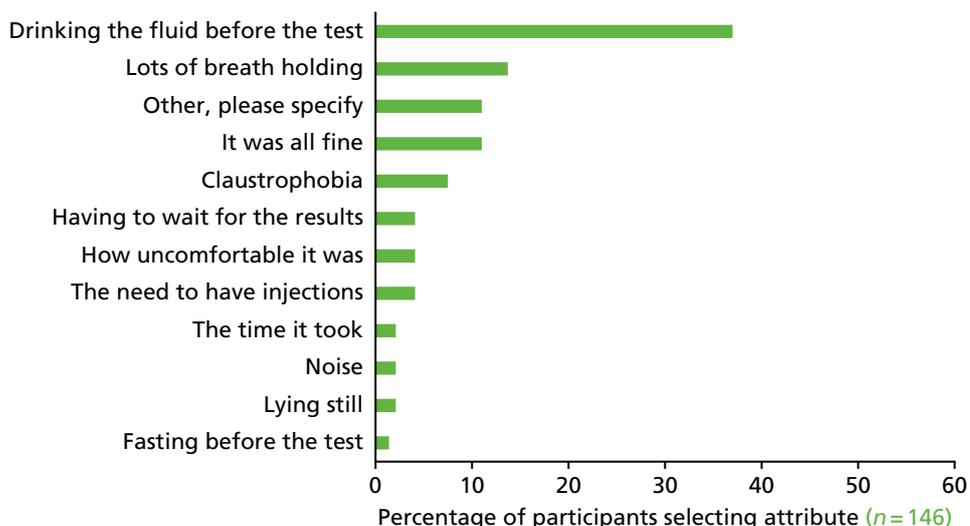


FIGURE 6 Least acceptable part of MRE. Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

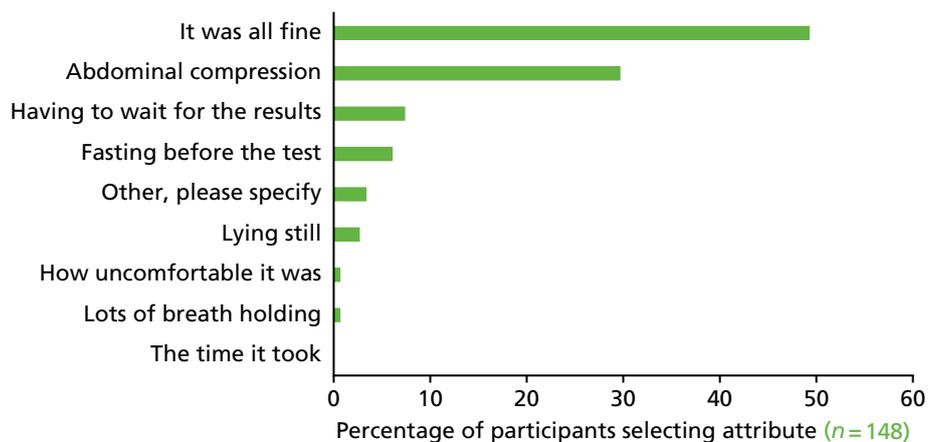


FIGURE 7 Least acceptable part of US. Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Burden of magnetic resonance enterography and ultrasonography

Burden scores for MRE and US are shown in *Table 24*. Overall burden scores were low for both tests. Participants reported higher burden during MRE than during US. This observation held true both overall and on the three subscales (discomfort, satisfaction and worry).

Perceived MRE scan burden was significantly higher among younger people ($p = 0.003$) and people with high levels of emotional distress ($p = 0.20$). Younger age ($p = 0.034$) and high levels of emotional distress ($p = 0.006$) were also associated with higher perceived burden of US.

Overall, 125 participants stated their preferred test. Of those participants, 100 (80%) preferred US to MRE.

Overall perceived importance of investigation attributes

Of the 25 attributes provided to participants, accuracy was rated as the most important attribute, followed by waiting time to diagnosis/treatment and the number of tests needed prior to final diagnosis (*Figure 8*).

Test-specific attributes, such as requirement to drink fluid, test discomfort, radiation exposure and fasting, were rated as less important and, generally, between 'a little bit important' and 'moderately important'.

TABLE 24 Perceived participant scan burden for MRE and US

Scale	Imaging technique, scan burden ^a				Wilcoxon signed-rank test
	MRE		US		
	Median score (IQR)	<i>n</i>	Median score (IQR)	<i>n</i>	
Overall	2.74 (1.35)	148	1.39 (0.89)	149	$z = 9.536; p < 0.001$
Discomfort subscale	3.08 (1.61)	148	1.33 (0.90)	149	$z = 9.558; p < 0.001$
Satisfaction subscale	1.86 (1.43)	147	1.00 (0.86)	149	$z = 0.704; p < 0.001$
Worry subscale	2.58 (1.63)	148	1.50 (1.42)	149	$z = 0.801; p < 0.001$

a 1 (low) to 7 (high).

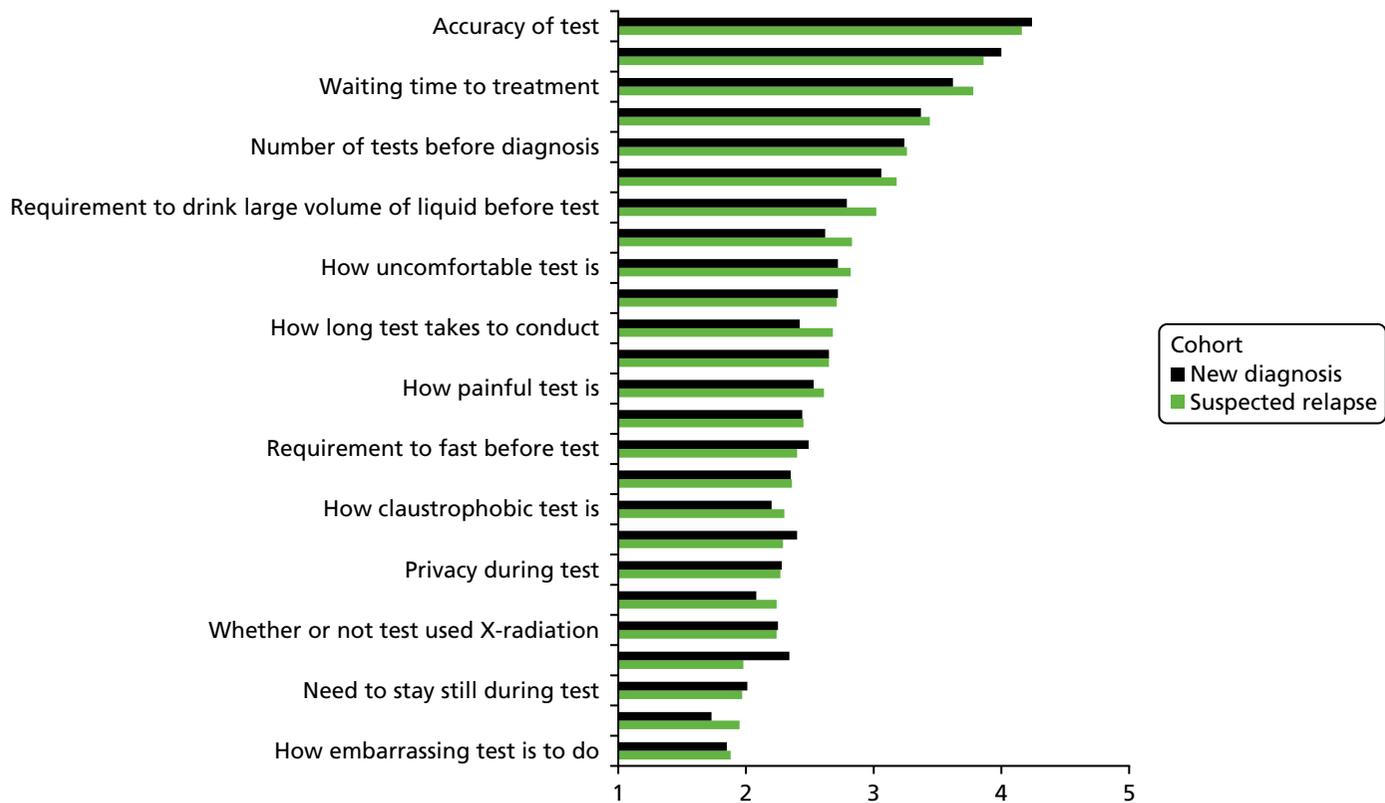


FIGURE 8 Perceived importance of different scan attributes. Reproduced with permission from Miles *et al.*⁶⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Chapter 8 Influence of oral contrast agent and ingested volume on small bowel distension and participant experience during magnetic resonance enterography

Introduction

Adequate distension of the small bowel is a prerequisite for MRE interpretation.⁶⁷ Collapsed bowel can both mimic CD and obscure CD. There are a variety of oral contrast agents used prior to MRI in clinical practice, such as polyethylene glycol⁶⁸ and mannitol,⁶⁹ and most are non-absorbable, retaining fluid in the bowel lumen to maintain distension throughout the small bowel volume.⁷⁰ A recent literature review by the European Society of Gastrointestinal and Abdominal Radiology (ESGAR) found no conclusive evidence supporting one oral contrast agent over another.⁵²

Participant experience questionnaire data from the METRIC trial (see *Chapter 7*) have clearly demonstrated that participants view the drinking of oral contrast and its subsequent effects, such as pain, bloating and diarrhoea, as the least acceptable part of MRE. However, it is unclear how participant experience is affected by the volume of oral contrast ingested.

The METRIC trial afforded an opportunity to investigate the influence of oral contrast agent and ingested volume on small bowel distension and participant experience during MRE in a subset of participants.

Methods

Participant experience of oral contrast

Recruited participants were asked to complete a questionnaire pertaining to their experience of various symptoms related to the ingestion of oral contrast on the day of their MRE examination. Specifically, they were asked to record feelings of fullness, regurgitation, vomiting, abdominal pains/spasms and diarrhoea and rate their tolerability as 'not at all tolerable', 'somewhat tolerable', 'moderately tolerable' or 'very tolerable' in response to the questions 'While you were drinking, during or just after the scan, did you experience any of the following? If so, how tolerable were they?' (see *Appendix 19, Table 58*).

Participants were asked to complete the questionnaire before they went home and leave it with a member of the local research team. The nature and volume of oral contrast ingested by the participants was recorded where possible by a member of the radiographer staff (see *Chapter 2*).

Grading of bowel distension quality on magnetic resonance enterography

Two post-FRCR radiology trainees who had completed their senior training block in gastrointestinal radiology graded the quality of bowel distension on the MRE data sets using the methods described by Sood *et al.*⁷¹ Specifically, each small bowel and proximal colonic segment (caecum, ascending colon and transverse colon), as defined in *Chapter 2*, was graded by each observer independently as 0 (segment not identified), 1 (segment identified but limited luminal opacification), 2 (lumen clearly opacified with contrast but only limited delineation of the wall), 3 (lumen opacified and distended with clear delineation of the wall throughout the majority of the segment), or 4 (good distension of the lumen with clear visualisation of the bowel wall throughout all the segment). The observers were free to use all sequences of the MRE examination in making their grading decision.

Analysis

The average excellent or good-quality distension was calculated for all participants regardless of oral contrast type. Excellent or good distension frequencies were calculated by participant and by segment based on the type of oral contrast ingested. Oral contrast agents were split into two groups: mannitol and polyethylene glycol. Frequencies and percentages for participant experience of feelings of fullness, regurgitation, vomiting, abdominal pain and diarrhoea were summarised by type of oral contrast and by volume of oral contrast ingested.

Outcomes

The primary outcomes were the segmental distension grades and participant symptom scores according to the type of oral contrast ingested. A secondary outcome was participant symptoms scores according to the volume of oral contrast ingested, which was recorded as < 1000 ml or ≥ 1000 ml.

Results

Complete data (oral contrast type, volume ingested, participant experience scores and MRE distension scores) were available for 66 participants. Of these, 38 ingested 2.5% mannitol, nine ingested 2% mannitol and an additional agent [either 0.2% locust bean gum or 2 scoops of carob gum (OptiFibre®, Nestlé, Vevey, Switzerland)] and 19 ingested polyethylene glycol 3350 (KLEAN-PREP®, Helsinn Healthcare, Lugano, Switzerland). For the purposes of analysis, all mannitol-based preparations were combined.

The overall distension scores for the two observers are given in *Table 25*.

Observer agreement for distension score was in general reasonable, although the terminal ileum was scored by observer 1 as better distended than by observer 2 (graded 3 or 4 in 56% of cases vs. 27% of cases, respectively).

The distension scores averaged across the two observers for each oral preparation type are given in *Table 26*.

Overall, there was no clear difference in the percentage of segments achieving adequate or good distension between mannitol- or polyethylene glycol-based preparations, allowing for the segment numbers across each group.

TABLE 25 Number of segments scored as excellent or good quality for distension according to the two observers (both oral contrast agent types combined)

Segment	Number of segments	Observer, n (%)		Average, n (%)
		1	2	
Overall	66	30 (45)	28 (42)	29 (44)
Duodenum	66	8 (12)	2 (3)	5 (8)
Jejunum	66	13 (20)	18 (27)	16 (23)
Ileum	66	42 (64)	42 (64)	43 (64)
Terminal ileum	66	37 (56)	18 (27)	28 (42)
Caecum	49	18 (37)	19 (39)	19 (38)
Ascending colon	60	36 (60)	33 (55)	35 (58)
Transverse colon	63	33 (52)	19 (30)	26 (41)

TABLE 26 Number of segments achieving good (score of 3) or excellent (score of 4) distension according to oral contrast agent

Segment	Oral contrast agent, n/N (%)	
	Mannitol	Polyethylene glycol
Overall	22/47 (47)	8/19 (42)
Duodenum	4/47 (9)	1/19 (5)
Jejunum	13/47 (28)	3/19 (16)
Ileum	31/47 (66)	12/19 (63)
Terminal ileum	20/47 (43)	8/19 (42)
Caecum	13/36 (36)	6/13 (46)
Ascending colon	24/43 (56)	11/17 (65)
Transverse colon	21/44 (48)	5/19 (26)

The ileum was the best-distended small bowel segment for both preparations, followed by the terminal ileum and jejunum. Specifically, distension of the jejunum, ileum and terminal ileum were rated as excellent or good in 28% (13/47), 66% (31/47) and 43% (20/47) of cases, respectively, for the mannitol group compared with 16% (3/19), 63% (12/19) and 42% (8/19), respectively, for the polyethylene glycol group. There was a suggestion that distension of the right colon (caecum and ascending colon) was a little better with polyethylene glycol than with mannitol, although the small sizes of the groups preclude a definitive conclusion.

Participant grading of symptoms related to oral contrast agent ingested is shown in *Table 27* and summarised in *Appendix 20, Figure 12*.

In general, participants found symptoms of vomiting and regurgitation as being either absent or very tolerable for both contrast agent types. However, symptoms of fullness were graded as moderately tolerable by 27 out of 46 (59%) ingesting mannitol-based contrast and by 12 out of 19 (63%) ingesting polyethylene glycol. Similarly, 26 out of 41 (63%) and 7 out of 15 (47%) graded their feeling of abdominal pain/spasm as moderately tolerable for mannitol-based and polyethylene glycol-based contrast, respectively. The number of participants who rated any symptom as not tolerable was very small in both groups.

Overall, 16 participants ingested < 1 l and 50 participants ingested ≥ 1 l of oral contrast. Symptoms by volume of contrast ingested are shown in *Table 28*.

TABLE 27 Participant symptoms according to oral contrast agent

Symptom	Oral contrast agent, n/N ^a (%)					
	Mannitol (N = 47)			Polyethylene glycol (N = 19)		
	Very tolerable	Moderately tolerable	Not tolerable	Very tolerable	Moderately tolerable	Not tolerable
A feeling of fullness	18/46 (39)	27/46 (59)	1/46 (2)	7/19 (37)	12/19 (63)	0/19 (0)
Regurgitation	35/38 (92)	2/38 (5)	1/38 (3)	12/12 (100)	0/12 (0)	0/12 (0)
Vomiting	36/37 (97)	1/37 (3)	0/37 (0)	12/12 (100)	0/12 (0)	0/12 (0)
Abdominal pain/spasms	14/41 (34)	26/41 (63)	1/41 (3)	5/15 (33)	7/15 (47)	3/15 (20)
Diarrhoea	25/45 (56)	19/45 (42)	1/45 (2)	9/18 (50)	7/18 (39)	2/18 (11)

^a Total number of participants reporting symptom.

Very tolerable = participant did not experience the symptom; moderately tolerable = somewhat/moderately tolerable; not tolerable = not at all tolerable.

TABLE 28 Participant symptoms according to volume of oral contrast agent ingested

Symptom	Volume of oral contrast agent ingested, <i>n/N^a</i> (%)						All participants (<i>N</i> = 66), <i>n/N^a</i> (%)		
	< 1 l (<i>N</i> = 16)			≥ 1 l (<i>N</i> = 50)					
	Very tolerable	Moderately tolerable	Not tolerable	Very tolerable	Moderately tolerable	Not tolerable	Very tolerable	Moderately tolerable	Not tolerable
A feeling of fullness	3/16 (19)	13/16 (81)	0/16 (0)	22/49 (45)	26/49 (53)	1/49 (2)	25/65 (38)	39/65 (60)	1/35 (2)
Regurgitation	12/12 (100)	0/12 (0)	0/12 (0)	35/38 (92)	2/38 (5)	1/38 (3)	47/50 (94)	2/50 (4)	1/50 (2)
Vomiting	12/12 (100)	0/12 (0)	0/12 (0)	36/37 (97)	0/37 (0)	1/37 (3)	48/49 (98)	0/49 (0)	1/49 (2)
Abdominal pain/ spasms	5/13 (38)	6/13 (46)	2/13 (15)	14/43 (33)	27/43 (63)	2/43 (5)	19/56 (34)	33/56 (59)	4/56 (7)
Diarrhoea	11/14 (79)	3/13 (21)	0/13 (0)	23/49 (47)	23/49 (47)	3/49 (6)	34/63 (54)	36/63 (41)	3/63 (5)

^a Total number of participants reporting symptom.

Very tolerable = participant did not experience the symptom; moderately tolerable = somewhat/moderately tolerable; not tolerable = not at all tolerable.

There was no evidence that the tolerability of oral contrast for symptoms of fullness, regurgitation, vomiting or abdominal pain was lower for those participants ingesting ≥ 1 l of oral contrast than for those participants ingesting < 1 l (test of proportions, $p > 0.05$). However, diarrhoea was rated as very tolerable by only 23 out of 49 (47%) participants ingesting ≥ 1 l of oral contrast compared with 11 out of 14 (79%) of those participants ingesting ≥ 1 l ($p = 0.04$).

Chapter 9 The impact of magnetic resonance enterography and ultrasonography on diagnostic confidence and patient management

Introduction

The comparative diagnostic accuracy of MRE and US has a crucial influence on clinical implementation. However, the impact of the tests on actual clinical decision-making is also important. Management of CD patients is complex; symptoms are rarely specific and, alongside patient preference, clinicians balance a range of considerations when defining management, including disease extent, activity and complications. As discussed in *Chapter 1*, the phenotype of CD is very variable, ranging from mild superficial ulceration to complex penetrating and stricturing disease, and cross-sectional imaging is frequently used alongside endoscopy to guide patient management over time.

There is currently very little research comparing the diagnostic and therapeutic impact of MRE and US on clinical decision-making, although there are some data on each modality in isolation. Hafeez *et al.*⁷² tested the impact of MRE in 51 participants with known or suspected CD. Gastroenterologists completed a pro forma detailing their diagnostic confidence for the presence and extent of CD along with the planned therapeutic strategy before and after MRE results were available. MRE significantly increased clinician diagnostic confidence for the presence or absence of SBCD, especially in those patients with a normal MRE result (62–84%) and influenced a change of therapeutic management in 61%. Similarly, in a retrospective service review, MRE and CTE influenced a change of management in 142 out of 311 (45.7%) participants with known or suspected IBD;⁷³ in another retrospective study of 347 participants,⁷⁴ MRE changed the assumed Montreal classification in 21.2% (location) and 24.6% (behaviour). Garcia-Bosch *et al.*⁷⁵ compared the diagnostic impact of MRE and colonoscopy in 100 participants. MRE alone was sufficient to change patient management in 80% of participants compared with in just 34% of participants for colonoscopy alone, and adding MRE to colonoscopy changed patient management in 28% of participants compared with in just 8% of participants after adding colonoscopy to MRE.

Wilkens *et al.*⁷⁶ investigated the impact of US in 115 participants with CD with temporally related US and colonoscopy examinations. Overall, 37 out of 40 participants with incomplete colonoscopy had additional disease detected by US, which changed management in 22 out of 29 participants with available data. In a prospective study of 49 participants with CD, US changed the clinical decisions of two gastroenterologists in 60% and 58% of participants, respectively.⁷⁷

The METRIC trial afforded an opportunity to compare the diagnostic and therapeutic impact of MRE and US across a range of patients and clinicians. This study differs from most studies in the current literature, which usually investigate the impact of adding one test to another, for example US to colonoscopy. Instead, the substudy was designed to compare patient management decisions based on clinical tests supplemented by either MRE or US with a reference standard based on all available test results from individual participants' pathways in the METRIC trial. The primary outcome was the number of times therapeutic decisions (split into prespecified decision groups) based on MRE or US differed from those decisions based on all available tests.

Methods

Gastroenterologists from each of the eight recruitment sites participated in the substudy. An initial study design was piloted at one site (UCLH), requiring regular face-to-face meetings between the gastroenterologist and radiologist, with full access to the complete patient record and MRE and US images. This process was noted to be prohibitively time-consuming and impractical for successful rollout to the remaining seven sites, and therefore the process was centralised by the Clinical Trial Unit, with use of electronic pro forma as detailed below.

Participant selection

A total of 184 out of 284 METRIC trial participants were randomised for inclusion in the substudy. To increase the statistical power, the case mix was enriched to increase the proportion with discrepancies between MRE and US. At the time of substudy design, it was possible to enrich based only on the presence of SBCD and colonic CD, as the segmental extent of disease and disease activity were not available. After enrichment, the percentages of cases with discrepancies between MRE and US compared with the full METRIC study cohort were as follows: for presence of SBCD alone, 21% in the substudy and 13% in the overall METRIC study cohort; for presence of colonic CD alone, 16% in the substudy and 20% in the overall METRIC study cohort; and for presence of both SBCD and colonic CD, 6% in the substudy and 7% in the overall METRIC study cohort.

Trial data as detailed below were collated in batches of 10 participants for individual gastroenterologist review. In each batch of 10 participants, $\approx 50\%$ were new diagnosis participants and $\approx 50\%$ were suspected relapse participants, and $\leq 50\%$ were participants for whom MRE and US disagreed on disease presence.

Data presentation and case report form completion

Fifteen consultant gastroenterologists, representative of clinicians making treatment decisions in NHS clinical practice, participated in the study. The same gastroenterologist reviewed all imaging and test data for an individual participant, so analysis could focus on the impact of different imaging test information provided to the same gastroenterologist.

Data were provided to gastroenterologists via a series of password-protected Microsoft Excel® 2016 version 1905 (Microsoft Corporation, Redmond, WA, USA) spreadsheets. The spreadsheets were populated via CRF data collected as part of the main METRIC study. A summary of the process is shown in *Figure 9*.

The first tab of the spreadsheet listed clinical data including cohort (new diagnosis or suspected relapse), length of symptoms, current symptoms (e.g. weight loss, obstructive symptoms, diarrhoea, pain), family history of CD, smoking status, current CD-related medication, history and nature of previous bowel surgery, and Montreal classification (suspected relapse participants only). The findings of colonoscopy (if performed during the main METRIC study) were also provided, summarised as disease presence and activity for six colonic segments and the terminal ileum. Based on these clinical (and endoscopic) data, the gastroenterologist completed a diagnostic confidence and therapeutic impact (DCTI) CRF based on one we had previously used successfully,⁷² documenting their diagnostic confidence for the presence of SBCD and colonic CD, and their therapeutic management from a range of options provided on the CRF.

The second tab of the spreadsheet contained the findings of MRE, US or a third small bowel test such as CTE or BaFT (if the participant had undergone this as part of the METRIC study). The imaging revelation order was randomised for each participant in a six-block design with mixed randomisation to ensure masking of allocation even for the last participant in each block. Mixed randomisation consisted of the random selection of one row from the six block options and the insertion of this in a random position in each block, giving a total of seven options in each block.

The presence and activity (if applicable) of CD for each small bowel and colonic segment was indicated based on the imaging report CRF completed as part of the main METRIC study. Disease complications

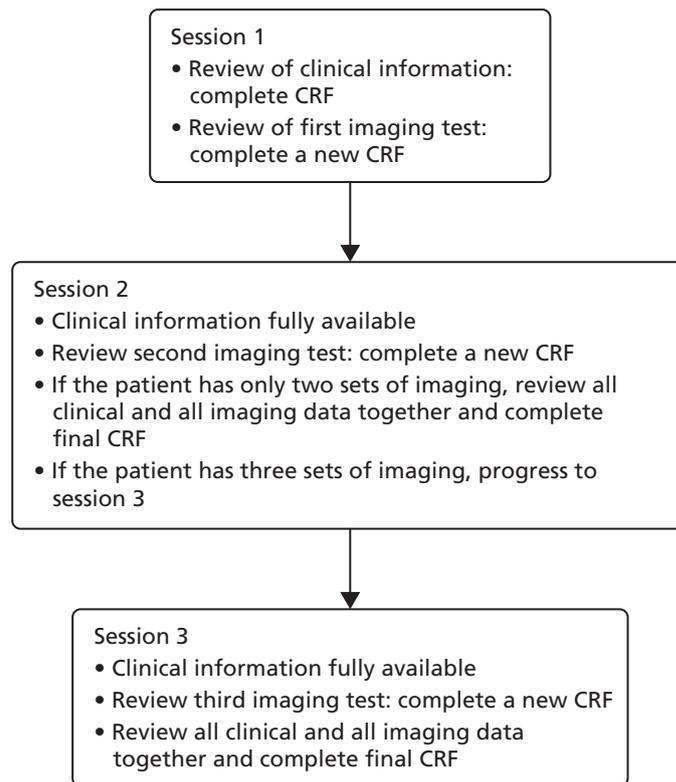


FIGURE 9 Diagnostic impact study flow.

including fistulae, abscess and strictures were also indicated. Once they had reviewed the findings of this imaging test (taking into account all the clinical information provided on the first tab), the gastroenterologist completed a new copy of the DCTI CRF taking into account the additional information (if any) provided by the imaging.

At least 2 weeks later, the results of the second imaging test (e.g. MRE if the first test had been US) were provided in the same format. Again, based on this imaging test together with the original clinical information provided in the first tab, the gastroenterologist completed a new copy of the DCTI CRF. If the participant had undergone a third imaging test, the process was repeated at least 2 weeks later when the results of the final small bowel imaging test were provided via the spreadsheet. The gastroenterologist then completed a final DCTI CRF based on all available information (i.e. all the clinical data together with the findings of MRE, US and any third small bowel imaging test, if applicable).

Analysis

The reference standard was defined as the patient management decision made after the final review of all test information provided, including clinical information, and the findings of MRE, US and any conventional imaging.

Patient management decisions based on clinical information plus MRE or US alone were compared with the reference standard in the same participant and classified as agreeing or disagreeing. A total of 158 participants were included in the final analysis: 165 participants had patient management reported, but seven had missing data for patient management decisions. Management decisions based on conventional imaging data were not analysed as extra conventional imaging was only completed in 20% (31/158) of participants in the substudy.

The 14 possible therapeutic options were collapsed into five prespecified categories (*Table 29*), selected during study design. Agreement or disagreement between treatment decisions based on the clinical information plus an individual imaging modality (MRE or US) and the reference standard were based on these five categories. The primary outcome was prespecified as the percentage difference between MRE and US agreements with the reference standard. Prespecified subgroups were new diagnosis participants, suspected relapse participants, disease-positive participants (any disease in small bowel or colon), disease-negative participants, participants with active disease (any activity in small bowel or colon) and participants with inactive disease. The subgroups were based on the full METRIC trial consensus reference standard.

Statistical analysis was performed in Stata 14.2 using comparison of paired proportions.

Results

A total of 158 participants were included: 74 new diagnosis and 84 suspected relapse participants. Conventional imaging tests were used for only 31 participants (20%) and so therapeutic decisions based on these data were not analysed. Therapeutic decisions based on clinical information and MRE agreed with the final reference standard treatment decision category in 122 out of 158 (77%) decisions and disagreed in 36 out of 158 (23%) decisions. Therapeutic decisions based on clinical information and US agreed with the reference standard treatment decision category in 124 out of 158 (78%) decisions and disagreed in 34 out of 158 (22%) decisions (*Table 30*).

TABLE 29 Grouping of therapeutic options

Categorisation	Category	Description on CRF
No medication change	1	Participant is on no medication for active CD and none will be added
No medication change	1	Maintain current medication for active CD
Medication dose change	2	Reduce dose of current medication for active CD
Medication dose change	2	Increase dose of current medication for active CD
Medication change/addition	3	Participant is not on medication for active CD but will be started on some
Medication change/addition	3	Maintain current medication for active CD and supplement with non-biological medication
Medication change/addition	3	Maintain current medication for active CD and ADD biological medication
Medication change/addition	3	Maintain current medication for active CD and ADD antibiotics
Medication change/addition	3	Change current medication for active CD to similar drug class (e.g. non-biological or biological)
Medication change/addition	3	Change current medication for active CD to different drug class (e.g. from non-biological or biological)
Medication cessation	4	Stop current medication for active CD
Medication cessation	4	Stop current medication for active CD and give antibiotics
Surgery	5	Refer to surgical therapy

TABLE 30 Concordance of therapeutic decisions based on MRE or US compared with the reference standard decision based on all information

Subgroup	Total number of decisions	Clinical information, <i>n</i> (%)				Difference in MRE–US agreement, % (95% CI)	Difference adjusted to main trial prevalence, % (95% CI)
		MRE		US			
		Agreement	Disagreement	Agreement	Disagreement		
All	158	122 (77)	36 (23)	124 (78)	34 (22)	–1 (–10 to 8)	0 (–4 to 5)
New diagnosis participants	74	61 (82)	13 (18)	63 (85)	11 (15)	–3 (–16 to 10)	0 (–6 to 5)
Suspected relapse participants	84	61 (73)	23 (27)	61 (73)	23 (27)	0 (–14 to 14)	1 (–6 to 7)
Disease-positive participants	151	119 (79)	32 (21)	117 (78)	34 (22)	1 (–8 to 11)	3 (–1 to 7)
Disease-negative participants	7	3 (43)	4 (57)	7 (100)	0 (0)	–57 (–108 to –6)	–56 (–73 to –38)
Participants with active disease	141	111 (78)	30 (22)	112 (79)	29 (21)	–1 (–10 to 9)	1 (–3 to 6)
Participants with inactive disease	17	11 (65)	6 (35)	12 (71)	5 (29)	–6 (–42 to 30)	–8 (–23 to 7)

The range of therapeutic decisions on MRE and US compared with the final reference standard treatment decision category is shown in *Tables 31* and *32*. The majority (111/158; 70%) of final reference standard treatment decisions were in category 3 (medication change/addition). Therapeutic decisions based on clinical information and MRE and US agreed with a final reference standard category 3 in 88 out of 111 (79%) and 92 out of 111 (83%) decisions, respectively.

Based on clinical information and MRE, 13 participants would have been referred for surgery. Of these, five (38%) would also have been referred for surgery based on the final reference standard treatment decision, but seven would have undergone a medication change/addition only and one would have no change in treatment. Based on the final reference standard treatment decision, of the 10 participants who would have been referred for surgery, four would have undergone a medication change/addition and one would have had no change in medication if the therapeutic decision had been based on clinical information and MRE only.

Based on clinical information and US only, 10 participants would have been referred for surgery, of whom six (60%) would have been referred for surgery based on the final reference standard treatment decision category, two would have had no medication change, one would have had a medication dose change and one would have undergone a medication change/addition only.

TABLE 31 Descriptive analysis of participant therapeutic decisions based on imaging for MRE against all tests, grouped by therapeutic options category

Grouping of therapeutic decision	Final therapeutic decision based on all tests (reference standard) (n)					Total (n)
	1	2	3	4	5	
1	23	0	13	0	1	37
2	0	6	1	0	0	7
3	7	0	88	0	4	99
4	0	0	2	0	0	2
5	1	0	7	0	5	13
Total	31	6	111	0	10	158

1, no medication change; 2, medication dose change; 3, medication change/addition; 4, medication cessation; 5, surgery.

TABLE 32 Descriptive analysis of participant therapeutic decisions based on imaging for US against all tests, grouped by therapeutic options category

Grouping of therapeutic decision	Final therapeutic decision based on all tests (reference standard) (n)					Total (n)
	1	2	3	4	5	
1	22	1	15	0	2	40
2	0	4	1	0	0	5
3	7	0	92	0	2	101
4	0	0	2	0	0	2
5	2	1	1	0	6	10
Total	31	6	111	0	10	158

1, no medication change; 2, medication dose change; 3, medication change/addition; 4, medication cessation; 5, surgery.

Chapter 10 Cost–utility analysis of magnetic resonance enterography versus ultrasonography for small bowel Crohn’s disease

Introduction

The aim of this substudy was to examine the cost-effectiveness of MRE compared with US for routine imaging of SBCD. Several published studies have evaluated the cost-effectiveness of different imaging strategies for CD,^{78–82} but none of these was UK based and none compared MRE with US.

Methods

Overview of economic evaluation

We undertook a cost–utility analysis to compare the costs and outcomes of MRE versus US based on data collected in the METRIC trial. The outcome measure was quality-adjusted life-years (QALYs), which combine length of life and quality of life, based on National Institute for Health and Care Excellence (NICE) recommendations.⁸³ Cost-effectiveness was expressed as incremental net monetary benefits (INMBs) and the cost analysis took a UK NHS and Personal Social Services perspective.⁸³ Costs were calculated in 2016–17 Great British pounds (GBP) (£) and inflated where necessary.⁸⁴ The time horizon was 1 year, reflecting the fact that national guidance recommends that treatment should be reassessed and follow-up documented at least every 12 months.^{85,86} Modelling beyond this time period was therefore felt to be unnecessary because disease status will be reassessed after 12 months and treatment revised as needed. Therefore, the cost and health impacts associated with the index MRE or US are unlikely to persist beyond 12 months. Given the time horizon, discounting was not applied to costs or outcomes. Separate cost–utility analyses were undertaken for the new diagnosis and suspected relapse cohorts in the METRIC trial.

Diagnostic and treatment stages

The management pathway for patients with CD can be divided into two stages: (1) the diagnostic stage, in which disease identification, location and activity is determined using clinical information, MRE, US and other tests and the initial treatment decision is made, and (2) the treatment stage, during which treatment is delivered, monitored, evaluated and amended as required. The former covers the time from suspected diagnosis or suspected relapse through diagnosis to treatment decision; the latter covers the time period following the treatment decision. From an economic evaluation point of view, the diagnostic stage will be different for MRE and US; the treatment stage may or may not be different, depending on whether or not MRE and US produce different diagnostic and therapeutic outcomes at the diagnostic stage. If there is agreement between MRE and US, the economic analysis can focus on the diagnostic stage only because the treatment stage will be no different between the two options. In this case, the cost-effectiveness of MRI versus US depends on the incremental cost (positive or negative) of each test at the diagnostic stage, and any differential impact of the two tests on health-related quality of life (HRQoL). Alternatively, if there is disagreement between the two tests, the economic analysis should include both the diagnostic and treatment stages because the cost-effectiveness of MRE versus US depends on differences in costs and QALYs at both stages. As noted in *Chapter 3*, there were significant differences in the sensitivity and specificity of MRE and US against the consensus reference standard in determining the extent of SBCD and also in the sensitivity of MRE and US in detecting the presence of SBCD. Based on the data in *Chapter 9*, MRE and US also have different impacts on clinical decision-making. Therapeutic decisions based on clinical information and MRE were different from the final reference standard treatment decision in 23% of cases. In the case of US, the figure was 22%, with differences in therapeutic decisions when comparing MRE and US directly.

Therefore, based on data from the METRIC trial, the economic analysis should incorporate both diagnostic and treatment stages.

Diagnostic versus therapeutic impacts

As noted, data are available from the METRIC trial on the differential impact of MRE and US on both diagnostic and therapeutic outcomes. Costs and QALYs during the treatment stage could be linked to either type of outcome. Focusing on differences in MRE and US on therapeutic outcomes, for example on differences in the types of medication taken and their dosages and referrals for surgery, will be directly related to treatment costs and outcomes following treatment. However, we focus on the diagnostic outcomes as the basis for the economic analysis for two reasons. The first reason is the larger number of participants included in the assessment of diagnostic outcomes (the primary outcome) than therapeutic outcomes (284 vs. 158 participants, respectively). This was a material difference given that separate analyses were undertaken for the new diagnosis and suspected relapse subgroups and that within these subgroups further disaggregation is needed in diagnostic or therapeutic categories (see *Diagnostic outcomes*). The second reason is that management of CD patients is complex: symptoms are rarely specific and clinicians balance a range of considerations when determining treatment, including disease extent, activity and complications, and participant preference. Although there are guidelines⁸⁵ about which treatments to use to induce and maintain disease remission, there is substantial clinical freedom about precise treatments to use for individual participants, with the potential for legitimate variations in treatment between clinicians for the same participant. Hence, variations in therapeutic decisions may reflect legitimate differences in treatment practices given the available clinical information and imaging data, as well as variations in the imaging data. For these reasons we used the diagnostic outcomes from the METRIC trial as the basis for the economic analysis.

Diagnostic outcomes

We created matrices grouping participants into the following five diagnostic categories:

1. multifocal/proximal SBCD, active disease
2. multifocal/proximal SBCD, inactive disease
3. isolated SBCD in the terminal ileal, active disease
4. isolated SBCD in the terminal ileal, inactive disease
5. no SBCD.

These categories were selected based on clinical opinion collected from the METRIC trial investigators and were chosen to mirror the primary and secondary outcomes of the METRIC trial in terms of identifying the presence of CD, as well as the location of CD, and whether or not it was active (see *Chapter 3*).

The per-participant diagnostic category, based separately on each imaging modality (MRE and US), was compared with the final diagnostic category based on the consensus reference standard in a series of five-by-five matrices, with each cell in a matrix containing the proportion of participants in that cell. Separate matrices were constructed for MRE and US for the new diagnosis and suspected relapse cohorts and were populated using data from the main trial (see *Chapter 3*). Weighted mean costs and QALYs for the treatment stage of the management pathway associated with MRE and US were based on the costs and QALYs accrued by participants in different cells/rows of the matrices and the proportion of participants in those cells/rows.

Measuring costs

We consider costs separately for the diagnostic and treatment stages of the management pathway. Potential cost components at the diagnostic stage are:

1. conventional investigations (e.g. ileocolonoscopy, histological and clinical) and conventional imaging tests (e.g. CT and BaFT)
2. treating adverse events associated with conventional investigations
3. MRE

4. treating adverse events associated with MRE
5. additional tests generated by MRE findings
6. US
7. treating adverse events associated with US
8. additional tests generated by US findings
9. MDT meetings to discuss test results and determine treatment where necessary.

We wanted to calculate the incremental cost of MRE versus US. Items 1, 2 and 9 were the same whether the participant had MRE or US and so were not included (in the case of item 9 we assumed that all participants who underwent MRE or US would be discussed at a MDT meeting, even if the test indicates that they have no SBCD). No adverse events related to MRE or US for SBCD were recorded in the METRIC trial (see *Chapter 3*), so items 4 and 7 were not included. No additional tests were recorded as generated by MRE or US for SBCD participants (see *Chapter 3*). Additional small bowel imaging tests triggered by discrepant MRE and US findings were not included as they are unlikely to reflect NHS practice when in general only one or other of MRE or US is used, so items 5 and 8 were also not included. Additional imaging tests during follow-up were included. As a consequence, the costs of the diagnostic stage were reduced to the costs of MRE and US (items 3 and 6).

The unit costs for MRE and US were taken from *2016/17 Reference Costs*.⁸⁷ The base-case unit cost of MRE was £180, based on the national average unit cost of MRI of one area with pre- and post-contrast delivered in an outpatient setting. This was varied in sensitivity analyses to the national lower- and upper-quartile unit costs (£127 and £192, respectively) and to the national average unit cost of MRI of one area without contrast in an outpatient setting (£139). The base-case unit cost of US was £52, based on the national average unit cost of US with a duration of < 20 minutes without contrast delivered in an outpatient setting. This was varied between the lower- and upper-quartile values (£37 and £60, respectively) and to the national average unit cost of US with a duration of ≥ 20 minutes without contrast delivered in an outpatient setting (£65).

At the treatment stage, resource use data were obtained on the following cost components for every participant in the trial, collected via CRFs and participant resource use diaries:

- medications for CD
- surgical procedures for CD
- hospital admissions for CD
- additional imaging/endoscopic investigations for CD
- other outpatient visits for CD
- primary care contacts for CD
 - visits to the general practitioner (GP) at the surgery
 - telephone calls to the GP
 - visits to the practice nurse at the surgery
 - nurse visits at home
 - telephone calls to the practice nurse.

Data were collected separately for months 1–3 and 4–6 after baseline. Unit costs were obtained from published and market sources,^{84,87–92} inflated where appropriate,⁸⁴ and multiplied by resource use collected in the trial. For CD medications, we accounted for the average cost of the induction phase (typically 8 weeks) and the maintenance phase (typically 26 weeks) apportioned over the 6-month follow-up based on national treatment recommendations.^{88–92} Where use of medications was recorded, we assumed that the treatment started at the beginning of the first follow-up period in which it was first recorded.

For participants recruited to the METRIC trial, actual treatment was based on all available clinical, endoscopic and imaging information. In particular, both MRE and US reports were made available to treating clinicians as part of the trial protocol (see *Chapter 2*). As a consequence, it was not possible to

calculate actual mean treatment costs per participant over the 6-month follow-up period for every cell of the five-by-five diagnostic matrices based on MRE and US findings separately. Instead, we calculated the mean cost per participant in each of the five diagnostic categories based on the consensus reference standard; we did this for new diagnosis and suspected relapse participants. We then calculated the mean costs per participant according to the proportion of participants placed in each category by MRE and US separately, as recorded by reporting practitioners during the trial. Again, this was done for both new diagnosis and suspected relapse participants. This assumes that if participants are placed in one of the five diagnostic categories based on MRE or US findings, their costs are the same as if they were placed in the same category by the consensus reference standard. We explored the sensitivity of the results to this assumption by varying follow-up costs during the treatment stage in sensitivity analyses.

Data were collected in the trial for up to 6 months after recruitment. We estimated costs over the 12-month time horizon by extrapolation, multiplying the costs incurred from months 4 to 6 by 3. In sensitivity analyses, as noted, we varied follow-up costs during the treatment stage and ran another sensitivity analysis using a 6-month time horizon only, with no extrapolation.

Measuring quality-adjusted life-years

We considered the impact of MRE and US on QALYs during both the diagnostic and treatment phases. MRE and US were only likely to affect HRQoL if they were associated with adverse events. As noted, there were no reported adverse events (see *Chapter 3*); therefore, we focused on QALYs accrued during the treatment stage.

Generic health status was measured using the EuroQoL-5 Dimensions, five-level version (EQ-5D-5L), which was completed by participants at baseline, 3 months and 6 months. This measure contains five dimensions (mobility, self-care, usual activities, pain and discomfort, anxiety and depression), with five levels in each dimension. Each EQ-5D-5L health state can be converted into a single summary index (utility value) by applying a formula that attaches weights to each of the levels in each dimension based on valuations by general population samples.⁹³ In accordance with the *NICE Position Statement on the EQ-5D-5L*,⁹⁴ we used a mapping function to convert the EQ-5D-5L into EuroQoL-5 Dimensions, three-level version (EQ-5D-3L), utility scores for the UK population to calculate utility values at each time point for every participant.⁹⁵ A utility value of 1 represents full health, a value of 0 is equivalent to death and negative values represent states worse than death. No participant died during the follow-up period.

As noted, actual participant treatment was based on all available clinical, endoscopic and imaging information. As a consequence, it was not possible to directly calculate QALYs per participant over the 6-month follow-up period for every cell of the five-by-five diagnostic matrices based on MRE and US findings separately. We therefore calculated mean utility values at each time point for the five diagnostic categories based on the consensus reference standard. We calculated QALYs over the 6-month period assuming a straight-line relation between the utility values at each time point (calling these '6-month utility QALYs'). We then also calculated QALYs assuming the baseline utility value persisted over the whole 6-month period (calling these 'baseline utility QALYs'). We assumed that 6-month utility QALYs would be higher than baseline utility QALYs as the former includes the effects of treatment. We then calculated the mean 6-month utility and baseline utility QALYs per participant in each of the five diagnostic categories based on the consensus reference standard; we did this separately for new diagnosis and suspected relapse participants.

We then calculated the mean per-participant QALYs for MRE and US based on the proportion of participants who were placed in each diagnostic category, by the reporting practitioners, by each test. We did this separately for new diagnosis and suspected relapse participants. Where the diagnostic category from MRE or US matched the diagnostic category assigned by the consensus reference standard, we allocated the 6-month utility QALY value to that outcome; where they did not match we assigned the

baseline utility QALY value. With the assumption that 6-month utility QALYs were higher than baseline utility QALYs, this approach means that where diagnostic category based on MRE or US disagrees with the consensus reference category, they will incur a QALY penalty. We ran sensitivity analyses using an alternative decision rule for assigning 6-month utility QALYs and baseline utility QALYs to the diagnostic outcomes and where we forced 6-month utility QALYs to be higher than baseline utility QALYs (which was not the case for every diagnostic category using the raw data).

We extended 6-month utility and baseline utility QALY estimates to 12 months using simple extrapolation, assuming that utility scores at 6 months remained constant during the period 7–12 months. As noted, in sensitivity analyses we explored the effect of using a 6-month time horizon only with no extrapolation, and of making alternative assumptions about utility values in the analysis.

Dealing with missing data

There was a substantial number of missing resource use data for primary care contacts and other outpatient attendances (> 50%), which were based on patient resource diaries. For other cost components, which relied on CRF collection, the extent of missing resource use data was low (< 5%). There were also missing utility values based on participant resource diaries at baseline (8%), 3 months (48%) and 6 months (64%). Multiple imputation was used to impute missing values for costs and utility values during the 6-month follow-up period. For costs, we imputed all the cost components included in the cost analysis. The cost variables we imputed were unit costs multiplied by resource use. For utilities, we imputed utility values at baseline, 3 months and 6 months. We included age, sex, study centre and participant cohort (new diagnosis or suspected relapsed cohort) in the imputation models as additional explanatory variables. We used multivariate multiple imputation by chained equations⁹⁶ to impute missing values. This method assumes that the data are missing at random and fills in the missing values up to n number of prespecified times, creating multiple duplicates of the estimated missing value. We created 20 imputed data sets and calculated mean values and SEs for inclusion in our model using Rubin's rules to include both within-imputation variance (accounting for uncertainty if the data were complete) and between-imputation variance (accounting for uncertainty about the missing data).⁹⁷ We repeated the imputation process several times using different random number seeds and the mean values and SEs did not vary appreciably.

Measuring cost-effectiveness

Mean costs, outcomes and net monetary benefits (NMBs) were compared between MRE and US for new diagnosis and suspected relapse participants separately. We calculated differences in mean costs, QALYs and INMBs between groups. NMBs for MRE and US were calculated as the mean QALYs per participant (Q) multiplied by the maximum willingness to pay for a QALY (R) minus the mean cost per participant (C):

$$\text{NMB}_i = Q_i \times R - C_i, \quad (1)$$

in which i is MRE or US. The treatment option with the highest NMB (either most positive or least negative) is preferred on cost-effectiveness grounds. The INMB was calculated as the difference in mean QALYs per participant based on MRE versus US multiplied by the maximum willingness to pay for a QALY minus the difference in mean cost per participant:

$$\text{INMB} = (Q_{\text{MRE}} - Q_{\text{US}}) \times R - (C_{\text{MRE}} - C_{\text{US}}) = \text{NMB}_{\text{MRE}} - \text{NMB}_{\text{US}}. \quad (2)$$

We used the lower bound of the cost-effectiveness threshold range recommended by NICE (£20,000)⁸³ as the maximum willingness to pay for a QALY (R) and varied this in sensitivity analyses. If the INMB was positive or negative then MRE or US, respectively, was preferred on cost-effectiveness grounds.

Sensitivity analyses

A probabilistic sensitivity analysis (PSA) was undertaken,⁸³ varying the following model parameters simultaneously:⁹⁸

- Costs of index MRE and US, modelled using a triangular distribution, with lower and upper boundaries of the distribution based on the lower-quartile and upper-quartile unit costs from NHS Reference Costs.⁸⁷
- Costs during follow-up, with separate costs measured for time periods 1–3 months and 4–6 months for each diagnostic category, and for new diagnosis and suspected relapse participants. These were modelled using gamma distributions, with parameter values α and β based on the mean values and SEs for each category from the study sample.
- Probabilities of each diagnostic category based on the consensus reference standard for MRE and US for new diagnosis and suspected relapse participants. These were modelled using Dirichlet distributions based on the proportion of participants in each category based on the consensus reference standard in the study sample.
- Utilities during follow-up, with separate utilities measured at baseline, 3 months and 6 months for each diagnostic category and for new diagnosis and suspected relapse participants. These were modelled using beta distributions, with parameter values α and β based on the mean values and SEs for each category from the study sample.
- Probability of being in each cell of the five-by-five diagnostic outcome matrices comparing the consensus reference standard separately with MRE and US for new diagnosis and suspected relapse participants. These were modelled using Dirichlet distributions based on the proportion of participants in each cell in the study sample.

After assigning distributions to each parameter, a random value from the corresponding distribution for each parameter was selected simultaneously. The selected values were used in the model to generate estimated values of the following output variables: mean cost, mean QALYs and the NMB associated with MRE and US, and the incremental costs and QALYs gained and INMB of MRE versus US. This was repeated 1000 times and the results for each simulation were noted. The 95% credibility intervals (CrIs) around the base-case values of each output variable were computed based on the 2.5th and 97.5th percentile values from the PSA. For each of the 1000 simulations, the proportion of times MRE or US had the highest NMB was calculated for a range of values for the maximum willingness to pay for 1 QALY, which was varied from £0 to £50,000. These were summarised graphically using cost-effectiveness acceptability curves. Note that the probability that US was cost-effective was 1 minus the probability that MRE was cost-effective.

We also undertook deterministic sensitivity analyses, varying the following model parameters one at a time:

- We calculated NMBs and INMBs using a cost-effectiveness threshold of £15,000 and £30,000 per QALY gained (base-case value of £20,000).
- We calculated costs and QALYs over a 6-month time horizon with no extrapolation beyond this period (base case was 12 months).
- We assigned QALYs to each cell in the five-by-five diagnostic outcome matrices based on whether or not MRE or US produced the same decision as the consensus reference standard only, with respect to having a diagnosis of SBCD or not – where the diagnostic outcome was the same in terms of having SBCD or not, the QALYs based on baseline, 3-month and 6-month utility scores were used; where the diagnostic outcome was not the same in terms of having SBCD or not, QALYs were based on baseline utilities only (and, in the base case, incorrect diagnostic outcomes in terms of location and activity of disease were also ‘penalised’).
- The cost of MRE was set at the lower-quartile (£127) and upper-quartile (£192) unit costs from NHS Reference Costs (base-case value of £180).⁸⁷ It was also set to the national average assuming that no contrast agents were used (£139).
- The cost of US was set at the lower-quartile (£37) and upper-quartile (£60) unit costs from NHS Reference Costs (base-case value of £52).⁸⁷ It was also set to the national average assuming US with a duration of ≥ 20 minutes (£65).

- Costs during follow-up were set at the lower and upper 95% confidence limits (base-case value set at sample mean).
- No imputation of missing cost and utility data during follow-up was undertaken (base case included values imputed using multiple imputation).
- Utility scores in each diagnostic category at 3 and 6 months were forced to be 0.05 and 0.1 higher, respectively, than the baseline value (base case used mean values of the actual utility scores at 3 and 6 months).
- Utility scores in each diagnostic category at 3 months and 6 months were forced to be 0.1 and 0.2 higher, respectively, than the baseline value (base case used mean values of the actual utility scores at 3 and 6 months).

Results

Diagnostic outcomes

Reflecting the primary and secondary outcomes of the study (see *Chapter 3*), there were differences in diagnostic accuracy between MRE and US against the consensus reference standard (*Tables 33–36*). In the new diagnosis cohort there was 75% (100/133) agreement between MRE and the consensus reference standard in terms of the five diagnostic categories (see *Table 33*), compared with 63% (84/133) agreement between US and the consensus reference standard (see *Table 34*). For the suspected relapse cohorts there was 68% (103/151) and 60% (90/151) agreement between MRE and the consensus reference standard and between US and the consensus reference standard, respectively (see *Tables 35 and 36*). The incremental cost-effectiveness is mainly dependent on the difference in the level of agreement between the two tests and the consensus reference standard.

Resource use and costs at the treatment stage

Data in *Appendix 21, Table 59*, show resource use and unit costs for each of the five diagnostic outcomes presented for new diagnosis and suspected relapse participants combined. Accounting for missing data using multiple imputation, mean total costs per participant varied over the two 3-month periods and the five diagnostic categories from £1252 to £7859 for new diagnosis participants and from £1290 to £4471 for the suspected relapse participants (*Table 37*). Note that for some categories the number of observations was small. It is difficult to identify trends in costs between the diagnostic categories and between the new diagnosis and suspected relapse cohorts. Values were broadly similar for the complete-case analysis with no multiple imputation (see *Appendix 22, Table 60*).

TABLE 33 Diagnostic accuracy matrices (new diagnosis cohort): MRE vs. consensus reference standard (133 participants)

Consensus	MRE					Total
	A	B	C	D	E	
A	0.098	0.008	0.045	0.000	0.008	0.158
B	0.008	0.000	0.000	0.000	0.000	0.008
C	0.060	0.000	0.496	0.000	0.075	0.632
D	0.000	0.000	0.030	0.000	0.008	0.038
E	0.000	0.000	0.008	0.000	0.158	0.165

A, multifocal/proximal SBCD, active disease; B, multifocal/proximal SBCD, inactive disease; C, isolated terminal ileal SBCD, active disease; D, isolated terminal ileal SBCD, inactive disease; E, no SBCD.

Shaded cells indicate agreement between MRE and consensus reference standard.

TABLE 34 Diagnostic accuracy matrices (new diagnosis cohort): US vs. consensus reference standard (133 participants)

Consensus	US					Total
	A	B	C	D	E	
A	0.090	0.000	0.045	0.008	0.015	0.158
B	0.000	0.000	0.000	0.008	0.000	0.008
C	0.075	0.008	0.406	0.030	0.113	0.632
D	0.008	0.000	0.023	0.000	0.008	0.038
E	0.008	0.000	0.023	0.000	0.135	0.165

A, multifocal/proximal SBCD, active disease; B, multifocal/proximal SBCD, inactive disease; C, isolated terminal ileal SBCD, active disease; D, isolated terminal ileal SBCD, inactive disease; E, no SBCD.
Shaded cells indicate agreement between US and consensus reference standard.

TABLE 35 Diagnostic accuracy matrices (suspected relapse cohort): MRE vs. consensus reference standard (151 participants)

Consensus	MRE					Total
	A	B	C	D	E	
A	0.113	0.000	0.013	0.000	0.013	0.139
B	0.020	0.007	0.000	0.000	0.000	0.026
C	0.099	0.007	0.391	0.020	0.040	0.556
D	0.013	0.000	0.046	0.007	0.020	0.086
E	0.007	0.007	0.007	0.007	0.166	0.192

A, multifocal/proximal SBCD, active disease; B, multifocal/proximal SBCD, inactive disease; C, isolated terminal ileal SBCD, active disease; D, isolated terminal ileal SBCD, inactive disease; E, no SBCD.
Shaded cells indicate agreement between MRE and consensus reference standard.

TABLE 36 Diagnostic accuracy matrices (suspected relapse cohort): US vs. consensus reference standard (151 participants)

Consensus	US					Total
	A	B	C	D	E	
A	0.079	0.007	0.040	0.000	0.013	0.139
B	0.013	0.007	0.000	0.000	0.007	0.026
C	0.066	0.013	0.364	0.026	0.086	0.556
D	0.020	0.007	0.020	0.013	0.026	0.086
E	0.033	0.000	0.020	0.007	0.132	0.192

A, multifocal/proximal SBCD, active disease; B, multifocal/proximal SBCD, inactive disease; C, isolated terminal ileal SBCD, active disease; D, isolated terminal ileal SBCD, inactive disease; E, no SBCD.
Shaded cells indicate agreement between US and consensus reference standard.

TABLE 37 Mean CD management costs and utilities per participant at the treatment stage

Disease description	Costs (£)				Utilities					
	Months 1–3		Months 4–6		Baseline		3 months		6 months	
	Mean (SE)	<i>n</i>								
New diagnosis cohort (N = 132)										
Multifocal/proximal SBCD										
Active disease	3807 (760)	21	1549 (483)	21	0.79 (0.04)	21	0.77 (0.05)	21	0.79 (0.05)	21
Inactive disease	7889	1	4437	1	0.79	1	0.79	1	0.80	1
Isolated terminal ileal SBCD										
Active disease	2379 (296)	84	1855 (371)	84	0.74 (0.03)	84	0.76 (0.03)	84	0.75 (0.03)	84
Inactive disease	2466 (1605)	5	1252 (1148)	5	0.76 (0.03)	5	0.78 (0.08)	5	0.61 (0.17)	5
No SBCD	1855 (511)	22	1513 (842)	22	0.82 (0.04)	22	0.77 (0.04)	22	0.77 (0.05)	22
Suspected relapse cohort (N = 150)										
Multifocal/proximal SBCD										
Active disease	4361 (715)	21	3379 (529)	21	0.79 (0.03)	21	0.79 (0.05)	21	0.76 (0.06)	21
Inactive disease	3423 (1915)	4	3308 (1112)	4	0.78 (0.05)	4	0.74 (0.08)	4	0.75 (0.09)	4
Isolated terminal ileal SBCD										
Active disease	3542 (377)	84	2114 (297)	84	0.75 (0.02)	84	0.79 (0.03)	84	0.76 (0.03)	84
Inactive disease	1290 (600)	13	1610 (668)	13	0.82 (0.02)	13	0.85 (0.05)	13	0.80 (0.05)	13
No SBCD	4471 (653)	29	2010 (418)	29	0.73 (0.03)	29	0.75 (0.04)	29	0.75 (0.05)	29

Costs are in 2016–17 GBP (£).

Data include values imputed using multiple imputation.

Utility values

Mean utility values at each follow-up point for each of the five diagnostic categories are shown in *Table 37*. There are no obvious trends in baseline utility score by diagnostic category; for some diagnostic categories utilities improved over time and in others they worsened. Mean utility scores at baseline for the new diagnosis cohort varied across a narrow range from 0.74 to 0.82, with the lowest value in those patients with active disease in the terminal ileum and the highest value in those patients with no SBCD. In the suspected relapse cohort, the range at baseline was 0.73 to 0.82, with the lowest value in those patients with no SBCD and the highest value in those patients with inactive disease in the terminal ileum. In the new diagnosis cohort, utilities stayed constant over 6 months for participants with active multifocal/proximal disease, improved in those patients with active disease of the terminal ileum and worsened for the other three groups. In the suspected relapse cohort, utilities improved in participants with disease of the terminal ileum (active and inactive) and in those patients with no SBCD but worsened in those patients with multifocal/proximal disease (active and inactive). Values were broadly similar for the complete-case analysis with no multiple imputation (see *Appendix 22, Table 60*).

Cost-utility analysis

For the new diagnosis cohort, mean costs per participant over 12 months were £7923 based on MRE categorisation and £7690 based on US categorisation, an incremental cost of MRE versus US of £233 (95% CrI -£612 to £1061) (*Table 38*). This figure is slightly higher than the difference in the costs between the two tests (£128). Total QALYs were the same to two decimal places, with the difference in QALYs gained between MRE and US being 0.0004 (95% CrI -0.016 to 0.018). The NMBs for MRE and US were not significantly different, at £7288 (95% CrI £4797 to £9111) and £7513 (95% CrI £4936 to £9392), respectively, giving an INMB of MRE compared with US that was not significantly different from zero [-£225 (95% CrI -£1085 to £713)].

For the suspected relapse cohort, the difference in incremental cost of MRE compared with US was £299 (95% CrI -£267 to £946). The difference in QALYs gained with MRE compared with US was -0.0001 (95% CrI -0.013 to 0.011). The NMBs for MRE and US were not significantly different [£4020 (95% CrI £2426 to £5488) and £4321 (95% CrI £2696 to £5787), respectively], with a non-significant INMB of MRE versus US of -£301 (95% CrI -£993 to £305). Detailed findings were qualitatively similar using a 6-month time horizon, with no extrapolation (see *Appendix 23, Table 61*). Utility values computed using the EQ-5D-5L value set for England are shown in *Appendix 24, Table 62*.

TABLE 38 Incremental cost-effectiveness of MRE vs. US

Cohort	Mean (95% CrI)		
	Total cost (£)	Total QALYs	NMB (£)
New diagnosis			
MRE	7923 (6178 to 10,187)	0.76 (0.72 to 0.79)	7288 (4797 to 9111)
US	7690 (5961 to 9980)	0.76 (0.72 to 0.79)	7513 (4936 to 9392)
Suspected relapse			
MRE	11,317 (9821 to 12,912)	0.77 (0.74 to 0.78)	4020 (2426 to 5488)
US	11,017 (9604 to 12,573)	0.77 (0.74 to 0.79)	4321 (2696 to 5787)
Cohort	Incremental cost (£)	QALYs gained	INMB (£)
New diagnosis			
MRE minus US	233 (-612 to 1061)	0.0004 (-0.016 to 0.018)	-225 (-1085 to 713)
Suspected relapse			
MRE minus US	299 (-267 to 946)	-0.0001 (-0.013 to 0.011)	-301 (-993 to 305)
NMB and INMB calculated at a cost-effectiveness threshold of £20,000 per QALY gained. Costs are in 2016-17 GBP (£).			

Sensitivity analyses

The cost-effectiveness acceptability curves shows that at a maximum willingness to pay for 1 QALY of £20,000, the probability that MRE is cost-effective in the new diagnosis cohort was 0.28 and the probability that US was cost-effective was 0.72 (Figure 10). For the suspected relapse cohort, the probabilities were 0.16 (MRE) and 0.84 (US). The curves shown in Figure 10 are relatively flat over the range of values on the x-axis, indicating that variations in the cost-effectiveness threshold shown do not affect relative cost-effectiveness appreciably. This suggests that the cost-effectiveness of MRE versus US is driven predominantly by differences in costs rather than QALYs (the cost-effectiveness threshold is used in the calculation of the INMBs to monetise the QALY gains). The gaps between the cost-effectiveness acceptability curves for MRE and US are relatively wide because an important driver of the difference in costs between the two options is the cost of each test (base-case values of £180 for MRE, £52 for US); given this difference and the parameters used in the PSA, the proportion of times in the 1000 simulations in the PSA where MRE was costlier than US was relatively small. The interpretation is that although the probability that US is cost-effective is higher than the probability that MRE is cost-effective, there remains no significant differences in cost, outcomes and NMBs overall between the two options.

The findings did not change appreciably when we undertook the deterministic sensitivity analyses. The INMBs were most sensitive to varying the costs during follow-up, and to forcing utilities at 3 and 6 months to be higher than the baseline value (Table 39), but in every case, for both the new diagnosis and suspected relapse cohorts, the incremental MNB was not significantly different from zero.

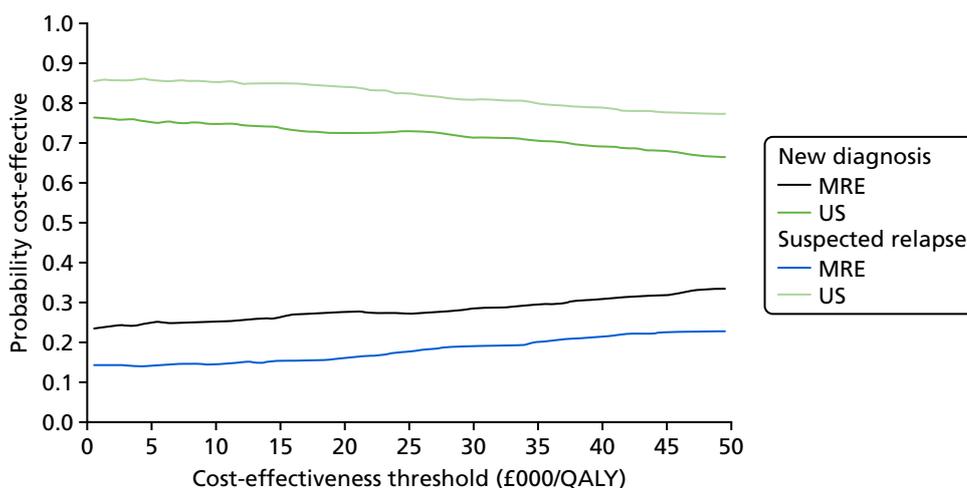


FIGURE 10 Cost-effectiveness acceptability curves showing the probability that MRE and US are cost-effective at different values of the cost-effectiveness threshold.

TABLE 39 Univariate sensitivity analyses: cost-effectiveness of MRE vs. US

Variable	INMB (£), MRE minus US, mean (95% CI)	
	New diagnosis cohort	Suspected relapse cohort
Base case	–225 (–1085 to 713)	–301 (–993 to 305)
Cost-effectiveness threshold of £15,000 per QALY gained	–227 (–1080 to 686)	–301 (–996 to 286)
Cost-effectiveness threshold of £30,000 per QALY gained	–221 (–1169 to 797)	–302 (–1045 to 354)
6-month time horizon	–147 (–610 to 368)	–209 (–537 to 113)
Utilities based on whether or not tests correctly diagnose SB CD	–241 (–1087 to 689)	–287 (–985 to 353)
Cost of MRE		
£127	–172 (–1052 to 671)	–248 (–939 to 414)
£139	–184 (–1105 to 852)	–260 (–895 to 368)
£192	–237 (–1071 to 731)	–313 (–1041 to 349)
Cost of US		
£37	–240 (–1087 to 602)	–316 (–930 to 322)
£60	–217 (–1104 to 687)	–293 (–904 to 335)
£65	–212 (–1236 to 667)	–288 (–948 to 291)
Costs during follow-up set at lower 95% confidence limit	–467 (–1103 to 169)	–388 (–993 to 161)
Costs during follow-up set at upper 95% confidence limit	192 (–1760 to 2225)	–210 (–1057 to 555)
No imputation of missing follow-up data	–244 (–2332 to 1511)	–474 (–2612 to 1091)
3-month and 6-month utilities in each diagnostic category forced to be 0.05 and 0.1 higher, respectively, than baseline value	–52 (–952 to 849)	–198 (–867 to 402)
3-month and 6-month utilities in each diagnostic category forced to be 0.1 and 0.2 higher, respectively, than baseline value	189 (–721 to 1158)	–197 (–1104 to 688)

All analyses are as for the base-case analysis with univariate adjustment of the parameters listed. INMB calculated at a cost-effectiveness threshold of £20,000 per QALY gained excepted where stated. Costs are in 2016–17 GBP (£).

Chapter 11 Discussion

Main trial

Cross-sectional imaging is fundamental for diagnosis and management of CD and is replacing barium fluoroscopic techniques, which have been the bedrock of small bowel imaging for many years. Dissemination of cross-sectional imaging, however, has occurred despite a paucity of supportive data from prospective multicentre studies recruiting consecutive and unselected participants. Emphasis is placed on MRE and US as they avoid ionising radiation. At the time of writing, the METRIC trial is the largest prospective multicentre study directly comparing diagnostic accuracy of MRE and US for the presence, extent and activity of CD in the same participants.

Overall, we found that both tests achieved high sensitivity for SBCD; sensitivity of MRE and US was 97% and 96%, and specificity was 92% and 84%, respectively. Sensitivity compares favourably with the estimates from the various available meta-analysis (see *Table 1*). For example, Dong *et al.*¹⁸ estimated sensitivity and specificity of US at 88% and 97%, respectively, and Liu *et al.*²³ reported corresponding figures of 86% and 93% for MRE. As noted in *Chapter 2*, the primary literature is relatively deficient, with the majority of studies small and single centre^{23,26} and few comparing tests directly in the same participant, despite this being advocated as optimal methodology for diagnostic accuracy studies.²⁹ For example, in a recent meta-analysis, Greenup *et al.*²¹ found that just 1 of 33 included studies compared MRE and US directly in the same participants. Our data suggest that both MRE and US are valid first-line investigations in detecting SBCD and are competitive with ileocolonoscopy for evaluating the terminal ileum. Indeed, against a colonoscopic reference standard (available in 186 participants) we found that MRE and US achieved 97% and 91% sensitivity, respectively, for terminal ileal disease.

However, our primary outcome was not only simple detection of SBCD but also correct segmental localisation (to the duodenum, jejunum, ileum and/or terminal ileum). Management of CD is contingent on the presence and extent of disease. For example, a participant with isolated active terminal ileal disease would be a candidate for either surgical resection or medical therapy,⁹⁹ but the presence of additional jejunal disease would probably tip the balance in favour of medical therapy. As would be expected, sensitivity for SBCD extent, although good (80% and 70% for MRE and US, respectively), was lower than simple detection. Importantly, although the final study cohort was smaller than originally planned (owing to a higher than expected number of dropouts, which were most commonly due to a final diagnosis other than CD), we met the requirements of a power calculation (210 participants with SBCD) and therefore could answer our primary research question with confidence.

Although both MRE and US performed well in the setting of the trial, we found that MRE had significantly higher sensitivity and specificity for small bowel extent and higher sensitivity for disease presence than US. The study was not powered to detect test accuracy differences for individual bowel segments; nevertheless, we found that MRE had significantly greater sensitivity than US for disease in the ileum (84% and 56% sensitivity, respectively, in 38 positive segments). Radiologists graded visualisation of the ileum as good in 231 out of 284 (81%) participants with MRE compared with 201 out of 284 (71%) participants with US, suggesting that some of the missed disease could have been missed owing to poor visualisation of ileal loops, for example secondary to a deep pelvic location or obscuration by bowel gas. In support, the overall number of perceptual errors noted by the consensus reference standard panel was small, and similar between MRE (9%) and US (14%). It should of course be noted that the ability of the panel to classify US perceptual errors was limited to the images taken by the practitioner at the time of the study. Practitioners were encouraged to provide an image or images of every bowel segment, but given the real-time interruption of US, these may not be fully representative of what was seen during the examination. Sensitivity of MRE and US was similar for jejunal disease (and > 60% for both tests), although only 13 participants had jejunal involvement, limiting

conclusions that can be drawn. It is interesting to note that jejunal visualisation was graded as better on US than on MRE [graded as good in 131/284 (46%) on MRE vs. 163/284 (57%) on US]. Jejunal loops are often poorly distended using conventional MRE protocols with oral contrast agent ingestion, but, with the caveat that there was a small number of relevant participants, our data are relatively reassuring that MRE can detect diseased jejunum with reasonable accuracy (achieving 71% sensitivity).

Magnetic resonance enterography was also significantly more sensitive than US in detecting active SBCD (96% vs. 90%, respectively). We required at least one independent activity marker before a participant could be classified as having active disease (either raised CRP or FC levels, ulceration at endoscopy or histologically proven acute inflammation). An advantage of this approach is that it reduces incorporation bias from the imaging tests themselves. However, some participants with very probable active disease on imaging did not necessarily have any of these independent markers, and therefore would have been classified as inactive. However, the number falling into this category was small (< 10 participants). Furthermore, this limitation would affect both MRE and US equally, so would not have an impact on our conclusions. The sensitivity data for MRE and US were once again at the upper limit of those estimated by meta-analysis data (see *Table 1*). Indeed, Puylaert *et al.*²⁶ estimated the sensitivity of US to be just 44%;²⁶ our data suggest that it is much better than that. However, we did not grade the level of activity (e.g. as mild, moderate or severe) and there is evidence that, for example, MRE has higher sensitivity for participants with 'frank' activity than those participants with less active disease.²² It is likely that our cohort was skewed towards participants with more active disease given the need for at least one independent marker of activity. Furthermore, the 133 new diagnosis participants will have been on treatment for a limited amount of time, if at all, meaning that their disease was more likely to be active. The suspected relapse cohort was also required to have either raised CRP/FC levels or endoscopic abnormality to be eligible if they did not exhibit obstructive symptoms. The high sensitivity and specificity of MRE and US for active SBCD probably reflects this disease spectrum. Sites measured FC locally. Although there may be differences in assay between sites, in all sites a FC level of > 100 µg/g was considered abnormal and was therefore used as part of the inclusion criteria. A higher FC level of > 250 µg/g was used to define active disease, because for all sites this was interpreted as active disease. A different level was used between the eligibility criteria and the definition of activity because for the latter we wanted to ensure high specificity, because for some participants they could be the only criteria for activity available to the consensus reference panel. Conversely, we used the same CRP cut-off point for eligibility and activity. We could have also stipulated a higher CRP level for disease activity, but at the time of trial design the research team considered the specificity of FC of > 100 µg/g to be less than a CRP level of > 8 mg/l, and so a higher cut-off point for only the former was chosen.

To 'label' a participant as having active disease, the imaging test must first identify the disease as present. The sensitivity difference between MRE and US for SBCD activity was very similar to that for disease presence (regardless of activity), suggesting that, once disease has been positively identified, both tests are probably similar in classifying disease as active or inactive.

We recruited approximately equally from two cohorts: those newly diagnosed with CD and those with established disease and suspected relapse. Both are clinically distinct and important, and may manifest with differing disease phenotypes; prevalence of stricturing and penetrating disease increases with time.¹⁰⁰ Noting that the METRIC trial was not powered to detect differences between these two cohorts, we found that sensitivity for SBCD was similar (both overall and at the small bowel segmental level), although specificity tended to be lower in those participants with suspected relapse, particularly for US. The reason for this is unclear, although it is possible that previously diseased bowel never completely goes back to 'normal' when the disease is quiescent, and this is manifested as subtle mural changes on US, leading to overdiagnosis. Future research should investigate the prognostic significance of residual mural changes in bowel achieving mucosal healing. Although the number of extraluminal complications was relatively small, MRE has numerically greater sensitivity for abscess and fistulae than US (71% vs. 43% and 86% vs. 52%, respectively). This tends to lend support to the opinion statement from the ESGAR/European Crohn's and Colitis Organisation (ECCO) international guideline group: following negative US, MRE or CTE should be performed in individuals still highly suspected of harbouring an underlying fistula or abscess.⁵

Our primary outcomes centred around detection and staging of CD in the small bowel, but secondary outcomes investigated the diagnostic performance of both tests in the colon. In this regard, overall we found no significant difference between MRE and US for colonic CD detection (64% and 73%, respectively). However, the sensitivity of both tests was considerably lower than for the small bowel. Furthermore, sensitivity for colonic CD extent was poor (22% for MRE and 17% for US). The inferior performance in the colon in part reflects the intrinsic limitations of both tests for detecting colonic CD. For example, MRE protocols were focused on the small bowel and did not include a colonic enema, which improves sensitivity for disease in the colon.⁵⁴ However, it also probably reflects the exquisite sensitivity of colonoscopy and histology for detection of subtle mucosal disease beyond the resolution of imaging. Colonoscopy was available to the consensus reference panel in 186 of the recruited participants and, against this standard of reference alone, MRE and US sensitivity for colonic CD presence dropped to 41% and 49%, respectively. It was not possible to insist on CapE in all participants (owing to limitations in access across sites, participant compliance, costs and risks of retention), but it would have been interesting to compare the sensitivity of the tests for SBCD against CapE (which, like colonoscopy, affords very detailed views of the bowel mucosa).^{101,102} It should be remembered that specificity of CapE is potentially low¹⁰³ and so in itself it is not a 'perfect' standard of reference.

We did find that US had higher sensitivity for colonic CD presence in the new diagnosis cohort than MRE (67% vs. 47%, respectively). Optimised colonic evaluation using MRE requires cleansing and fluid distension⁵⁴ not readily achieved using standard MRE protocols, whereas, in general, US relies on evaluating the manually compressed unprepared colon wall. We can therefore surmise that US has greater sensitivity than MRE for colonic changes associated with a new diagnosis, although it still falls short of colonoscopy. Both tests had higher sensitivity for colonic CD in the suspected relapse group than the new diagnosis cohort. This may in part reflect the greater availability of colonoscopy to the consensus panel in the new diagnosis cohort (with increased detection of subtle disease), but it is also likely that accumulative bowel damage following repeated episodes of colonic inflammation in chronic disease may be better appreciated on cross-sectional imaging owing to more discernible changes in the structure and thickness of the colon wall. In terms of colonic segmental disease detection, MRE had significantly greater sensitivity in the rectum than US (44% vs. 22%, respectively). This is not unexpected given the difficulties in US interrogation of the rectum owing to its deep posterior pelvic location. In support, visualisation was rated as poor in 56% (new diagnosis) and 46% (suspected relapse) of participants using US compared with 30% (new diagnosis) and 18% (suspected relapse), respectively, on MRE.

Our primary analysis considered equivocal findings on MRE and US as positive given the implications for clinical care of an equivocal result: either treatment of the participant or employment of an additional arbiter test. Overall, the number of equivocal scores was low for SBCD presence [3% (9/284) for MRE, 6% (17/284) for US]. In general, when practitioners scored the presence of SBCD as equivocal, it was usually positive at the consensus reference standard, although this was more mixed in the colon, particularly for MRE. We performed a sensitivity analysis in which equivocal findings were treated as negative. As would be expected, sensitivity for SBCD extent fell (from 80% to 75% for MRE and from 70% to 60% for US), but there remained a statistically significant difference between the two tests. Specificity rose for US (from 81% to 90%) but was little changed for MRE (96% to 95%). The data suggest that equivocal findings on US are just that, and that considering them as positive will increase sensitivity at the expense of a fall in specificity. However, for MRE, considering equivocal findings as positive will increase sensitivity with apparently little effect on specificity.

We recruited across eight NHS sites. Although all practitioners were representative of those performing and interpreting US and MRE in clinical practice, as would be expected there was a range of practitioner experience for one modality or the other. There was also a range of monthly case volumes for each modality between the sites. Interpretation of MRE and US performance according to recruitment site is complex, particularly because the study was not powered to detect differences between sites. Most sites report both a higher sensitivity and a higher specificity for MRE than for US, in keeping with the overall trial results. However, at one site (Portsmouth), the sensitivity of US was significantly higher than that of MRE. Portsmouth has the highest volume of US of all the recruitment sites (80 cases per month) and is a known centre of US

excellence. Conversely, it has the lowest monthly volume of MREs of all recruitment sites. This result would be compatible with the hypothesis that increased test (in this case US) sensitivity can be achieved in high-volume sites with specialist expertise. However, some caution must be exercised when interpreting the data. In particular, of the 48 participants recruited by Portsmouth, only two had no SBCD, meaning that the estimate of specificity of 0% (95% CI 0% to 66%) is not reliable owing to the small numbers of participants. Nevertheless, the impact of training and case volume on diagnostic accuracy requires further investigation.

Our protocol had a number of strengths. Ascertaining the true standalone diagnostic accuracy of an individual imaging test is possible only in the absence of external influences to radiological decision-making. Interpretation of MRE or US is likely to be influenced by knowledge of clinical parameters and findings of other imaging tests. We employed a robust blinding protocol so that practitioners interpreting MRE and US were blinded to the findings of all other tests and clinical information, other than the cohort to which the participant had been recruited, and surgical history. There is no single test that could be employed as a standalone reference standard for the presence, extent and activity of CD. 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.*³¹ used MRE without any additional reference standard in many recruits. The potential for incorporation bias is self-evident. In such circumstances, the construct reference standard paradigm (panel diagnosis), which incorporates multiple data sources with clinical outcome, is recommended.³² 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.³² To reduce incorporation bias, participants without supplementary small bowel imaging underwent a third small bowel investigation whenever discrepancy between MRE and US arose. This was available to the panel for 48 out of the 53 (91%) participants for whom MRE and US were discrepant for SBCD presence or location. It is notable that, when our analysis was limited to an ileocolonoscopy reference standard, any differences in accuracy between MRE and US closely mirrored those differences found using the consensus panel reference, which is reassuring in terms of our reference standard methodology. Our statistical analysis was designed to address which is the best test choice for an individual participant. We used modelling methods (rather than a two-by-two approach) to allow interpretation from the participant-specific perspective (i.e. for a participant, what would be the average test sensitivity or specificity and average difference between diagnosis using MRE or US). This uses a bivariate multilevel participant-specific (conditional) random-effects modelling approach. In this model, a random intercept is fitted per participant so that the variability is linked within a participant. The allowance for a different intercept per participant reflects that participants might have a different level of difficulty in diagnosis. In the METRIC trial, all participants had both tests, so this linkage of difficulty of diagnosis within a participant was appropriate. Using a logistic regression allows analysis of subgroups using covariates, for example for new diagnosis and suspected relapse participants. This conserved power compared to using a two-by-two approach, which was important as the study was not powered to detect effects in the separate subgroups.

There are of course limitations to consider in addition to those limitations already discussed. Although we employed a multimodality consensus reference standard, including 6 months of participant follow-up, the robustness or otherwise will in part rely in the actual tests available to the panel, because they all have limitations. Colonoscopy was available in 186 participants and CapE in 18, together with a mix of other tests including CTE, BaFT and magnetic resonance enteroclysis. It was not deemed practical to employ an invasive test, such as CapE or push enteroscopy, in all participants as part of a pragmatic multicentre trial for reasons of accessibility, cost, participant compliance and risks. To insist on these tests (even if available) would probably also have introduced spectrum bias due to differences between participants consenting to them compared with those participants declining. The METRIC trial was conceived as a large pragmatic trial¹⁰⁴ because 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 practitioners reporting NHS small bowel imaging in terms of training and experience. We specifically avoided using a small number of highly experienced practitioners because they would not represent a national workforce. However, as suggested by our individual site data, specialist practitioners working in

high-volume practices may achieve sensitivities in excess of our findings. Imaging was interpreted according to local clinical practice to mirror 'real-world' procedures and enhance generalisability of our results. We acknowledge that blinding practitioners to individual participant history does not mirror usual clinical practice, but this precaution was necessary to isolate diagnostic test accuracy as far as possible. Recruited participants were representative of those participants undergoing MRE and US in daily practice, although we did exclude pregnant women and participants with contraindications to MRI. Our results are therefore highly likely to extrapolate across the NHS and similar health-care settings.

As noted previously, recruited participants to both cohorts had an a priori high likelihood of SBCD and the METRIC trial findings should be interpreted in this context. The utility of the tests in those participants with non-specific abdominal symptoms to confirm or refute a diagnosis of CD is unknown.

In conclusion, we found that both US and MRE achieve excellent diagnostic accuracy for the extent and activity of SBCD in both new diagnosis and suspected relapse participants. Both tests are valid first-line investigations. In a NHS setting, however, the sensitivity and specificity of MRE exceed those of US significantly.

Diagnostic benefit of oral contrast administration: small intestine contrast-enhanced ultrasonography

We found that the addition of oral contrast (SICUS) held no diagnostic advantage over conventional US for the extent of SBCD in our cohort of 64 participants (both tests achieving 71% sensitivity). There was weak evidence that specificity for colonic CD extent may be improved (from 82% to 92%), but this was not statistically significant.

Our data are relatively consistent with the previous literature, which has shown relatively small improvements in US disease detection by the addition of oral contrast. For example, Calabrese *et al.*⁴¹ reported similar accuracy for SBCD for US and SICUS (96% vs. 100%) in 28 participants, although an additional four sites of jejunal disease were found by SICUS. Similarly, in a 102-participant study, Parente *et al.*⁴³ found a small difference in sensitivity between SICUS and conventional US for disease detection (96.1% vs. 91.4%, respectively). However, in a study including 57 participants with CD, Pallotta *et al.*⁴² did report an increase in sensitivity with the use of oral contrast (sensitivity for SBCD against a BaFT reference was 98% for SICUS compared with 87% for conventional US). However, there are more consistent data that SICUS improves the diagnosis of small bowel strictures. Calabrese *et al.*⁴¹ reported a 27% increase in stricture detection by SICUS compared with conventional US (67% vs. 94%, respectively) and Parente *et al.*⁴³ reported that SICUS and conventional US had sensitivities for stricture detection of 89% and 74%, respectively. In our cohort, there were 14 small bowel segments with stenosis causing obstruction according to the consensus reference standard. We found no evidence that SICUS offered any advantage over US; both techniques correctly identified 9 of the 14 segments. There were fewer false-positive diagnoses of obstructing stenosis on SICUS than on US (7 vs. 11, respectively). The diagnosis of stricture/stenosis remains somewhat controversial, but our definition requiring the upstream bowel to be obstructed concurs with recent consensus guidelines.¹⁰⁵

As part of our study design we ensured that both conventional US and SICUS were performed by the same practitioner so as not to introduce reader variability as a cofounder. There is a potential risk of bias with this approach, as the practitioner was unblinded to their first US interpretation when performing SICUS. It may have been preferable to randomise the order of conventional US and SICUS, but this was not practical; the priority was to perform the main METRIC study US and many participants preferred both examinations on the same day. To aid participant compliance, we allowed the MRE oral contrast to be used as the distension agent for SICUS, which happened in around two-thirds of recruited participants. Because of the delay in transferring participants from the MRI scanner to the US suite, it is possible that some contrast had left the small bowel and filled the colon such that distension was potentially inferior to that achieved by a dedicated SICUS study. Although our cohort size of 64 is reasonable in the context of

the existing literature, it is probably too small to detect anything but large differences between US and SICUS. However, we found essentially identical sensitivity and specificity between the two tests for SBCD extent.

In summary, we found no evidence that SICUS increases diagnosed accuracy for SBCD extent compared with conventional US. Stenosis detection was not improved by SICUS but there were few false-positive diagnoses.

Interobserver variation in the interpretation of enteric ultrasonography and magnetic resonance enterography

Interobserver variation in US interpretation was tested across seven practitioners from two recruitment sites. Overall, we found substantial agreement between two reads and the consensus reference standard for detection of SBCD in both new diagnosis and suspected relapse cohorts. Specifically, both reads agreed with the consensus reference standard in 9 out of 11 (82%) new diagnosis participants (PABAK 0.64) and 22 out of 27 (81%) suspected relapse participants (PABAK 0.63). In fact, agreement was even higher when the 'correctness' of the diagnosis against the reference standard was put to one side (reads disagreed with each other for disease presence or absence in only 1 out of 11 and 4 out of 27 new diagnosis and suspected relapse participants, respectively). Agreement for disease extent against the consensus reference was a little lower. Out of 152 small bowel segments, both reads disagreed with the reference standard in five segments and one read disagreed with the reference standard in a further 14. Overall, the two reads agreed on the presence or absence of disease in 138 out of 152 (91%) small bowel segments. Parente *et al.*⁴³ report very good agreement for the location of SBCD between two experienced practitioners ($\kappa = 0.91$). Our data are generally consistent with this finding, although, as noted previously, agreement for disease extent was not as good as for disease presence alone. Clearly, the need to match segmental location adds to the risk of observer disagreement.

Agreement for colonic CD presence was fair in the new diagnosis cohort and moderate in the suspected relapse cohort. As noted in the main METRIC study results (see *Chapter 3*), US (and MRE) (1) have lower sensitivity for colonic CD than for SBCD and (2) tend to have higher sensitivity for colonic CD in the suspected relapse cohort. Overall, our interobserver data reflect this. Again, putting to one side agreement with the final consensus, reads agreed with each other for colonic CD presence or absence in 10 out of 11 new diagnosis participants and 23 out of 27 suspected relapse participants, identical to that of the small bowel. To date, there are few data in the literature on agreement for colonic CD using US, but our data suggest that it is similar to that of the small bowel, although intrinsic sensitivity is lower, particularly in newly diagnosed individuals. In reality, the vast majority of newly diagnosed patients will undergo colonoscopy, which clearly remains the best test to detect colonic involvement.

Our data should be viewed with some caution. Owing to difficulties in accommodating the interobserver study at recruitment sites, only two sites and seven practitioners participated, of whom two performed the vast majority of reads. Thus, our data may not be generalisable across a larger number of practitioners and institutions.

The design of our MRE interobserver study was different from that of the US study in part because MRE data sets can be collated centrally and interpreted retrospectively, unlike US, which requires real-time hands-on interpretation. Each MRE data set was read by three observers rather than two and we used a total of 73 data sets with a different mix of new diagnosis and suspected relapse participants. It is therefore difficult to directly compare across the two studies. However, our analyses of the MRE interobserver study averaged the data across pairs of reads (similar to the paired reads of the US study), so broad observations can be made. Overall, we found no evidence that interobserver agreement was better with MRE than US. Agreement for SBCD presence was fair for the new diagnosis cohort and moderate for the suspected relapse cohort. Like US, MRE agreement fell when considering disease extent rather than just presence. One potentially

interesting observation is that, unlike US, MRE agreement tended to be better in disease-negative rather than disease-positive participants. In the main METRIC study, specificity of US was significantly lower than of MRE for SBCD extent. The data therefore suggest that the risk of false-positive diagnosis might be higher with US than with MRE. The level of reader agreement we report is similar to that of Jensen *et al.*,⁴⁷ who also reported moderate agreement for the presence or absence of SBCD across four readers interpreting 50 MRE data sets ($\kappa = 0.48$, 54% agreement). All our readers were reporting MRE in clinical practice and most took part in the main METRIC study and are therefore representative of those individuals interpreting MRE in the NHS. There is, however, good evidence of a learning curve in MRE interpretation,^{106,107} so experience is important, as for any imaging technique.

In conclusion, we found interobserver agreement broadly similar between US and MRE, ranging from fair to substantial for SBCD presence. In general, agreement for both modalities was a little better in the small bowel than in the colon.

Influence of sequence selection on magnetic resonance enterography diagnostic accuracy

Consensus guidelines currently recommend multisequence MRE protocols including T1-weighted sequences acquired after administration of intravenous gadolinium contrast.⁵² The evidence in support of such complex protocols, however, is relatively sparse and most guidelines are mainly based on expert opinion.⁵² Such protocols typically take 30–40 minutes to acquire, which is relatively time-consuming for a limited resource such as MRI. In addition, in recent years there have been increasing concerns about routine use of intravenous gadolinium injections with potentially detrimental long-term retention in the brain.⁵⁵ To date, most research into MRI sequence protocol optimisation has investigated replacing post-intravenous contrast sequences with diffusion-weighted imaging.⁵⁸

Using a locked sequential-read study design, we investigated the impact of three sequence combinations on reader accuracy for SBCD extent (T2-weighted and steady-state free precession gradient echo images alone, T2-weighted and steady-state free precession gradient echo images with diffusion-weighted images and, finally, T2-weighted and steady-state free precession gradient echo images, diffusion-weighted images and contrast-enhanced images combined). In keeping with several studies in the literature,^{58,59} we found that there was no diagnostic benefit to adding contrast-enhanced sequences to a combination of T2-weighted and steady-state free precession gradient echo images and diffusion-weighted images. However, we in fact found that adding diffusion-weighted imaging offered no advantage over simple T2-weighted and steady-state free precession gradient echo images alone. Perhaps more surprisingly, sensitivity was significantly lower when readers used the full combination of all sequences (including diffusion-weighted and contrast-enhanced sequences) than when they just interpreted T2-weighted and steady-state free precession gradient echo images on their own. To our knowledge, such an observation has not been made in the literature. In a study of 59 participants, Maccioni *et al.*¹⁰⁸ directly compared T2-weighted images with post-contrast T1-weighted images and found them to be almost identical in terms of disease detection (95% vs. 93%, respectively), although T2-weighted images had higher sensitivity for stenosis. However, the authors only reported the effect of combining sequences on detection of extra-intestinal complications and not disease identification. Similarly, Low *et al.*¹⁰⁹ compared post-contrast T1-weighted sequences alone with T2-weighted images alone in 28 participants and reported that post-contrast T1-weighted images had a higher per-participant sensitivity (100% vs. 60%). Our finding that a combination of all sequences decreases sensitivity (and specificity) for SBCD extent is therefore of great interest. The data suggest that contrast-enhanced sequences are in some way misleading readers such that in some participants they change a correct diagnosis to an incorrect one. When sensitivity drops, we can speculate that, although readers may suspect disease on T2-weighted images, for example, a lack of avid contrast enhancement may falsely reassure them that there is no disease. Conversely, apparently increased enhancement in the bowel wall could suggest to radiologists that there is disease present when there is not (pseudo enhancement of normal bowel is well described, for example in areas of collapse and in the

jejunum).³⁵ There was weak evidence that diffusion-weighted imaging could have a similar effect, although this was not statistically significantly inferior to T2-weighted and steady-state free precession gradient echo images. It is possible that the combination of diffusion-weighted imaging and contrast-enhanced T1-weighted imaging in combination are synergetic in misleading radiologists. For example, both can be abnormal in lymphoid hyperplasia, which is often a normal finding.¹¹⁰ We did not test the combination of T2-weighted and steady-state free precession gradient echo images and T1-weighted images without diffusion to fully address this possibility.

In a second part of the study, radiologists interpreting the main METRIC study MRE noted if and how diffusion-weighted and contrast-enhanced images changed their diagnosis and/or diagnostic confidence. Overall, radiologists changed their diagnosis in a minority of participants (6% for diffusion-weighted images and 5% for contrast-enhanced sequences), although diagnostic confidence was increased in 64% (diffusion-weighted images) and 70% (contrast-enhanced sequences) of cases. These data support the concept that diffusion-weighted and contrast-enhanced sequences are not viewed as essential by radiologists in the majority of MRE examinations.

In summary, addition of diffusion-weighted images and contrast-enhanced sequences do not improve diagnostic accuracy for SBCD extent compared with a combination of T2-weighted and steady-state free precession gradient echo images alone. Indeed, multisequence protocols may be detrimental to diagnostic accuracy. Radiologists state that these sequences change their diagnosis in a minority of participants.

Magnetic resonance enterography and ultrasonography to diagnose Crohn's disease: participant acceptability, perceived burden and preferences

To date, a detailed evaluation of participant experience during MRE and US is lacking in the literature. Using a multifaceted questionnaire completed by 159 participants recruited to the METRIC trial, we found that in general both MRE and US are reasonably well tolerated by participants and attract relatively low burden scores and the majority of participants are willing to undergo either test again. However, on nearly every measure, MRE was rated less favourably than US. This is perhaps not unexpected given the attributes of the two tests. US requires little participant preparation and is performed without the need to enter a scanner, often at rest, while the participant breathes normally. Conversely, MRE requires prior participant preparation with an oral contrast load, after which the participant must lie flat in a relatively enclosed MRI scanner and undertake frequent breath holds as the images are acquired. Participants rated the drink before MRE as the worst part of the examination, together with symptoms produced such as abdominal pain (see *Chapter 8*). A large number of participants found US 'fine', with no worst part, although, for those participants listing a worst part, abdominal compression was the most frequently stated. In terms of recovery, 18% of participants took > 1 day after MRE, compared with just 2% after US. Using a similar burden questionnaire to the METRIC study, Evans *et al.*⁶⁶ reported that a 1-hour whole-body MRI for cancer staging actually attracted lower burden scores than we found for MRE (a score of 2.21 vs. 2.72). Although comparison should be made with caution given the different participant demographics, the data do again suggest that the oral contrast drink is more of a burden to participants than the MRI scan itself.

However, although MRE was less well tolerated than US and preferred by a minority of participants, 88% still rated it as very or fairly acceptable and 91% were willing to undergo it again. Another important finding was that participants value many other test attributes over scan burden. Notably, diagnostic accuracy was rated as the most important, followed by waiting time to diagnosis/treatment and number of tests needed prior to final diagnosis. Being able to receive the result immediately after the test was also rated important. In general, negative physical test attributes, such as requirement to drink fluid, test discomfort and fasting, were rated as less important, and generally between 'a little bit important' and 'moderately important'. We did not specifically indicate to participants the relative diagnostic accuracy of MRE and US (because the main outcome of the METRIC trial was unknown at the time) but presumed

that most considered them largely equivalent. Our findings mirror those of other diagnostic tests, such as CT colonography, where, again, participants rate diagnostic accuracy as more important than test discomfort.¹¹¹

This study does have limitations. Our response rate was just under 50%, which is perhaps understandable given the length and complexity of the questionnaire, and consistent with that reported in other similar questionnaire studies.⁶⁶ Non-responders tended to be younger than responders. By definition, participants had 'opted in' to a trial of MRE and US, so may not be representative of the general participant population; it would be interesting to sample the view of participants who declined to take part in the METRIC trial to ascertain if their views are similar to those participants who took part.

In summary, both MRE and US are generally well tolerated by participants. However, participant burden and recovery are significantly inferior for MRE than for US. Although a majority of participants would opt to undergo US rather than MRE, participants rate other scan attributes, notably diagnostic accuracy, as more important than test discomfort.

Influence of oral contrast agent and ingested volume on small bowel distension and participant experience during magnetic resonance enterography

Small bowel distension with an oral contrast agent improves accuracy of MRE compared with a non-prepared examination.⁶⁷ Many different oral contrast agents are described in the literature,^{68,69,112,113} with no single preferred or recommended agent,⁵² although there is evidence that distension quality is related to the osmolarity of the ingested agent.⁷⁰

In the METRIC study, recruitment sites used either mannitol (alone or with various additives) or polyethylene glycol. Using a previously published distension quality grading system,⁷¹ we found no strong evidence that distension quality was any different between the two agents. Excellent- or good-quality distension was achieved in 65% and 43% of ileal and terminal ileal segments, respectively, with mannitol-based agents, compared with 63% and 39%, respectively, with polyethylene glycol. As expected, jejunal distension was inferior to that of the ileum using enterography protocols (where oral contrast agents are ingested as opposed to infused via nasojejunal tubes); excellent or good distension was achieved in 27% and 16% of jejunal segments with mannitol solutions and polyethylene glycol, respectively. There was moderate distension of the right colon with both agents. We did not assess distension in the left colon, but previous work has shown that detection of colonic CD is improved after colonic distension with a water enema compared with unprepared colon.^{114,115} As noted in the main METRIC study results, diagnostic performance of MRE was inferior in the colon compared with in the small bowel.

A second consideration as to the choice of optimal oral contrast agent prior to MRE is the participant symptom profile. *Chapter 7* shows that participants rate ingesting oral contrast agents prior to MRE as the worst part of the examination, with 18% taking > 1 day to recover. We collated feelings of fullness, regurgitation, vomiting, abdominal pain and diarrhoea in a subset of participants at the time of their MRE examinations. Overall, the vast majority of participants found these symptoms either very tolerable or moderately tolerable, with a minority describing them as 'not tolerable'. However, there were differences in tolerability between the various symptoms. Regurgitation and vomiting were rarely problematic, but feeling of fullness, abdominal pain and to a lesser extent diarrhoea were less well tolerated. We found no evidence that either mannitol-based agents or polyethylene glycol was better tolerated than the other. For example, 59% of participants rated abdominal pain as moderately tolerable after mannitol-based oral contrast compared with 63% after polyethylene glycol. Although proportionally more participants reported abdominal pain after ingesting mannitol (63% rating it as moderately tolerable compared with 47% of participants who rated polyethylene glycol as moderately tolerable), this was not a statistically significant difference. There is relatively little in the literature comparing the symptom profiles between oral contrast

agents in participant groups. Most work has been performed in small volunteer studies. For example, in a small study of 12 volunteers, Ajaj *et al.*¹¹³ suggested that sorbitol produces less symptoms than mannitol.

We also found little evidence that the volume of oral contrast agent ingested influenced tolerability, and 50 out of 66 participants managed to ingest > 1 l. The only symptom that was statistically more severe in the group ingesting > 1 l of oral contrast agent was diarrhoea, rated as very tolerable by 79% of those participants ingesting < 1 l compared with 47% of those participants ingesting > 1 l. This is perhaps expected given that most of the oral contrast agent is not absorbed by the gut such that larger ingested volumes would be expected to result in increased diarrhoea. Again, there is relatively little in the literature documenting participants' symptoms according to volume of contrast ingested, and most data are derived from volunteer studies. In a study of 10 volunteers, Ajaj *et al.*¹¹⁶ reported that although a volume of 1 l was preferred over 1.5 l, there was no difference in the side-effect profile of the different volumes. Kuehler *et al.*,¹¹⁷ in a study of six volunteers, reported a greater symptom burden (notably abdominal pain) with ingested volumes of 1800 ml than with small volumes. In clinical practice, however, ingested volumes prior to MRE rarely exceed 1500 ml.

Our study has limitations. Because METRIC trial sites implemented their local MRE protocol as part of the main study, only two types of oral contrast agents were available for comparison and, even then, numbers in the polyethylene glycol group were small such that we would have been able to detect only large differences in distension quality and side-effect profiles. Similarly, the majority of participants were able to ingest > 1 l of contrast, with only a minority ingesting less. All sites provided > 1.5 l of oral contrast and encouraged participants to ingest as much as they could. Ultimately, however, the volume ingested differs between participants according to their individual tolerance. We cannot be sure if participants drinking < 1 l had a lower tolerance of symptoms than those participants drinking > 1 l, or if in fact the oral contrast volume does not affect symptoms other than diarrhoea.

In summary, as oral contrast agents prior to MRE, mannitol-based solutions and polyethylene glycol give comparable distension quality and produced similar side-effect profiles. Participants ingesting > 1 l of oral contrast agent tend to experience only more diarrhoea than those participants ingesting less.

Diagnostic impact

Overall, we found no major difference between MRE and US on therapeutic decision-making. Both tests agreed with a final therapeutic decision based on all tests in > 75% of cases. To date and to the best of our knowledge, there has been no previous study comparing the impact of MRE and US on participant management, but in general our data concur with the previous literature, which shows that both MRE and US have positive effects on participant management. As noted in *Chapter 9*, previous research has shown that both MRE and US change participant management in a large proportion of participants.^{72–74,76,77}

We deliberately enriched our cohort with participants in whom MRE and US were discrepant for SBCD extent, but, even by 'stressing' the clinicians with discordant imaging results, we found no clear differences between the impact of the two tests. It should be acknowledged that our trial design was 'artificial' in that clinicians were provided with clinical and endoscopic data in the form of an electronic summary. They were also unable to review the MRE and US images, although this better reflects clinical practice where radiological reports form the mainstay of outpatient management. The management of CD, however, is complex, with a range of therapeutic options, and clinician decision-making is based not only on disease presence and activity but also on severity of participant symptoms, medical history and participant choice. It is possible that by employing this methodology we were unable to detect more-subtle differences in patient management than would be apparent had we for example tested the impact of the imaging live in a MDT or outpatient clinic. Our initial intention was to make use of a 'mini MDT', involving a gastroenterologist and radiologist testing the impact of imaging live and face to face, with full access to the patient electronic record, endoscopy and

imaging. This was attempted at one recruitment site but proved to be very time-consuming and impracticable to implement across all sites.

There is a perception that clinicians find MRE images more intuitive to interpret given their familiarity with cross-sectional imaging. Of course, this is not necessarily the case for gastroenterologists trained in US. In any event, our data suggest that, at least in the context of this study design, clinicians believed or 'trusted' US and MRE results to the same extent and did not favour one over the other.

This study does have limitations. The restrictions of using an electronic format are noted in the paragraphs above. For ease of data analysis, we grouped treatment decisions into five groups. The majority of outcomes were in decision group 3 (medication change or addition), which included several different therapeutic options. Although our a priori analysis plan was designed to highlight major differences in treatment plans based on the two tests, a more granular analysis of category 3 decisions may have been informative.

In conclusion, we found no significant difference in the impact of MRE and US on clinician therapeutic decision-making based on our prespecified grouping of treatment plans.

Cost-utility analysis of magnetic resonance enterography versus ultrasonography for small bowel Crohn's disease

Our economic analysis using the METRIC trial data to evaluate the cost-effectiveness of MRE compared with US for imaging the small bowel showed that both options had similar costs and QALYs: for both new diagnosis and suspected relapse participants, the incremental costs of MRE versus US per participant were positive but quantitatively small, with wide CIs, and therefore not significantly different from zero; the QALY differences between MRE and US were quantitatively negligible and also not significantly different from zero. Together, these translated into small incremental negative NMBs for MRE versus US that were not significantly different from zero. Although the negative INMBs for MRE versus US indicate a trend towards US over MRE, we conclude, given the small non-significant differences in costs and QALYs between the two options, that it is not possible to recommend US or MRE on cost-effectiveness grounds. There were a range of assumptions underpinning our analyses, and the main findings were also borne out in sensitivity analyses. The findings indicate that there is no reason to prefer MRE or US on the basis of differences in HRQoL or on economic grounds; other factors should be taken into account when deciding which option to use for imaging in SBCD, for example diagnostic accuracy, availability of and access to different imaging modalities, and patient preferences.

There are two main limitations of the economics analysis, which should be considered when viewing the findings. The first is that treatment of participants recruited to the METRIC trial was based on all available clinical, endoscopic and imaging information. In particular, for ethics reasons, the findings of both MRE and US reports were made available to treating clinicians as part of the trial protocol. This led to difficulties in determining the costs and benefits of MRE and US when they were discrepant as we did not collate data on 'clean' therapeutic decisions based on MRE or US alone. The therapeutic impact of MRE and US were modelled (see *Chapter 9*) but, given the smaller number of participants in this study and the variability in treatment practices between individual gastroenterologists, we felt it to be more appropriate to use the diagnostic outcomes as the basis for the economic analysis. We explored the impact of assumptions in our sensitivity analyses and the findings did not change appreciably. The second main limitation is that data were collected at the treatment stage for 6 months only; the time horizon for the economic analysis was 12 months and simple extrapolation methods were used. We evaluated the sensitivity of the findings to alternative assumptions about costs and QALYs at the treatment stage, and ran analyses using a 6-month time horizon only, and the overall conclusions did not change.

Overall conclusions

- When tested in a prospective multicentre trial setting, both MRE and US have good accuracy for detecting the presence and extent of SBCD in newly diagnosed participants and in those participants with established disease and suspected relapse. However, in this setting, sensitivity and specificity of MRE for SBCD extent, presence and activity is significantly greater than for US.
- Modelled diagnostic impact is similar between MRE and US.
- There are no significant differences between MRE and US in terms of costs and QALYs and overall cost-effectiveness.
- The addition of oral contrast prior to US (SICUS) does not improve the accuracy for SBCD extent or stenosis detection in comparison with conventional US.
- Both MRE and US are deemed acceptable by the majority of participants, although US induces less participant burden and is generally preferred over MRE. However, participants rank diagnostic accuracy as more important than test burden, so choice of test should involve a dialogue between clinicians and participants, considering the full range of test attributes.
- There was variable agreement between radiologists in interpretation of both MRE and US, particularly for disease extent.
- We found no evidence that one oral contrast agent is better than another in achieving good bowel distension during MRE or in reducing patient symptom load.
- Addition of diffusion-weighted imaging holds no diagnostic advantage over simple MRE protocols based only on T2-weighted and steady-state free precession gradient echo images. Addition of diffusion-weighted imaging and post-contrast images may actually be detrimental to sensitivity for SBCD extent, which should be considered when designing MRE protocols.

Implications for practice

In a NHS setting, MRE has significantly higher sensitivity and specificity than US for the presence and extent of SBCD, although both tests are valid first-line investigations. US performs better than MRE for detection of colonic CD in those participants newly diagnosed but both tests are less accurate than colonoscopy in the large bowel.

Both MRE and US are deemed acceptable by the majority of participants, although US induces less participant burden and is generally preferred. However, participants rank diagnostic accuracy as more important than test burden. The choice of small bowel imaging should involve dialogue between participants, clinicians and radiologists. Both tests show variable agreement in interpretation between practitioners, particularly for disease extent, which should be considered part of implementation. We found no evidence that one oral contrast agent is better than another in achieving good bowel distension during MRE or reducing participant symptom load. We also found no evidence that SICUS increases diagnostic accuracy for SBCD extent compared with conventional US. Addition of diffusion-weighted imaging does not improve the accuracy of MRE protocols based only on T2-weighted and steady-state free precession gradient echo images, and post-contrast images may be detrimental to sensitivity, which requires further investigation. Modelled diagnostic impact on clinician therapeutic strategy was similar between MRE and US. There is no reason to prefer MRE or US on economic grounds.

Recommendations for future research

- The role of US in targeted follow-up of CD patients with an established disease phenotype as an alternative to MRE. The METRIC trial blinded practitioners to clinical history and prior disease phenotype. US is well tolerated by patients and in those patients with an established and stable disease phenotype may be an effective tool in monitoring patient status.

- The utility of MRE and US in treatment response assessment and prediction of response. Treatment decisions in CD are dependent on the presence of active disease. Both imaging tests have potential as powerful non-invasive tools to monitor the efficacy or otherwise of medical therapy.
- The most clinically effective and cost-effective cross-sectional imaging investigation in patients with non-specific abdominal symptoms to confirm or refute diagnosis of CD. Many patients present with non-specific gastrointestinal symptoms and, although pre-test probability of CD is low, still undergo investigation, often with cross-sectional imaging. Diagnostic pathways are not yet defined.
- The impact of dedicated training programs and clinical case volumes on practitioner accuracy. The METRIC trial found moderate interobserver agreement for both tests and some variability in test performance between recruitment sites. The training requirements and scan volumes required to achieve and maintain practitioner competency for both tests is relatively unknown.

Acknowledgements

The project was supported by researchers at the NIHR UCLH Biomedical Research Centre and NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham.

The investigators are grateful to Biotronics3D (London, UK), which hosted the MRE and US images on its online server.

Metric trial investigators

Clinical Trial Unit

Jade Dyer, Prinitha Veeramalla, Sue Tebbs and Steve Hibbert.

Radiologists/sonographers

Richard Beable, Hannah Lambie, Rachel Hyland, Roger Lapham, Helen Bungay, Maggie Betts, Niall Power, Rajapandian Ilangovan, Uday Patel, Evgenia Mainta, Phillip Lung, Francois Porte, James Pilcher, Jonny Vlahos, Rebecca Greenhalgh, Anita Wale, Harbir Sidhu, Shonit Punwani, Hameed Rafiee and Gillian Duncan.

Consultant clinicians/histopathologists

Fergus Thursby-Pelham, Richard Ellis, Anthony O'Connor, Nigel Scott, Ian Johnston, Mani Naghibi, Morgan Moorghen, Adriana Martinez, Christopher Alexakis, Farooq Rahman, Simona de Caro, Shameer Metha, Rosa Vega and Craig Mowat.

Research nurses/radiographers/local co-ordinators

Nicola Gibbons, Claire Ward, Doris Quartey, Deborah Scrimshaw, Simona Fourie, Anisur Rahman, Teresita Beeston, Wivijin Piga, Joey Clemente, Roman Jastrub, Mairead Tennent and Caron Innes.

Oversight committees

The TSC and independent Data Monitoring Committee met at least annually and included those individuals mentioned below.

Trial Steering Committee

Vicky Goh (chairperson), Andrea Marshall (statistician), Ilan Jacobs (patient representative) and James Lindsay (subject expert).

Independent Data Monitoring Committee

Tim Orchard (chairperson), Doh-Mu Koh (subject expert) and Chris Rogers (statistician).

Contributions of authors

Stuart A Taylor (Professor of Medical Imaging) was the chief investigator, conceived the study design, contributed to the protocol writing and study management as a member of the TMG, interpreted trial imaging and performed the initial drafting and final editing of the report.

Sue Mallett (Senior Statistician) helped conceive the study design, contributed to the protocol writing and study management as member of the TMG, wrote the statistical analysis plan, led the statistical analysis and contributed to the writing and editing of the final report.

Gauraang Bhatnagar (Consultant Radiologist) contributed to the study design, contributed to management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Stephen Morris (Senior Health Economist) contributed to the study design and protocol writing, contributed to the health economic analysis plan, provided health economic analysis support and reviewed the final report.

Laura Quinn (Statistician) contributed to the statistical analysis plan, statistical analysis and writing and editing of the final report.

Florian Tomini (Health Economist) contributed to the health economic analysis plan, provided health economic analysis support and reviewed the final report.

Anne Miles (Health Psychologist) contributed to the study design, provided clinical advice, analysed the health psychology data and contributed to the final report.

Rachel Baldwin-Cleland (Lead Radiographer) contributed to data acquisition, study management as a member of the TMG and the final report.

Stuart Bloom (Consultant Gastroenterologist) contributed to the study design, contributed to study management as a member of the TMG, provided clinical advice, contributed to data acquisition and contributed to the final report.

Arun Gupta (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Peter John Hamlin (Consultant Gastroenterologist) contributed to the study design, provided clinical advice, contributed to data acquisition and contributed to the final report.

Ailsa L Hart (Consultant Gastroenterologist) contributed to the study design, provided clinical advice, contributed to data acquisition and contributed to the final report.

Antony Higginson (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Ilan Jacobs (Independent Patient Representative) contributed to the study design, study management as a member of the TMG and the final report.

Sara McCartney (Consultant Gastroenterologist) contributed to the study design, provided clinical advice, contributed to data acquisition and contributed to the final report.

Charles D Murray (Consultant Gastroenterologist) contributed to the study design, provided clinical advice, contributed to data acquisition and contributed to the final report.

Andrew AO Plumb (Consultant Radiologist) contributed to data analysis and interpretation, interpreted trial imaging, collected trial data and contributed to the final report.

Richard C Pollok (Consultant Gastroenterologist) contributed to the study design, contributed to study management as a member of the TMG, provided clinical advice, contributed to data acquisition and contributed to the final report.

Manuel Rodriguez-Justo (Consultant Histopathologist) contributed to the study design, contributed to the protocol writing, provided histopathological advice, contributed to data acquisition and contributed to the final report.

Zainib Shabir (CTU Lead) contributed to the study design, contributed to study management as a member of the TMG and as CTU lead, contributed to data acquisition and contributed to the final report.

Andrew Slater (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Damian Tolan (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Simon Travis (Consultant Gastroenterologist) contributed to the study design, contributed to study management as a member of the TMG, provided clinical advice, contributed to data acquisition and contributed to the final report.

Alastair Windsor (Consultant Surgeon) contributed to the study design, contributed to the protocol writing, provided surgical advice, contributed to data acquisition and contributed to the final report.

Peter Wylie (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Ian Zealley (Consultant Radiologist) contributed to the study design, contributed to study management as a member of the TMG, interpreted trial imaging, collected trial data and contributed to the final report.

Steve Halligan (Consultant Radiologist) contributed to the study design, protocol writing and study management as a member of the TMG, interpreted trial imaging and contributed to the final report.

Publications

Taylor S, Mallett S, Bhatnagar G, Bloom S, Gupta A, Halligan S, *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 Gastroenterol* 2014;**14**:142.

Taylor SA, Mallett S, Bhatnagar G, Baldwin-Cleland R, Bloom S, Gupta A, *et al.* 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 Gastroenterol Hepatol* 2018;**3**:548–58.

Miles A, Bhatnagar G, Halligan S, Gupta A, Tolan D, Zealley I, *et al.* Magnetic resonance enterography, small bowel ultrasound and colonoscopy to diagnose and stage Crohn's disease: patient acceptability and perceived burden. *Eur Radiol* 2019;**29**:1083–93.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to available anonymised data may be granted following review.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2018;**390**:2769–78. [https://doi.org/10.1016/S0140-6736\(17\)32448-0](https://doi.org/10.1016/S0140-6736(17)32448-0)
2. Luces C, Bodger K. Economic burden of inflammatory bowel disease: a UK perspective. *Expert Rev Pharmacoecon Outcomes Res* 2006;**6**:471–82. <https://doi.org/10.1586/14737167.6.4.471>
3. Ghosh N, Premchand P. A UK cost of care model for inflammatory bowel disease. *Frontline Gastroenterol* 2015;**6**:169–74. <https://doi.org/10.1136/flgastro-2014-100514>
4. Gomollón F, Dignass A, Annese V, Tilg H, Van Assche G, Lindsay JO, *et al.* 3rd European Evidence-based Consensus on the Diagnosis and Management of Crohn's Disease 2016: Part 1: Diagnosis and medical management. *J Crohns Colitis* 2017;**11**:3–25. <https://doi.org/10.1093/ecco-jcc/jjw168>
5. Panes J, Bouhnik Y, Reinisch W, Stoker J, Taylor SA, Baumgart DC, *et al.* Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. *J Crohns Colitis* 2013;**7**:556–85. <https://doi.org/10.1016/j.crohns.2013.02.020>
6. Chatu S, Subramanian V, Pollok RC. Meta-analysis: diagnostic medical radiation exposure in inflammatory bowel disease. *Aliment Pharmacol Ther* 2012;**35**:529–39. <https://doi.org/10.1111/j.1365-2036.2011.04975.x>
7. Chatu S, Poullis A, Holmes R, Greenhalgh R, Pollok RC. Temporal trends in imaging and associated radiation exposure in inflammatory bowel disease. *Int J Clin Pract* 2013;**67**:1057–65. <https://doi.org/10.1111/ijcp.12187>
8. Estay C, Simian D, Lubascher J, Figueroa C, O'Brien A, Quera R. Ionizing radiation exposure in patients with inflammatory bowel disease: are we overexposing our patients? *J Dig Dis* 2015;**16**:83–9. <https://doi.org/10.1111/1751-2980.12213>
9. Gandhi NS, Baker ME, Goenka AH, Bullen JA, Obuchowski NA, Remer EM, *et al.* Diagnostic accuracy of CT enterography for active inflammatory terminal ileal Crohn disease: comparison of full-dose and half-dose images reconstructed with FBP and half-dose images with SAFIRE. *Radiology* 2016;**280**:436–45. <https://doi.org/10.1148/radiol.2016151281>
10. Zakeri N, Pollok RC. Diagnostic imaging and radiation exposure in inflammatory bowel disease. *World J Gastroenterol* 2016;**22**:2165–78. <https://doi.org/10.3748/wjg.v22.i7.2165>
11. Sonnenberg A, Erckenbrecht J, Peter P, Niederau C. Detection of Crohn's disease by ultrasound. *Gastroenterology* 1982;**83**:430–4.
12. Panés J, Bouzas R, Chaparro M, García-Sánchez V, Gisbert JP, Martínez de Guereñu B, *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. *Aliment Pharmacol Ther* 2011;**34**:125–45. <https://doi.org/10.1111/j.1365-2036.2011.04710.x>
13. Fraquelli M, Sarno A, Girelli C, Laudi C, Buscarini E, Villa C, *et al.* Reproducibility of bowel ultrasonography in the evaluation of Crohn's disease. *Dig Liver Dis* 2008;**40**:860–6. <https://doi.org/10.1016/j.dld.2008.04.006>
14. 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. *J Clin Gastroenterol* 1994;**19**:31–5. <https://doi.org/10.1097/00004836-199407000-00009>

15. Ahmed O, Rodrigues DM, Nguyen GC. Magnetic resonance imaging of the small bowel in Crohn's disease: a systematic review and meta-analysis. *Can J Gastroenterol Hepatol* 2016;**2016**:7857352. <https://doi.org/10.1155/2016/7857352>
16. 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 Liver* 2017;**11**:62–72. <https://doi.org/10.5009/gnl16015>
17. Church PC, Turner D, Feldman BM, Walters TD, Greer ML, Amitai MM, *et al.* Systematic review with meta-analysis: magnetic resonance enterography signs for the detection of inflammation and intestinal damage in Crohn's disease. *Aliment Pharmacol Ther* 2015;**41**:153–66. <https://doi.org/10.1111/apt.13024>
18. Dong J, Wang H, Zhao J, Zhu W, Zhang L, Gong J, *et al.* Ultrasound as a diagnostic tool in detecting active Crohn's disease: a meta-analysis of prospective studies. *Eur Radiol* 2014;**24**:26–33. <https://doi.org/10.1007/s00330-013-2973-0>
19. Fraquelli M, Colli A, Casazza G, Paggi S, Colucci A, Massironi S, *et al.* Role of US in detection of Crohn disease: meta-analysis. *Radiology* 2005;**236**:95–101. <https://doi.org/10.1148/radiol.2361040799>
20. Giles E, Barclay AR, Chippington S, Wilson DC. Systematic review: MRI enterography for assessment of small bowel involvement in paediatric Crohn's disease. *Aliment Pharmacol Ther* 2013;**37**:1121–31. <https://doi.org/10.1111/apt.12323>
21. 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?'. *Inflamm Bowel Dis* 2016;**22**:1246–61. <https://doi.org/10.1097/MIB.0000000000000727>
22. Horsthuis K, Bipat S, Stokkers PC, Stoker J. Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. *Eur Radiol* 2009;**19**:1450–60. <https://doi.org/10.1007/s00330-008-1287-0>
23. 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. *Acad Radiol* 2017;**24**:1216–25. <https://doi.org/10.1016/j.acra.2017.04.013>
24. Qiu Y, Mao R, Chen BL, Li XH, He Y, Zeng ZR, *et al.* Systematic review with meta-analysis: magnetic resonance enterography vs. computed tomography enterography for evaluating disease activity in small bowel Crohn's disease. *Aliment Pharmacol Ther* 2014;**40**:134–46. <https://doi.org/10.1111/apt.12815>
25. 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. <https://doi.org/10.1148/radiol.2471070611>
26. Puylaert CA, Tielbeek JA, Bipat S, Stoker J. Grading of Crohn's disease activity using CT, MRI, US and scintigraphy: a meta-analysis. *Eur Radiol* 2015;**25**:3295–313. <https://doi.org/10.1007/s00330-015-3737-9>
27. 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 Med Res Methodol* 2003;**3**:25. <https://doi.org/10.1186/1471-2288-3-25>
28. Whiting PF, Rutjes AW, Westwood ME, Mallett S, QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013;**66**:1093–104. <https://doi.org/10.1016/j.jclinepi.2013.05.014>

29. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Ann Intern Med* 2013;**158**:544–54. <https://doi.org/10.7326/0003-4819-158-7-201304020-00006>
30. Parente F, Maconi G, Bollani S, Anderloni A, Sampietro G, Cristaldi M, et al. Bowel ultrasound in assessment of Crohn's disease and detection of related small bowel strictures: a prospective comparative study versus X ray and intraoperative findings. *Gut* 2002;**50**:490–5. <https://doi.org/10.1136/gut.50.4.490>
31. Castiglione F, Mainenti PP, De Palma GD, Testa A, Bucci L, Pesce G, et al. Noninvasive diagnosis of small bowel Crohn's disease: direct comparison of bowel sonography and magnetic resonance enterography. *Inflamm Bowel Dis* 2013;**19**:991–8. <https://doi.org/10.1097/MIB.0b013e3182802b87>
32. 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 Technol Assess* 2007;**11**(50). <https://doi.org/10.3310/hta11500>
33. Hafeez R, Greenhalgh R, Rajan J, Bloom S, McCartney S, Halligan S, Taylor SA. Use of small bowel imaging for the diagnosis and staging of Crohn's disease: a survey of current UK practice. *Br J Radiol* 2011;**84**:508–17. <https://doi.org/10.1259/bjr/65972479>
34. Taylor S, Mallett S, Bhatnagar G, Bloom S, Gupta A, Halligan S, 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 Gastroenterol* 2014;**14**:142. <https://doi.org/10.1186/1471-230X-14-142>
35. Tolan DJ, Greenhalgh R, Zealley IA, Halligan S, Taylor SA. MR enterographic manifestations of small bowel Crohn disease. *Radiographics* 2010;**30**:367–84. <https://doi.org/10.1148/rg.302095028>
36. Taylor SA, Halligan S, Goh V, Morley S, Bassett P, Atkin W, Bartram CI. Optimizing colonic distention for multi-detector row CT colonography: effect of hyoscine butylbromide and rectal balloon catheter. *Radiology* 2003;**229**:99–108. <https://doi.org/10.1148/radiol.2291021151>
37. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 2005;**19**(Suppl. A):5A–36A. <https://doi.org/10.1155/2005/269076>
38. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980;**315**:514. [https://doi.org/10.1016/S0140-6736\(80\)92767-1](https://doi.org/10.1016/S0140-6736(80)92767-1)
39. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;**21**:835–52. <https://doi.org/10.1002/sim.1058>
40. Taylor SA, Mallett S, Bhatnagar G, Baldwin-Cleland R, Bloom S, Gupta A, et al. 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 Gastroenterol Hepatol* 2018;**3**:548–58. [https://doi.org/10.1016/s2468-1253\(18\)30161-4](https://doi.org/10.1016/s2468-1253(18)30161-4)
41. Calabrese E, La Seta F, Buccellato A, Virdone R, Pallotta N, Corazziari E, Cottone M. Crohn's disease: a comparative prospective study of transabdominal ultrasonography, small intestine contrast ultrasonography, and small bowel enema. *Inflamm Bowel Dis* 2005;**11**:139–45. <https://doi.org/10.1097/00054725-200502000-00007>
42. Pallotta N, Tomei E, Viscido A, Calabrese E, Marcheggiano A, Caprilli R, Corazziari E. Small intestine contrast ultrasonography: an alternative to radiology in the assessment of small bowel disease. *Inflamm Bowel Dis* 2005;**11**:146–53. <https://doi.org/10.1097/00054725-200502000-00008>

43. Parente F, Greco S, Molteni M, Anderloni A, Sampietro GM, Danelli PG, *et al.* Oral contrast enhanced bowel ultrasonography in the assessment of small intestine Crohn's disease. A prospective comparison with conventional ultrasound, X ray studies, and ileocolonoscopy. *Gut* 2004;**53**:1652–7. <https://doi.org/10.1136/gut.2004.041038>
44. Pallotta N, Vincoli G, Montesani C, Chirletti P, Pronio A, Caronna R, *et al.* Small intestine contrast ultrasonography (SICUS) for the detection of small bowel complications in Crohn's disease: a prospective comparative study versus intraoperative findings. *Inflamm Bowel Dis* 2012;**18**:74–84. <https://doi.org/10.1002/ibd.21678>
45. Kumar S, Hakim A, Alexakis C, Chhaya V, Tzias D, Pilcher J, *et al.* Small intestinal contrast ultrasonography for the detection of small bowel complications in Crohn's disease: correlation with intraoperative findings and magnetic resonance enterography. *J Gastroenterol Hepatol* 2015;**30**:86–91. <https://doi.org/10.1111/jgh.12724>
46. Tielbeek JA, Makanyanga JC, Bipat S, Pendsé DA, Nio CY, Vos FM, *et al.* Grading Crohn disease activity with MRI: interobserver variability of MRI features, MRI scoring of severity, and correlation with Crohn disease endoscopic index of severity. *AJR Am J Roentgenol* 2013;**201**:1220–8. <https://doi.org/10.2214/AJR.12.10341>
47. Jensen MD, Ormstrup T, Vagn-Hansen C, Østergaard L, Rafaelsen SR. Interobserver and intermodality agreement for detection of small bowel Crohn's disease with MR enterography and CT enterography. *Inflamm Bowel Dis* 2011;**17**:1081–8. <https://doi.org/10.1002/ibd.21534>
48. Dillman JR, Smith EA, Sanchez R, DiPietro MA, Dehkordy SF, Adler J, *et al.* Prospective cohort study of ultrasound–ultrasound and ultrasound–MR enterography agreement in the evaluation of pediatric small bowel Crohn disease. *Pediatr Radiol* 2016;**46**:490–7. <https://doi.org/10.1007/s00247-015-3517-3>
49. Schleder S, Pawlik M, Wiggermann P, Ott C, Fichtner-Feigl S, Müller-Wille R, *et al.* Interobserver agreement in MR enterography for diagnostic assessment in patients with Crohn's disease. *Rofo* 2013;**185**:992–7. <https://doi.org/10.1055/s-0033-1335445>
50. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;**46**:423–9. [https://doi.org/10.1016/0895-4356\(93\)90018-V](https://doi.org/10.1016/0895-4356(93)90018-V)
51. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74. <https://doi.org/10.2307/2529310>
52. Taylor SA, Avni F, Cronin CG, Hoeffel C, Kim SH, Laghi A, *et al.* The first joint ESGAR/ ESPR consensus statement on the technical performance of cross-sectional small bowel and colonic imaging. *Eur Radiol* 2017;**27**:2570–82. <https://doi.org/10.1007/s00330-016-4615-9>
53. Steward MJ, Punwani S, Proctor I, Adjei-Gyamfi Y, Chatterjee F, Bloom S, *et al.* Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. *Eur J Radiol* 2012;**81**:2080–8. <https://doi.org/10.1016/j.ejrad.2011.07.013>
54. Rimola J, Ordás I, Rodriguez S, García-Bosch O, Aceituno M, Llach J, *et al.* Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. *Inflamm Bowel Dis* 2011;**17**:1759–68. <https://doi.org/10.1002/ibd.21551>
55. Olchowy C, Cebulski K, Łasecki M, Chaber R, Olchowy A, Kałwak K, Zaleska-Dorobisz U. The presence of the gadolinium-based contrast agent depositions in the brain and symptoms of gadolinium neurotoxicity: a systematic review. *PLOS One* 2017;**12**:e0171704. <https://doi.org/10.1371/journal.pone.0171704>

56. Dohan A, Taylor S, Hoeffel C, Barret M, Allez M, Dautry R, *et al.* Diffusion-weighted MRI in Crohn's disease: current status and recommendations. *J Magn Reson Imaging* 2016;**44**:1381–96. <https://doi.org/10.1002/jmri.25325>
57. Choi SH, Kim KW, Lee JY, Kim KJ, Park SH. Diffusion-weighted magnetic resonance enterography for evaluating bowel inflammation in Crohn's disease: a systematic review and meta-analysis. *Inflamm Bowel Dis* 2016;**22**:669–79. <https://doi.org/10.1097/MIB.0000000000000607>
58. Seo N, Park SH, Kim KJ, Kang BK, Lee Y, Yang SK, *et al.* MR enterography for the evaluation of small-bowel inflammation in Crohn disease by using diffusion-weighted imaging without intravenous contrast material: a prospective noninferiority study. *Radiology* 2016;**278**:762–72. <https://doi.org/10.1148/radiol.2015150809>
59. Neubauer H, Pabst T, Dick A, Machann W, Evangelista L, Wirth C, *et al.* Small-bowel MRI in children and young adults with Crohn disease: retrospective head-to-head comparison of contrast-enhanced and diffusion-weighted MRI. *Pediatr Radiol* 2013;**43**:103–14. <https://doi.org/10.1007/s00247-012-2492-1>
60. Miles A, Bhatnagar G, Halligan S, Gupta A, Tolan D, Zealley I, *et al.* Magnetic resonance enterography, small bowel ultrasound and colonoscopy to diagnose and stage Crohn's disease: patient acceptability and perceived burden. *Eur Radiol* 2019;**29**:1083–93. <https://doi.org/10.1007/s00330-018-5661-2>
61. Bretthauer M, Kaminski MF, Løberg M, Zauber AG, Regula J, Kuipers EJ, *et al.* Population-based colonoscopy screening for colorectal cancer: a randomized clinical trial. *JAMA Intern Med* 2016;**176**:894–902. <https://doi.org/10.1001/jamainternmed.2016.0960>
62. Plumb AA, Ghanouni A, Rainbow S, Djedovic N, Marshall S, Stein J, *et al.* Patient factors associated with non-attendance at colonoscopy after a positive screening faecal occult blood test. *J Med Screen* 2017;**24**:12–19. <https://doi.org/10.1177/0969141316645629>
63. Casati J, Toner BB, de Rooy EC, Drossman DA, Maunder RG. Concerns of patients with inflammatory bowel disease: a review of emerging themes. *Dig Dis Sci* 2000;**45**:26–31. <https://doi.org/10.1023/A:1005492806777>
64. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, Rutter C. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;**27**:191–7. <https://doi.org/10.1017/S0033291796004242>
65. Salmon P, Shah R, Berg S, Williams C. Evaluating customer satisfaction with colonoscopy. *Endoscopy* 1994;**26**:342–6. <https://doi.org/10.1055/s-2007-1008988>
66. Evans RE, Taylor SA, Beare S, Halligan S, Morton A, Oliver A, *et al.* Perceived patient burden and acceptability of whole body MRI for staging lung and colorectal cancer: comparison with standard staging investigations. *Br J Radiol* 2018;**91**:20170731. <https://doi.org/10.1259/bjr.20170731>
67. Jesuratnam-Nielsen K, Løgager VB, Rezanavaz-Gheshlagh B, Munkholm P, Thomsen HS. Plain magnetic resonance imaging as an alternative in evaluating inflammation and bowel damage in inflammatory bowel disease: a prospective comparison with conventional magnetic resonance follow-through. *Scand J Gastroenterol* 2015;**50**:519–27. <https://doi.org/10.3109/00365521.2014.1003398>
68. Laghi A, Paolantonio P, Iafrate F, Borrelli O, Dito L, Tomei E, *et al.* MR of the small bowel with a biphasic oral contrast agent (polyethylene glycol): technical aspects and findings in patients affected by Crohn's disease. *Radiol Med* 2003;**106**:18–27.
69. Ajaj W, Goehde SC, Schneemann H, Ruehm SG, Debatin JF, Lauenstein TC. Dose optimization of mannitol solution for small bowel distension in MRI. *J Magn Reson Imaging* 2004;**20**:648–53. <https://doi.org/10.1002/jmri.20166>

70. Borthne AS, Abdelnoor M, Storaas T, Pierre-Jerome C, Kløw NE. Osmolarity: a decisive parameter of bowel agents in intestinal magnetic resonance imaging. *Eur Radiol* 2006;**16**:1331–6. <https://doi.org/10.1007/s00330-005-0063-7>
71. Sood RR, Joubert I, Franklin H, Doyle T, Lomas DJ. Small bowel MRI: comparison of a polyethylene glycol preparation and water as oral contrast media. *J Magn Reson Imaging* 2002;**15**:401–8. <https://doi.org/10.1002/jmri.10090>
72. Hafeez R, Punwani S, Boulos P, Bloom S, McCartney S, Halligan S, Taylor SA. Diagnostic and therapeutic impact of MR enterography in Crohn's disease. *Clin Radiol* 2011;**66**:1148–58. <https://doi.org/10.1016/j.crad.2010.12.018>
73. Patel NS, Pola S, Muralimohan R, Zou GY, Santillan C, Patel D, *et al.* Outcomes of computed tomography and magnetic resonance enterography in clinical practice of inflammatory bowel disease. *Dig Dis Sci* 2014;**59**:838–49. <https://doi.org/10.1007/s10620-013-2964-7>
74. Lang G, Schmiegel W, Nicolas V, Brechmann T. Impact of small bowel MRI in routine clinical practice on staging of Crohn's disease. *J Crohns Colitis* 2015;**9**:784–94. <https://doi.org/10.1093/ecco-jcc/jjv106>
75. García-Bosch O, Ordás I, Aceituno M, Rodríguez S, Ramírez AM, Gallego M, *et al.* Comparison of diagnostic accuracy and impact of magnetic resonance imaging and colonoscopy for the management of Crohn's disease. *J Crohns Colitis* 2016;**10**:663–9. <https://doi.org/10.1093/ecco-jcc/jjw015>
76. Wilkens R, Novak KL, Lebeuf-Taylor E, Wilson SR. Impact of intestinal ultrasound on classification and management of Crohn's disease patients with inconclusive colonoscopy. *Can J Gastroenterol Hepatol* 2016;**2016**:8745972. <https://doi.org/10.1155/2016/8745972>
77. Novak K, Tanyingoh D, Petersen F, Kucharzik T, Panaccione R, Ghosh S, *et al.* Clinic-based point of care transabdominal ultrasound for monitoring Crohn's disease: impact on clinical decision making. *J Crohns Colitis* 2015;**9**:795–801. <https://doi.org/10.1093/ecco-jcc/jjv105>
78. Levesque BG, Cipriano LE, Chang SL, Lee KK, Owens DK, Garber AM. Cost effectiveness of alternative imaging strategies for the diagnosis of small-bowel Crohn's disease. *Clin Gastroenterol Hepatol* 2010;**8**:261–7. <https://doi.org/10.1016/j.cgh.2009.10.032>
79. Cipriano LE, Levesque BG, Zaric GS, Loftus EV, Sandborn WJ. Cost-effectiveness of imaging strategies to reduce radiation-induced cancer risk in Crohn's disease. *Inflamm Bowel Dis* 2012;**18**:1240–8. <https://doi.org/10.1002/ibd.21862>
80. Goldfarb NI, Pizzi LT, Fuhr JP, Salvador C, Sikirica V, Kornbluth A, Lewis B. Diagnosing Crohn's disease: an economic analysis comparing wireless capsule endoscopy with traditional diagnostic procedures. *Dis Manag* 2004;**7**:292–304. <https://doi.org/10.1089/dis.2004.7.292>
81. Maconi G, Bolzoni E, Giussani A, Friedman AB, Duca P. Accuracy and cost of diagnostic strategies for patients with suspected Crohn's disease. *J Crohns Colitis* 2014;**8**:1684–92. <https://doi.org/10.1016/j.crohns.2014.08.005>
82. Dubinsky MC, Johanson JF, Seidman EG, Ofman JJ. Suspected inflammatory bowel disease: the clinical and economic impact of competing diagnostic strategies. *Am J Gastroenterol* 2002;**97**:2333–42. <https://doi.org/10.1111/j.1572-0241.2002.05988.x>
83. National Institute for Health Care Excellence (NICE). *Process and Methods Guides. Guide to the Methods of Technology Appraisal*. London: NICE; 2013.
84. Curtis L, Burns A. *Unit Costs of Health and Social Care 2017*. Canterbury: Personal Social Services Research Unit, University of Kent; 2017.
85. National Institute for Health Care Excellence (NICE). *Crohn's Disease: Management. Clinical Guideline [CG152]*. London: NICE; 2016.

86. Royal College of Physicians (RCP). *National Clinical Audit of Biological Therapies: UK Inflammatory Bowel Disease (IBD) Audit – Annual Report September 2016*. London: RCP; 2016.
87. NHS Improvement. *2016/17 Reference Costs*. London: NICE; 2017.
88. National Institute for Health Care Excellence (NICE). *Infliximab and Adalimumab for the Treatment of Crohn's Disease. Technology Appraisal Guidance [TA187]*. London: NICE; 2010.
89. National Institute for Health Care Excellence (NICE). *Vedolizumab for Treating Moderately to Severely Active Crohn's Disease After Prior Therapy. Technology Appraisal Guidance [TA352]*. London: NICE; 2015.
90. National Institute for Health Care Excellence (NICE). *Infliximab, Adalimumab and Golimumab for Treating Moderately to Severely Active Ulcerative Colitis After the Failure of Conventional Therapy. Technology Appraisal Guidance [TA329]*. London: NICE; 2015.
91. National Institute for Health Care Excellence (NICE). *Vedolizumab for Treating Moderately to Severely Active Ulcerative Colitis. Technology Appraisal Guidance [TA342]*. London: NICE; 2015.
92. Joint Formulary Committee. *British National Formulary*. 75th edn. London: BMJ Group and Pharmaceuticals Press; 2018.
93. Featherstone RL, Dobson J, Ederle J, Doig D, Bonati LH, Morris S, *et al*. Carotid artery stenting compared with endarterectomy in patients with symptomatic carotid stenosis (International Carotid Stenting Study): a randomised controlled trial with cost-effectiveness analysis. *Health Technol Assess* 2016;**20**:1–94. <https://doi.org/10.3310/hta20200>
94. EQ-5D. *NICE Position Statement on the EQ-5D-5L*. 2017. URL: <https://euroqol.org/nice-position-statement-on-the-eq-5d-5l/> (accessed 12 January 2017).
95. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, *et al*. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;**15**:708–15. <https://doi.org/10.1016/j.jval.2012.02.008>
96. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;**16**:219–42. <https://doi.org/10.1177/0962280206074463>
97. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–99. <https://doi.org/10.1002/sim.4067>
98. Briggs ASM, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.
99. Ponsioen CY, de Groof EJ, Eshuis EJ, Gardenbroek TJ, Bossuyt PMM, Hart A, *et al*. Laparoscopic ileocaecal resection versus infliximab for terminal ileitis in Crohn's disease: a randomised controlled, open-label, multicentre trial. *Lancet Gastroenterol Hepatol* 2017;**2**:785–92. [https://doi.org/10.1016/S2468-1253\(17\)30248-0](https://doi.org/10.1016/S2468-1253(17)30248-0)
100. 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. <https://doi.org/10.1136/gut.49.6.777>
101. Dionisio PM, Gurudu SR, Leighton JA, Leontiadis GI, Fleischer DE, Hara AK, *et al*. Capsule endoscopy has a significantly higher diagnostic yield in patients with suspected and established small-bowel Crohn's disease: a meta-analysis. *Am J Gastroenterol* 2010;**105**:1240–8. <https://doi.org/10.1038/ajg.2009.713>
102. Jensen MD, Nathan T, Rafaelsen SR, Kjeldsen J. Diagnostic accuracy of capsule endoscopy for small bowel Crohn's disease is superior to that of MR enterography or CT enterography. *Clin Gastroenterol Hepatol* 2011;**9**:124–9. <https://doi.org/10.1016/j.cgh.2010.10.019>

103. Solem CA, Loftus EV, Fletcher JG, Baron TH, Gostout CJ, Petersen BT, *et al.* Small-bowel imaging in Crohn's disease: a prospective, blinded, 4-way comparison trial. *Gastrointest Endosc* 2008;**68**:255–66. <https://doi.org/10.1016/j.gie.2008.02.017>
104. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;**350**:h2147. <https://doi.org/10.1136/bmj.h2147>
105. Bruining DH, Zimmermann EM, Loftus EV, Sandborn WJ, Sauer CG, Strong SA, Society of Abdominal Radiology Crohn's Disease-Focused Panel. Consensus recommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography in patients with small bowel Crohn's disease. *Gastroenterology* 2018;**154**:1172–94. <https://doi.org/10.1053/j.gastro.2017.11.274>
106. Tielbeek JA, Bipat S, Boellaard TN, Nio CY, Stoker J. Training readers to improve their accuracy in grading Crohn's disease activity on MRI. *Eur Radiol* 2014;**24**:1059–67. <https://doi.org/10.1007/s00330-014-3111-3>
107. Puylaert CA, Tielbeek JA, Bipat S, Boellaard TN, Nio CY, Stoker J. Long-term performance of readers trained in grading Crohn disease activity using MRI. *Acad Radiol* 2016;**23**:1539–44. <https://doi.org/10.1016/j.acra.2016.08.006>
108. Maccioni F, Bruni A, Viscido A, Colaiacomo MC, Cocco A, Montesani C, *et al.* MR imaging in patients with Crohn disease: value of T2- versus T1-weighted gadolinium-enhanced MR sequences with use of an oral superparamagnetic contrast agent. *Radiology* 2006;**238**:517–30. <https://doi.org/10.1148/radiol.2381040244>
109. Low RN, Sebrechts CP, Politoske DA, Bennett MT, Flores S, Snyder RJ, Pressman JH. Crohn disease with endoscopic correlation: single-shot fast spin-echo and gadolinium-enhanced fat-suppressed spoiled gradient-echo MR imaging. *Radiology* 2002;**222**:652–60. <https://doi.org/10.1148/radiol.2223010811>
110. Plumb AA, Pendsé DA, McCartney S, Punwani S, Halligan S, Taylor SA. Lymphoid nodular hyperplasia of the terminal ileum can mimic active Crohn disease on MR enterography. *AJR Am J Roentgenol* 2014;**203**:W400–7. <https://doi.org/10.2214/AJR.13.12055>
111. 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 Expect* 2009;**12**:18–26. <https://doi.org/10.1111/j.1369-7625.2008.00520.x>
112. Maccioni F, Viscido A, Marini M, Caprilli R. MRI evaluation of Crohn's disease of the small and large bowel with the use of negative superparamagnetic oral contrast agents. *Abdom Imaging* 2002;**27**:384–93. <https://doi.org/10.1107/s00261-001-0119-3>
113. Ajaj W, Goehde SC, Schneemann H, Ruehm SG, Debatin JF, Lauenstein TC. Oral contrast agents for small bowel MRI: comparison of different additives to optimize bowel distension. *Eur Radiol* 2004;**14**:458–64. <https://doi.org/10.1007/s00330-003-2177-0>
114. Ajaj W, Lauenstein TC, Langhorst J, Kuehle C, Goyen M, Zoepf T, *et al.* Small bowel hydro-MR imaging for optimized ileocecal distension in Crohn's disease: should an additional rectal enema filling be performed? *J Magn Reson Imaging* 2005;**22**:92–100. <https://doi.org/10.1002/jmri.20342>
115. Friedrich C, Fajfar A, Pawlik M, Hoffstetter P, Rennert J, Agha A, *et al.* Magnetic resonance enterography with and without biphasic contrast agent enema compared to conventional ileocolonoscopy in patients with Crohn's disease. *Inflamm Bowel Dis* 2012;**18**:1842–8. <https://doi.org/10.1002/ibd.22843>

116. Ajaj W, Goyen M, Schneemann H, Kuehle C, Nuefer M, Ruehm SG, *et al.* Oral contrast agents for small bowel distension in MRI: influence of the osmolarity for small bowel distention. *Eur Radiol* 2005;**15**:1400–6. <https://doi.org/10.1007/s00330-005-2711-3>
117. Kuehle CA, Ajaj W, Ladd SC, Massing S, Barkhausen J, Lauenstein TC. Hydro-MRI of the small bowel: effect of contrast volume, timing of contrast administration, and data acquisition on bowel distention. *AJR Am J Roentgenol* 2006;**187**:W375–85. <https://doi.org/10.2214/AJR.05.1079>
118. Office of Health Economics. *Syntax for EQ-5D-5L Value Set for England.zip*. London: Office of Health Economics; 2014. URL: www.ohe.org/file/syntax-eq-5d-5l-value-set-englandzip (accessed 27 September 2018).

Appendix 1 Magnetic resonance enterography sequence protocol

TABLE 40 Magnetic resonance enterography sequence protocol

Minimum	Optional
Coronal steady-state free precession gradient echo sequences without fat saturation	Axial steady-state free precession gradient echo sequences without fat saturation
Hyoscine butylbromide (20 mg intravenously)	Axial fast spin-echo T2-weighted sequence with fat saturation
Axial and coronal fast spin-echo T2-weighted sequences without fat saturation	Axial contrast-enhanced coronal T1-weighted sequences with fat saturation (60–70 seconds post injection)
Coronal fast spin-echo T2-weighted sequence with fat saturation	Coronal steady-state free precession gradient echo dynamic motility sequences
Axial diffusion-weighted images (<i>b</i> -values 50 and 600)	
Non-enhanced coronal T1-weighted sequence with fat saturation followed by contrast-enhanced coronal T1-weighted sequences with fat saturation (60–70 seconds post injection)	

Appendix 2 Magnetic resonance enterography oral contrast agent by recruitment site

TABLE 41 Magnetic resonance enterography oral contrast agent by recruitment site

Recruitment site	Oral preparation
UCLH, London	2.5% mannitol
St Mark's Hospital, Harrow	2% mannitol and 2 scoops of carob gum
Royal Free Hospital, London	2.5% mannitol
Queen Alexandra Hospital, Portsmouth	69 g of polyethylene glycol (3350/l)
Leeds General Infirmary, Leeds	2.5% mannitol
Ninewells Hospital, Dundee	2.0% mannitol and 0.2% locust bean gum
John Radcliffe Hospital, Oxford	69 g of polyethylene glycol (3350/l)
St George's Hospital, London	2.5% mannitol

Appendix 3 Summary of study outcomes by disease site and cohort

TABLE 42 Summary of study outcomes by disease site and cohort

Diagnostic end points	SBCD only	SBCD and colonic CD	Colonic CD only
Identification and localisation of disease (active or inactive)	<p>Primary outcome (difference in sensitivity only, per participant)</p> <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants <p>Secondary outcome 1 (difference in specificity, per participant)</p> <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants 	<p>Secondary outcome 1 additional analyses</p> <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants 	
Identification of active disease	<p>Secondary outcome 2</p> <ol style="list-style-type: none"> Per participant Per terminal ileum segment <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants 		<p>Secondary outcome 2</p> <ol style="list-style-type: none"> (a) Per participant and (b) Per colonic segment <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants
Identification of disease (active or inactive)	<p>Secondary outcome 3 subgroup</p> <ol style="list-style-type: none"> Per participant Per terminal ileum segment <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants 	<p>Secondary outcome 3</p> <ol style="list-style-type: none"> Per participant Per terminal ileum segment <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants 	<p>Secondary outcome 3</p> <ol style="list-style-type: none"> Per segment in colonoscopic reference only group <ul style="list-style-type: none"> Subgroup new diagnosis and suspected relapse participants

Appendix 4 Definition of agreement with reference standard for disease extent

TABLE 43 Definition of agreement with reference standard for disease extent

Primary outcome: test accuracy for disease extent (correct identification and localisation)	Correct identification of disease presence	Test accurate for disease extent?
Yes (TP)	Yes: disease correctly identified	Yes: all segments identified
No (FN)	Yes: disease correctly identified	No: one or more segments missed
No (FN)	Yes: disease correctly identified	No: incorrect segment(s) identified
No (FN)	No: no disease identified when present	No: no disease identified when present
Yes (TN)	Yes: correctly identified no disease present	Yes: correctly identified no disease present
No (FP)	No: disease in index, not in reference	No: disease in index, not in reference

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Appendix 5 Recruitment sites and recruitment numbers and withdrawals per site

TABLE 44 Recruitment sites, recruitment numbers and withdrawals per site

Recruitment site	Total participants, <i>n</i> (%)							
	Screened	Recruited			Withdrawn	Included		
		New diagnosis	Suspected relapse	Both cohorts		New diagnosis	Suspected relapse	Both cohorts
UCLH, London	177	66 (39)	69 (41)	135 (40)	19 (36)	52 (39)	64 (42)	116 (41)
St Mark's Hospital, Harrow	78	8 (5)	16 (10)	24 (7)	4 (8)	5 (4)	15 (10)	20 (6)
Royal Free Hospital, London	7	1 (0)	2 (1)	3 (1)	1 (2)	1 (1)	1 (1)	2 (1)
Queen Alexandra Hospital, Portsmouth	66	32 (19)	27 (16)	59 (18)	9 (18)	28 (20)	22 (15)	50 (18)
Leeds General Infirmary, Leeds	69	29 (17)	22 (13)	51 (15)	4 (8)	27 (20)	20 (13)	47 (17)
Ninewells Hospital, Dundee	71	11 (6)	15 (9)	26 (8)	3 (6)	9 (7)	14 (9)	23 (7)
John Radcliffe Hospital, Oxford	39	15 (9)	11 (7)	26 (8)	7 (14)	9 (7)	10 (7)	19 (7)
St George's Hospital, London	11	6 (4)	5 (3)	11 (3)	4 (8)	2 (2)	5 (3)	7 (3)
Total	518	168	167	335	51	133	151	284

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Appendix 6 Presenting symptoms of the final study cohort

TABLE 45 Presenting symptoms of the final study cohort

Presenting symptom ^a	Cohort, <i>n</i> (%)	
	New diagnosis	Suspected relapse
Diarrhoea		
No blood	75 (56)	81 (54)
Blood	39 (29)	20 (13)
Weight loss	58 (44)	53 (35)
Abdominal pain	106 (80)	119 (79)
Perianal sepsis	11 (8)	9 (6)
Obstructive symptoms	22 (17)	47 (31)
Cutaneous fistulation	0 (0)	10 (7)
Fever	15 (11)	12 (8)
Nocturnal symptoms	19 (14)	29 (19)
Uveitis	6 (5)	9 (6)
Erythema nodosum	4 (3)	6 (4)
Arthropathy	11 (8)	25 (17)
Mouth ulcers	17 (13)	15 (10)
Other	49 (37)	40 (26)

^a Participants often presented with more than one symptom or sign.

Appendix 7 Presence and activity of small bowel and colonic Crohn's disease according to individual bowel segments

TABLE 46 Presence and activity of SBCD and colonic CD according to individual bowel segments

Bowel segment	Cohort, <i>n</i> (%)	
	New diagnosis	Suspected relapse
Terminal ileum		
Present	108 (81)	109 (72)
Active	102 (94)	94 (86)
Ileum		
Present	17 (13)	21 (14)
Active	16 (94)	16 (76)
Jejunum		
Present	7 (5)	6 (4)
Active	7 (100)	6 (100)
Duodenum		
Present	7 (5)	1 (1)
Active	7 (100)	1 (100)
Caecum		
Present	51 (38)	27 (18)
Active	48 (94)	25 (93)
Ascending colon		
Present	42 (32)	25 (17)
Active	41 (98)	24 (96)
Transverse colon		
Present	41 (31)	20 (13)
Active	39 (95)	19 (95)
Descending colon		
Present	35 (26)	24 (16)
Active	34 (97)	23 (96)
Sigmoid colon		
Present	47 (35)	29 (19)
Active	47 (100)	27 (93)
Rectum		
Present	35 (26)	19 (13)
Active	34 (97)	18 (95)

Appendix 8 Quality of segmental visualisation on magnetic resonance enterography and ultrasonography according to the participant cohort

TABLE 47 Quality of segmental visualisation on MRE and US (new diagnosis cohort)

Segment	Scan quality, <i>n</i> (%)							
	MRE (<i>N</i> = 133)				US (<i>N</i> = 133)			
	Good	Moderate	Poor	NA/excised ^a	Good	Moderate	Poor	NA/excised ^a
Duodenum	45 (34)	73 (55)	15 (11)	0 (0)	41 (31)	67 (50)	23 (17)	2 (2)
Jejunum	58 (44)	59 (44)	16 (12)	0 (0)	84 (63)	42 (32)	7 (5)	0 (0)
Ileum	109 (82)	21 (16)	3 (2)	0 (0)	100 (75)	24 (18)	9 (7)	0 (0)
Terminal ileum	107 (80)	21 (16)	5 (4)	0 (0)	106 (80)	20 (15)	7 (5)	0 (0)
Caecum	79 (59)	37 (28)	15 (11)	2 (2)	91 (68)	35 (26)	6 (5)	1 (1)
Ascending colon	77 (58)	41 (31)	15 (11)	0 (0)	96 (72)	33 (25)	4 (3)	0 (0)
Transverse colon	65 (49)	40 (30)	27 (20)	1 (1)	89 (67)	40 (30)	4 (3)	0 (0)
Descending colon	58 (44)	43 (32)	31 (23)	1 (1)	92 (69)	33 (25)	8 (6)	0 (0)
Sigmoid colon	47 (35)	52 (39)	33 (25)	1 (1)	55 (41)	59 (44)	19 (14)	0 (0)
Rectum	43 (32)	49 (37)	40 (30)	1 (1)	11 (8)	47 (35)	74 (56)	1 (1)

NA, not applicable.
^a There was some inconsistency in how resected segments were scored among practitioners.

TABLE 48 Quality of segmental visualisation on MRE and US (suspected relapse cohort)

Segment	Scan quality, <i>n</i> (%)									
	MRE					US				
	Number of segments scored	Good	Moderate	Poor	NA/excised ^a	Number of segments scored	Good	Moderate	Poor	NA/excised ^a
Duodenum	151	58 (38)	75 (50)	18 (12)	0 (0)	151	32 (21)	90 (60)	27 (18)	2 (1)
Jejunum	151	73 (48)	55 (36)	23 (15)	0 (0)	151	79 (52)	62 (41)	8 (5)	2 (1)
Ileum	151	122 (81)	26 (17)	3 (2)	0 (0)	151	101 (67)	43 (28)	6 (4)	1 (1)
Terminal ileum	151	108 (72)	18 (12)	9 (6)	16 (10)	151	100 (66)	18 (12)	12 (8)	21 (14)
Caecum	92	62 (67)	28 (30)	2 (2)	0 (0)	92	61 (67)	18 (20)	5 (5)	8 (9)
Ascending colon	134	84 (63)	45 (34)	4 (3)	1 (1)	134	94 (70)	27 (20)	9 (7)	4 (3)
Transverse colon	146	78 (53)	55 (38)	12 (8)	1 (1)	146	81 (55)	55 (38)	8 (5)	2 (1)
Descending colon	147	68 (46)	58 (39)	20 (14)	1 (1)	147	92 (63)	46 (31)	7 (5)	2 (1)
Sigmoid colon	146	63 (43)	59 (40)	23 (16)	1 (1)	146	69 (47)	61 (42)	16 (11)	0 (0)
Rectum	149	57 (38)	63 (42)	27 (18)	2 (1)	149	14 (9)	62 (42)	68 (46)	5 (3)

NA, not applicable.

^a There was some inconsistency in how resected segments were scored among practitioners.

Appendix 9 Imaging and endoscopic data available to the consensus panel in participants with discrepancy for small bowel disease presence or location between magnetic resonance enterography and ultrasonography

TABLE 49 Imaging and endoscopic data available to the consensus panel in participants with discrepancy for SBCD presence or location between MRE and US

Test	Cohort, n (%)	
	New diagnosis (N = 24)	Suspected relapse (N = 29)
MRE (repeated)	1 (4)	3 (10)
US (repeated)	0 (0)	1 (3)
Colonoscopy	24 (100)	13 (45)
Gastroscopy	2 (8)	1 (3)
Sigmoidoscopy	2 (8)	2 (7)
CapE	3 (13)	2 (7)
CTE	0 (0)	1 (3)
CT abdomen and/or pelvis	2 (8)	1 (3)
MR enteroclysis	0 (0)	1 (3)
MRI abdomen and/or pelvis	0 (0)	2 (7)
BaFT	4 (17)	9 (31)
Barium enteroclysis	0 (0)	1 (3)
Hydrosonography	7 (29)	6 (21)
White cell scan	0 (0)	0 (0)
Other	3 (13)	7 (24)
No third test performed	2 (8)	3 (10)

Appendix 10 Raw data for the primary outcome and selected secondary outcomes

TABLE 50 Raw data for the primary outcome and selected secondary outcomes

Trial outcome	Outcome description	Test	Participant group	n						
				TP	FN	FP	TN	DP	DN	Total
Primary outcome: SBCD extent	SBCD extent	MRE	All	171	62	5	46	233	51	284
		US	All	152	81	13	38	233	51	284
		MRE	New diagnosis	79	32	1	21	111	22	133
		US	New diagnosis	69	42	4	18	111	22	133
		MRE	Suspected relapse	92	30	4	25	122	29	151
		US	Suspected relapse	83	39	9	20	122	29	151
Secondary outcome 1: disease extent	SBCD and colonic CD extent	MRE	All	125	145	4	10	270	14	284
		US	All	96	174	6	8	270	14	284
		MRE	New diagnosis	51	82	0	0	133	0	133
		US	New diagnosis	37	96	0	0	133	0	133
		MRE	Suspected relapse	74	63	4	10	137	14	151
		US	Suspected relapse	59	78	6	8	137	14	151
	Colonic CD extent	MRE	All	35	94	17	138	129	155	284
		US	All	29	100	17	138	129	155	284
		MRE	New diagnosis	17	60	6	50	77	56	133
		US	New diagnosis	10	67	7	49	77	56	133
		MRE	Suspected relapse	18	34	11	88	52	99	151
		US	Suspected relapse	19	33	10	89	52	99	151
Secondary outcome 2: active disease	SBCD active disease	MRE	All	187	22	20	55	209	75	284
		US	All	167	42	24	51	209	75	284
		MRE	New diagnosis	93	11	6	23	104	29	133
		US	New diagnosis	83	21	8	21	104	29	133
		MRE	Suspected relapse	94	11	14	32	105	46	151
		US	Suspected relapse	84	21	16	30	105	46	151

continued

TABLE 50 Raw data for the primary outcome and selected secondary outcomes (continued)

Trial outcome	Outcome description	Test	Participant group	n						Total
				TP	FN	FP	TN	DP	DN	
Secondary outcome 3: disease presence	TI active segment colonoscopy	MRE	All	85	15	45	41	100	86	186
		US	All	70	30	48	38	100	86	186
		MRE	New diagnosis	61	8	26	28	69	54	123
		US	New diagnosis	50	19	30	24	69	54	123
		MRE	Suspected relapse	24	7	19	13	31	32	63
		US	Suspected relapse	20	11	18	14	31	32	63
	Colonic CD active disease	MRE	All	73	53	18	140	126	158	284
		US	All	75	51	13	145	126	158	284
		MRE	New diagnosis	37	39	6	51	76	57	133
		US	New diagnosis	40	36	5	52	76	57	133
		MRE	Suspected relapse	36	14	12	89	50	101	151
		US	Suspected relapse	35	15	8	93	50	101	151
	SBCD presence per participant	MRE	All	210	23	5	46	233	51	284
		US	All	193	40	13	38	233	51	284
		MRE	New diagnosis	99	12	1	21	111	22	133
		US	New diagnosis	92	19	4	18	111	22	133
		MRE	Suspected relapse	111	11	4	25	122	29	151
		US	Suspected relapse	101	21	9	20	122	29	151
	TI segment disease presence consensus	MRE	All	191	26	6	61	217	67	284
		US	All	177	40	11	56	217	67	284
		MRE	New diagnosis	95	13	2	23	108	25	133
		US	New diagnosis	89	19	4	21	108	25	133
		MRE	Suspected relapse	96	13	4	38	109	42	151
		US	Suspected relapse	88	21	7	35	109	42	151
SBCD and colonic CD presence	MRE	All	188	82	4	10	270	14	284	
	US	All	174	96	6	8	270	14	284	
	MRE	New diagnosis	80	53	0	0	133	0	133	
	US	New diagnosis	81	52	0	0	133	0	133	
	MRE	Suspected relapse	108	29	4	10	137	14	151	
	US	Suspected relapse	93	44	6	8	137	14	151	

TABLE 50 Raw data for the primary outcome and selected secondary outcomes (continued)

Trial outcome	Outcome description	Test	Participant group	n						
				TP	FN	FP	TN	DP	DN	Total
TI segment disease presence colonoscopy	MRE	All	89	16	44	37	105	81	186	
		US	79	26	47	34	105	81	186	
	MRE	New diagnosis	62	9	26	26	71	52	123	
		US	56	15	28	24	71	52	123	
	MRE	Suspected relapse	27	7	18	11	34	29	63	
		US	23	11	19	10	34	29	63	
Colonic CD presence	MRE	All	76	53	17	138	129	155	284	
		US	84	45	17	138	129	155	284	
	MRE	New diagnosis	37	40	6	50	77	56	133	
		US	47	30	7	49	77	56	133	
	MRE	Suspected relapse	39	13	11	88	52	99	151	
		US	37	15	10	89	52	99	151	

DN, disease negative; DP, disease positive; FN, false negative; FP, false positive; TI, terminal ileum; TN, true negative; TP, true positive.

Appendix 11 Potential impact of staging small bowel Crohn's disease presence with either magnetic resonance enterography and ultrasonography in a theoretical 1000-participant cohort

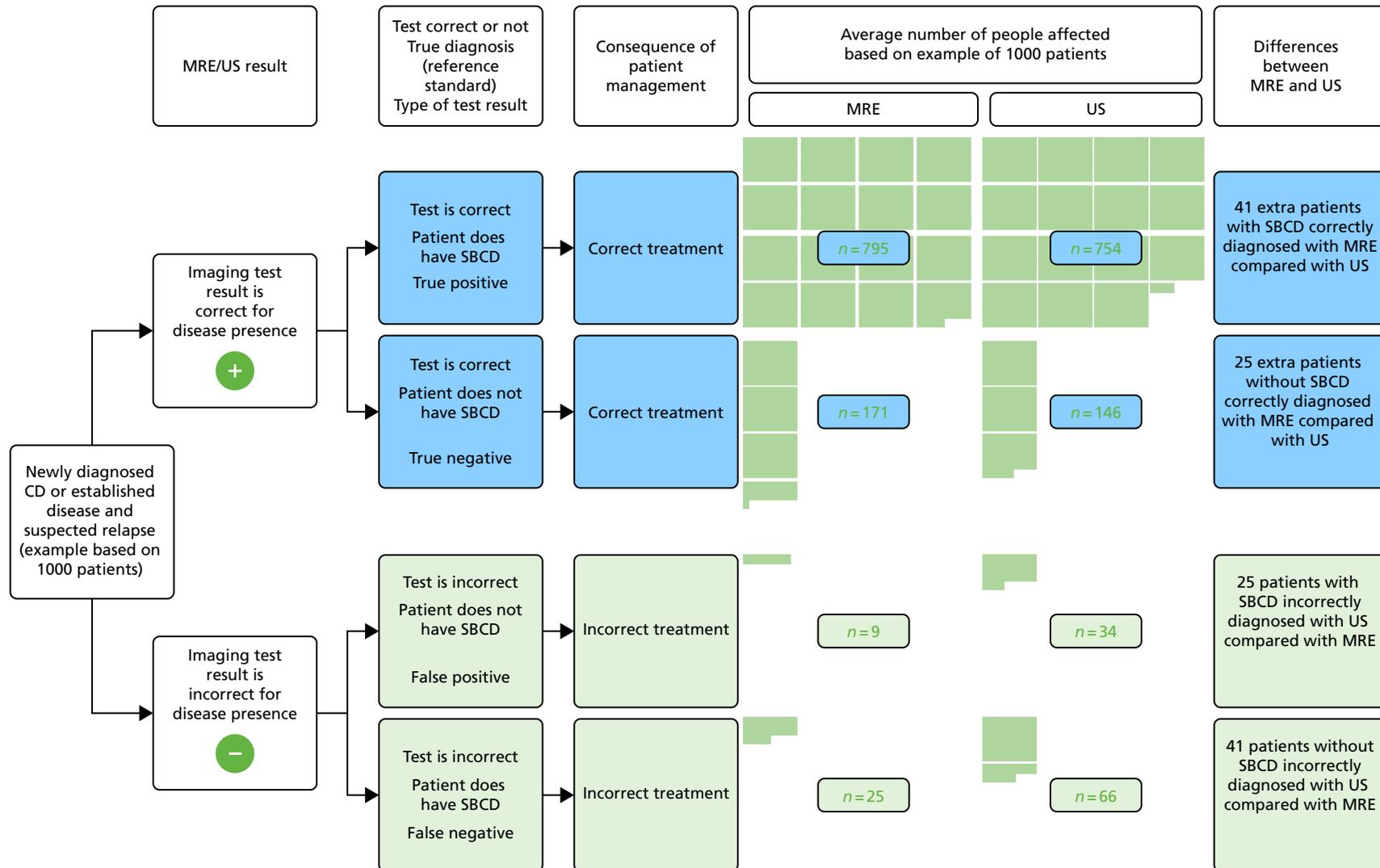


FIGURE 11 Potential impact of staging SBCD presence with either MRE or US in a theoretical 1000-participant cohort. Numbers of hypothetical participants are calculated from sensitivity and specificity results comparing test accuracy within individual participants.

Appendix 12 Per-segment sensitivity and specificity for disease presence and extent against the consensus reference standard, according to participant cohort

TABLE 51 Per-segment sensitivity and specificity for disease presence and extent against the consensus reference standard according to participant cohort. Equivocal MRE and US results considered disease positive

Segment	Cohort															
	New diagnosis (N = 133)								Suspected relapse (N = 151)							
	Disease positive (n)	Disease negative (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			Disease positive (n)	Disease negative (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)		
		MRE	US	Difference	MRE	US	Difference			MRE	US	Difference	MRE	US	Difference	
Small bowel segments																
Duodenum ^a	7	126	29 (8 to 64)	29 (8 to 64)	0 (-14 to 14; 1.000)	100 (97 to 100)	99 (96 to 100)	1 (-2 to 3; 1.000)	1	150	0 (0 to 79)	0 (0 to 79)	0 (-100 to 100; 1.000)	100 (98 to 100)	99 (95 to 100)	1 (-1 to 4; 0.500)
Jejunum ^a	7	126	43 (16 to 75)	57 (25 to 84)	-14 (-76 to 47; 1.000)	99 (96 to 100)	97 (93 to 99)	2 (-2 to 5; 0.625)	6	145	100 (61 to 100)	67 (30 to 90)	33 (-21 to 88; 0.500)	97 (93 to 99)	99 (96 to 100)	-2 (-6 to 2; 0.375)
Ileum	17	116	79 (51 to 93)	47 (23 to 72)	32 (0 to 65; 0.051)	95 (87 to 98)	94 (85 to 97)	1 (-4 to 7; 0.615)	21	130	89 (65 to 97)	64 (39 to 83)	25 (-1 to 52; 0.062)	91 (83 to 96)	92 (84 to 96)	-1 (-7 to 5; 0.833)
Terminal ileum	108	25	96 (89 to 99)	93 (83 to 97)	3 (-2 to 9; 0.203)	98 (84 to 100)	94 (72 to 99)	4 (-5 to 14; 0.394)	109	42	96 (89 to 99)	91 (81 to 97)	5 (-1 to 10; 0.114)	97 (87 to 99)	93 (78 to 98)	4 (-4 to 12; 0.314)
Colonic segments^b																
Caecum	51	82	37 (25 to 51)	39 (27 to 53)	-2 (-17 to 13; 0.796)	97 (89 to 99)	88 (79 to 93)	9 (0 to 17; 0.047)	27	65	63 (44 to 79)	59 (40 to 76)	4 (-15 to 23; 0.705)	97 (88 to 99)	94 (85 to 98)	3 (-4 to 10; 0.413)
Ascending	42	91	38 (25 to 53)	38 (25 to 53)	0 (-25 to 25; 1.000)	95 (89 to 98)	90 (82 to 95)	5 (-2 to 13; 0.127)	25	109	68 (48 to 83)	68 (48 to 83)	0 (-19 to 19; 1.000)	97 (92 to 99)	94 (88 to 98)	3 (-2 to 7; 0.255)
Transverse	41	92	39 (25 to 55)	37 (23 to 52)	2 (-10 to 15; 0.705)	99 (93 to 100)	96 (89 to 98)	3 (0 to 7; 0.079)	20	126	60 (38 to 79)	60 (38 to 79)	0 (-31 to 31; 1.000)	95 (90 to 98)	93 (88 to 97)	2 (-3 to 6; 0.479)
Descending	35	98	40 (25 to 57)	37 (23 to 54)	3 (-12 to 18; 0.705)	50 (50 to 50)	96 (90 to 98)	-46 (-50 to -42; 0.000)	24	123	71 (50 to 85)	46 (27 to 65)	25 (4 to 46; 0.019)	97 (92 to 99)	95 (89 to 97)	2 (-2 to 7; 0.255)
Sigmoid	47	86	30 (18 to 44)	28 (17 to 42)	2 (-15 to 19; 0.809)	98 (91 to 99)	93 (85 to 97)	5 (-2 to 11; 0.153)	29	117	72 (54 to 86)	69 (50 to 83)	3 (-17 to 24; 0.739)	95 (89 to 98)	93 (87 to 97)	2 (-4 to 8; 0.564)
Rectum	35	98	34 (21 to 51)	11 (4 to 27)	23 (7 to 39; 0.005)	97 (91 to 99)	94 (87 to 97)	3 (-2 to 8; 0.255)	19	130	63 (40 to 81)	42 (23 to 64)	21 (-2 to 45; 0.079)	97 (92 to 99)	93 (87 to 96)	4 (-2 to 9; 0.163)

a Analysis for duodenum and jejunum uses McNemar's test owing to small number of new participants with disease presence.

b Analysis for individual colonic segments uses a population average approach to compare imaging accuracy for individual colon segments. Segments by consensus reference standard.

Appendix 13 Sensitivity and specificity for terminal ileal and colonic Crohn's disease presence and extent (regardless of activity) versus ileocolonoscopy reference, according to participant cohort

TABLE 52 Sensitivity and specificity for terminal ileal and colonic CD presence and extent (regardless of activity) vs. ileocolonoscopy reference, according to participant cohort

Disease location	Cohort															
	New diagnosis (N = 123)								Suspected relapse (N = 63)							
	Disease positive (n)	Disease negative (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			Disease positive (n)	Disease negative (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)		
		MRE	US	Difference	MRE	US	Difference			MRE	US	Difference	MRE	US	Difference	
Terminal ileum	71	52	98 (91 to 100)	94 (81 to 98)	4 (-2 to 10; 0.200)	50 (23 to 77)	42 (18 to 70)	8 (-20 to 37; 0.560)	34	29	94 (74 to 99)	83 (50 to 96)	11 (-7 to 30; 0.232)	25 (6 to 64)	19 (4 to 57)	6 (-23 to 35; 0.676)
Colonic CD extent ^a	80	43	3 (1 to 14)	1 (0 to 7)	2 (-2 to 6; 0.237)	94 (77 to 99)	86 (63 to 96)	8 (-5 to 21; 0.239)	29	34	1 (0 to 0.11)	3 (0 to 19)	-2 (-8 to 3; 0.392)	95 (76 to 99)	93 (72 to 99)	2 (-7 to 11; 0.688)
Colonic CD presence	80	43	39 (23 to 58)	52 (33 to 70)	-13 (-34 to 7; 0.207)	94 (80 to 99)	86 (65 to 96)	8 (-5 to 20; 0.237)	29	34	47 (19 to 76)	41 (16 to 72)	6 (-28 to 40; 0.723)	95 (80 to 99)	93 (75 to 99)	2 (-7 to 11; 0.684)
Colonic segments^b																
Caecum	58	65	23 (13 to 35)	26 (16 to 39)	-3 (-17 to 10; 0.617)	88 (77 to 94)	79 (67 to 87)	9 (0 to 19; 0.052)	15	36	20 (7 to 47)	20 (7 to 47)	0 (-19 to 19; 1.000)	44 (29 to 61)	42 (27 to 58)	3 (-7 to 12; 0.563)
Ascending	49	74	24 (14 to 38)	20 (11 to 34)	4 (-6 to 14; 0.412)	90 (81 to 95)	81 (71 to 88)	9 (1 to 18; 0.030)	13	47	31 (12 to 59)	31 (12 to 59)	0 (-21 to 21; 1.000)	83 (69 to 91)	81 (67 to 90)	2 (-7 to 11; 0.655)
Transverse	4	82	25 (14 to 40)	27 (15 to 42)	-2 (-13 to 8; 0.655)	91 (83 to 96)	90 (82 to 95)	1 (-4 to 7; 0.655)	13	50	23 (8 to 52)	15 (4 to 45)	8 (-7 to 22; 0.299)	94 (83 to 98)	90 (78 to 96)	4 (-1 to 9; 0.150)
Descending	40	83	27 (16 to 43)	25 (14 to 41)	2 (-10 to 15; 0.706)	96 (89 to 99)	91 (83 to 96)	5 (0 to 9; 0.041)	18	45	28 (12 to 52)	22 (9 to 47)	6 (-5 to 16; 0.305)	93 (81 to 98)	95 (84 to 99)	-2 (-7 to 2; 0.313)
Sigmoid	54	69	24 (15 to 37)	24 (15 to 37)	0 (-15 to 15; 1.000)	96 (87 to 99)	96 (87 to 99)	0 (-7 to 7; 1.000)	20	42	25 (11 to 48)	40 (21 to 62)	-15 (-40 to 10; 0.242)	90 (77 to 96)	90 (77 to 96)	0 (-9 to 9; 1.000)
Rectum	45	78	29 (18 to 44)	13 (6 to 27)	16 (2 to 29; 0.027)	98 (90 to 99)	95 (87 to 98)	3 (-4 to 9; 0.413)	16	47	19 (6 to 45)	13 (3 to 39)	6 (-15 to 27; 0.561)	96 (84 to 99)	92 (79 to 97)	4 (-4 to 13; 0.313)

^a Agreement with reference standard for disease presence and segmental location.
^b Analysis for individual colonic segments uses a population average approach. Participants or colonic segments by ileocolonoscopy reference standard.

Appendix 14 Per-participant sensitivity and specificity the presence of active disease against the consensus reference standard, according to participant cohort

TABLE 53 Per-participant sensitivity and specificity for the presence of active disease against the consensus reference standard, according to participant cohort

Disease location	Cohort																
	New diagnosis (N = 133)									Suspected relapse (N = 151)							
	Participants with active disease (n)	Participants with inactive disease (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			Participants with active disease (n)	Participants with inactive disease (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			
		MRE	US	Difference	MRE	US	Difference			MRE	US	Difference	MRE	US	Difference		
Active SB/CD	104	29	96 (90 to 99)	90 (79 to 96)	6 (0 to 13; 0.056)	90 (68 to 98)	83 (56 to 95)	7 (-11 to 25; 0.453)	105	46	96 (90 to 99)	90 (79 to 96)	6 (0 to 13; 0.056)	79 (57 to 91)	73 (51 to 88)	6 (-14 to 25; 0.584)	
Active colonic CD	76	57	48 (30 to 66)	55 (36 to 72)	-7 (-28 to 14; 0.522)	96 (88 to 99)	97 (90 to 99)	-1 (-5 to 4; 0.720)	50	101	83 (63 to 93)	81 (59 to 92)	2 (-14 to 19; 0.779)	96 (89 to 99)	98 (93 to 99)	-2 (-5 to 2; 0.309)	
Active SB/CD and colonic CD ^a	130	3	64 (50 to 77)	59 (44 to 72)	5 (-10 to 20; 0.512)	0 (0 to 56)	0 (0 to 56)	0 (-33 to 33; 1.000)	121	30	88 (78 to 94)	73 (59 to 84)	15 (3 to 26; 0.012)	40 (25 to 58)	40 (25 to 58)	0 (-22 to 22; 1.000)	

^a Agreement with reference standard for presence of active disease to identify participant as having disease in small bowel, colon or both. Analysis uses McNemar's test for specificity due to small number of new diagnosis participants with inactive disease.

Participants by consensus reference standard.

Reproduced from Taylor *et al.*⁴⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

Appendix 15 Sensitivity and specificity for presence of active terminal ileal and colonic Crohn's disease versus ileocolonoscopy reference, according to participant cohort

TABLE 54 Sensitivity and specificity for the presence of active terminal ileal and colonic CD vs. ileocolonoscopy reference, according to participant cohort

Disease location	Cohort																
	New diagnosis (N = 123)									Suspected relapse (N = 63)							
	Participants with active disease (n)	Participants with inactive disease (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			Participants with active disease (n)	Participants with inactive disease (n)	Sensitivity, % (95% CI; p-value)			Specificity, % (95% CI; p-value)			
		MRE	US	Difference	MRE	US	Difference			MRE	US	Difference	MRE	US	Difference		
Active terminal ileum disease	69	54	99 (92 to 100)	89 (70 to 96)	10 (-1 to 21; 0.085)	54 (27 to 79)	38 (16 to 66)	16 (-11 to 44; 0.250)	31	32	93 (69 to 99)	78 (42 to 95)	15 (-8 to 38; 0.213)	31 (9 to 67)	37 (11 to 73)	-6 (-39 to 27; 0.708)	
Active colonic CD	69	54	40 (21 to 63)	46 (25 to 68)	-6 (-29 to 17; 0.629)	93 (78 to 98)	94 (88 to 98)	-1 (-9 to 6; 0.754)	21	42	55 (18 to 87)	37 (10 to 76)	18 (-23 to 59; 0.754)	95 (78 to 95)	95 (78 to 95)	0 (-8 to 8; 1.000)	

Participants by ileocolonoscopy reference standard.

Appendix 16 Perceptual errors for disease detection (regardless of activity) on magnetic resonance enterography and ultrasonography against the consensus reference (both cohorts combined)

TABLE 55 Perceptual errors for disease detection (regardless of activity) on MRE and US against the consensus reference (both cohorts combined)

Disease status ^b	Total participants scored	Perceptual errors, ^a n (%)			
		MRE		US	
		Perceptual error present	Perceptual error absent	Perceptual error present	Perceptual error absent
SBCD					
Present	233	22 (9)	211 (91)	33 (14)	200 (86)
Absent	51	0 (0)	51 (100)	0 (0)	51 (100)
Total	284	22 (8)	262 (92)	33 (12)	251 (88)
Colonic CD					
Present	129	25 (19)	104 (81)	23 (18)	106 (82)
Absent	155	0 (0)	155 (100)	0 (0)	155 (100)
Total	284	25 (9)	259 (91)	23 (8)	261 (92)

a Per participant considered separately for small bowel and colon locations.

b By consensus reference standard.

Appendix 17 Per-reader sensitivity and specificity for small bowel Crohn's disease extent (ultrasonography)

TABLE 56 Per-reader sensitivity and specificity for SBCD extent (US)

Reader	Reads per reader (n)	Disease positive (n)			Disease negative (n)			Sensitivity (%)	Specificity (%)
		Disease-positive participants ^a	TP	FN	Disease-negative participants ^a	FP	TN		
1	4	3	3	0	1	0	1	100	100
2	1	1	0	1	0	0	0	0	NA ^b
3	33	26	19	7	7	1	7	73	100
4	1	1	1	0	0	0	0	100	NA ^b
5	4	2	2	0	2	0	2	100	100
6	7	5	3	2	2	2	0	60	0
7	26	22	16	6	4	1	3	73	75

FN, false negative; FP, false positive; NA, not applicable; TN, true negative; TP, true positive.

a By consensus reference standard.

b Not applicable for sensitivity/specificity per reader when there are no disease-positive/disease-negative participants by reader.

Appendix 18 Per-reader sensitivity and specificity for small bowel Crohn's disease extent (magnetic resonance enterography)

TABLE 57 Per-reader sensitivity and specificity for SBCE extent (MRE)

Reader	Reads per reader	Disease positive (n)			Disease negative (n)			Sensitivity (%)	Specificity (%)
		Disease-positive participants ^a	TP	FN	Disease-negative participants ^a	FP	TN		
1	1	1	1	0	0	0	0	100	NA ^b
2	14	13	7	6	1	0	1	54	100
3	7	4	3	1	3	0	3	75	100
4	2	2	2	0	0	0	0	100	NA ^b
5	11	9	4	5	2	1	1	44	50
6	15	13	7	6	2	0	2	54	100
7	14	10	6	4	4	0	4	60	100
8	12	9	4	5	3	1	2	44	67
9	15	14	9	5	1	0	1	64	100
10	1	1	1	0	0	0	0	100	NA ^b
11	1	1	1	0	0	0	0	100	NA ^b
12	19	15	7	8	4	0	4	47	100
13	12	11	7	4	1	0	1	64	100
14	1	1	1	0	0	0	0	100	NA ^b
15	15	11	5	6	4	0	4	45	100
16	1	1	0	1	0	0	0	0	NA ^b
17	2	2	1	1	0	0	0	50	NA ^b
18	1	0	0	0	1	0	1	NA ^b	100
19	11	10	8	2	1	0	1	80	100
20	5	5	3	2	0	0	0	60	NA ^b
21	10	8	4	4	2	0	2	50	100
22	5	4	4	0	1	0	1	100	100
23	13	9	2	7	4	0	4	22	100
24	2	0	0	0	2	0	2	NA ^b	100
25	5	5	2	3	0	0	0	40	NA ^b
26	11	10	8	2	1	0	1	80	100
27	13	8	4	4	5	1	4	50	80

FN, false negative; FP, false positive; NA, not applicable; TN, true negative; TP, true positive.

^a By consensus reference standard.

^b Not applicable for sensitivity/specificity per reader when there are no disease-positive/disease-negative participants by reader.

Appendix 19 Participant symptom record sheet following ingestion of oral contrast prior to magnetic resonance enterography

TABLE 58 Participant symptom record sheet following ingestion of oral contrast prior to MRE

Symptom	Very tolerable	Moderately tolerable	Somewhat tolerable	Not at all tolerable	I did not experience this symptom
A feeling of fullness	<input type="checkbox"/>				
Regurgitation	<input type="checkbox"/>				
Vomiting	<input type="checkbox"/>				
Abdominal pain/spasms	<input type="checkbox"/>				
Diarrhoea	<input type="checkbox"/>				

Appendix 20 Bar chart of participant symptoms according to the oral contrast agent

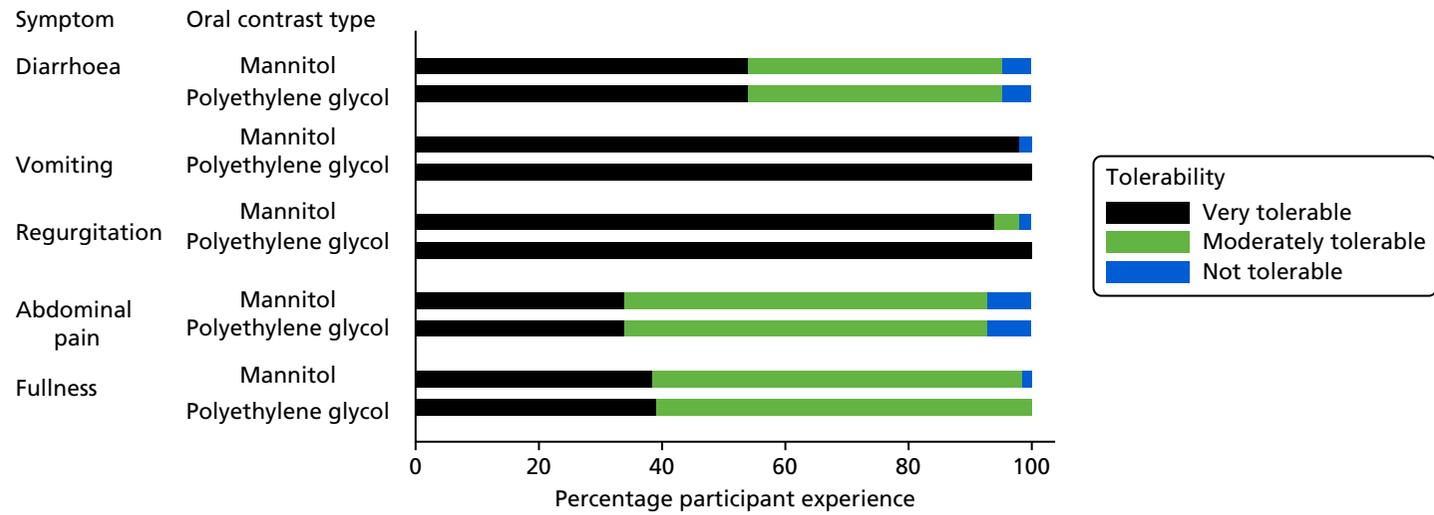


FIGURE 12 Bar chart of participant symptom tolerability by oral contrast type.

Appendix 21 Resource use by diagnostic outcome and unit costs

TABLE 59 Resource use by diagnostic outcome and unit costs

Variable	Multifocal/proximal SBCD				Isolated terminal ileal SBCD				No multifocal or isolated SBCD (N = 51)		Unit cost (£)
	Active disease (N = 42)		Inactive disease (N = 5)		Active disease (N = 168)		Inactive disease (N = 18)		Frequency (SD)	n	
	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n			
Medications, months 1–3											
5-ASAs (tablets)	0.190 (0.397)	42	0.000 (0.000)	5	0.167 (0.374)	168	0.278 (0.461)	18	0.204 (0.407)	49	76
5-ASAs (enemas)	0.048 (0.216)	42	0.000 (0.000)	5	0.024 (0.153)	168	0.167 (0.383)	18	0.061 (0.242)	49	220
Prednisolone	0.095 (0.297)	42	0.000 (0.000)	5	0.155 (0.363)	168	0.056 (0.236)	18	0.163 (0.373)	49	77
Azathioprine	0.429 (0.501)	42	0.400 (0.548)	5	0.429 (0.496)	168	0.222 (0.428)	18	0.327 (0.474)	49	28
6-MP	0.095 (0.297)	42	0.200 (0.447)	5	0.089 (0.286)	168	0.111 (0.323)	18	0.082 (0.277)	49	538
Infliximab	0.167 (0.377)	42	0.400 (0.548)	5	0.149 (0.357)	168	0.056 (0.236)	18	0.224 (0.422)	49	5984
Adalimumab	0.262 (0.445)	42	0.200 (0.447)	5	0.155 (0.363)	168	0.111 (0.323)	18	0.163 (0.373)	49	4522
Other medications	0.595 (0.734)	42	0.400 (0.548)	5	0.560 (0.920)	168	0.944 (1.305)	18	0.612 (0.862)	49	82
Medications, months 4–6											
5-ASAs (tablets)	0.175 (0.385)	40	0.000 (0.000)	5	0.172 (0.378)	163	0.313 (0.479)	16	0.143 (0.354)	49	76
5-ASAs (enemas)	0.000 (0.000)	40	0.000 (0.000)	5	0.006 (0.078)	163	0.063 (0.250)	16	0.000 (0.000)	49	220
Prednisolone	0.050 (0.221)	40	0.200 (0.447)	5	0.067 (0.252)	163	0.000 (0.000)	16	0.061 (0.242)	49	77
Azathioprine	0.400 (0.496)	40	0.400 (0.548)	5	0.417 (0.495)	163	0.188 (0.403)	16	0.286 (0.456)	49	28
6-MP	0.100 (0.304)	40	0.200 (0.447)	5	0.080 (0.272)	163	0.125 (0.342)	16	0.082 (0.277)	49	538
Infliximab	0.150 (0.362)	40	0.600 (0.548)	5	0.147 (0.355)	163	0.000 (0.000)	16	0.184 (0.391)	49	5984
Adalimumab	0.300 (0.464)	40	0.200 (0.447)	5	0.160 (0.367)	163	0.188 (0.403)	16	0.143 (0.354)	49	4522
Vedolizumab	0.050 (0.221)	40	0.000 (0.000)	5	0.000 (0.000)	163	0.000 (0.000)	16	0.000 (0.000)	49	8121
Other medications	0.600 (0.928)	40	0.400 (0.548)	5	0.337 (0.764)	163	0.750 (1.238)	16	0.306 (0.683)	49	73

Variable	Multifocal/proximal SBCD				Isolated terminal ileal SBCD				No multifocal or isolated SBCD (N = 51)		Unit cost (£)
	Active disease (N = 42)		Inactive disease (N = 5)		Active disease (N = 168)		Inactive disease (N = 18)		Frequency (SD)	n	
	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	
Primary care contacts, months 1–3											
Visits to GP	1.667 (1.500)	9	0	0	0.698 (1.170)	53	1.400 (1.517)	5	0.667 (0.970)	18	38
Telephone call to GP	0.667 (0.866)	9	0	0	0.453 (1.280)	53	0.200 (0.447)	5	0.278 (0.575)	18	15
Visit to nurse	2.556 (2.744)	9	0	0	0.528 (1.137)	53	0.000 (0.000)	5	1.778 (3.021)	18	19
Visit from nurse	0.000 (0.000)	9	0	0	0.094 (0.405)	53	0.200 (0.447)	5	0.000 (0.000)	18	41
Telephone call to nurse	2.222 (3.993)	9	0	0	1.189 (2.202)	53	0.400 (0.548)	5	0.611 (1.037)	18	8
Primary care contacts, months 4–6											
Visits to GP	3.500 (3.536)	2	0	0	0.750 (1.296)	36	1.667 (2.082)	3	0.667 (2.000)	9	38
Telephone call to GP	2.000 (2.828)	2	0	0	0.361 (0.990)	36	0.333 (0.577)	3	0.000 (0.000)	9	15
Visit to nurse	1.500 (0.707)	2	0	0	0.639 (1.376)	36	0.333 (0.577)	3	0.000 (0.000)	9	19
Visit from nurse	0.000 (0.000)	2	0	0	0.306 (1.527)	36	0.000 (0.000)	3	0.000 (0.000)	9	41
Telephone call to nurse	2.000 (2.828)	2	0	0	0.778 (2.099)	36	1.000 (1.732)	3	0.111 (0.333)	9	8
Surgical procedures, months 1–3											
Anal fistula	0.024 (0.156)	41	0.000 (0.000)	5	0.000 (0.000)	168	0.000 (0.000)	18	0.000 (0.000)	49	710
EUA	0.024 (0.156)	41	0.000 (0.000)	5	0.000 (0.000)	168	0.056 (0.236)	18	0.020 (0.143)	49	710
Ileal resection	0.000 (0.000)	41	0.000 (0.000)	5	0.018 (0.133)	168	0.000 (0.000)	18	0.000 (0.000)	49	5131
Lay open	0.024 (0.156)	41	0.000 (0.000)	5	0.000 (0.000)	168	0.000 (0.000)	18	0.000 (0.000)	49	710
Other	0.098 (0.374)	41	0.000 (0.000)	5	0.060 (0.237)	168	0.000 (0.000)	18	0.041 (0.200)	49	4825

continued

TABLE 59 Resource use by diagnostic outcome and unit costs (continued)

Variable	Multifocal/proximal SBCD				Isolated terminal ileal SBCD				No multifocal or isolated SBCD (N = 51)		Unit cost (£)
	Active disease (N = 42)		Inactive disease (N = 5)		Active disease (N = 168)		Inactive disease (N = 18)		Frequency (SD)	n	
	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	
Surgical procedures, months 4–6											
Anal fistula	0.000 (0.000)	39	0.000 (0.000)	5	0.000 (0.000)	164	0.000 (0.000)	16	0.020 (0.143)	49	710
EUA	0.000 (0.000)	39	0.000 (0.000)	5	0.000 (0.000)	164	0.000 (0.000)	16	0.020 (0.143)	49	710
Ileal resection	0.000 (0.000)	39	0.000 (0.000)	5	0.006 (0.078)	164	0.000 (0.000)	16	0.000 (0.000)	49	5131
Lay open	0.000 (0.000)	39	0.000 (0.000)	5	0.018 (0.134)	164	0.000 (0.000)	16	0.000 (0.000)	49	710
Other	0.000 (0.000)	39	0.000 (0.000)	5	0.055 (0.277)	164	0.063 (0.250)	16	0.020 (0.143)	49	4825
Hospital admissions, months 1–3											
Flare of CD	0.071 (0.463)	42	0.200 (0.447)	5	0.060 (0.238)	167	0.000 (0.000)	18	0.143 (0.500)	49	1647
Infection	0.024 (0.154)	42	0.000 (0.000)	5	0.000 (0.000)	167	0.000 (0.000)	18	0.000 (0.000)	49	2207
Other	0.119 (0.395)	42	0.200 (0.447)	5	0.048 (0.214)	167	0.000 (0.000)	18	0.041 (0.200)	49	2460
Hospital admissions, months 4–6											
Flare of CD	0.077 (0.480)	39	0.000 (0.000)	5	0.006 (0.079)	162	0.000 (0.000)	15	0.020 (0.143)	49	1647
Other	0.026 (0.160)	39	0.000 (0.000)	5	0.037 (0.189)	162	0.000 (0.000)	15	0.041 (0.286)	49	2200
Additional imaging/endoscopic investigations, months 1–3											
BaFT	0.048 (0.216)	42	0.000 (0.000)	5	0.042 (0.228)	168	0.056 (0.236)	18	0.082 (0.277)	49	126
CT abdomen and/or pelvis	0.071 (0.342)	42	0.200 (0.447)	5	0.071 (0.281)	168	0.000 (0.000)	18	0.000 (0.000)	49	112
CTE	0.071 (0.342)	42	0.000 (0.000)	5	0.036 (0.186)	168	0.000 (0.000)	18	0.020 (0.143)	49	97
CapE	0.095 (0.297)	42	0.000 (0.000)	5	0.012 (0.109)	168	0.000 (0.000)	18	0.020 (0.143)	49	1170
Colonoscopy	0.262 (0.445)	42	0.000 (0.000)	5	0.315 (0.479)	168	0.111 (0.323)	18	0.224 (0.422)	49	912
Flexible sigmoidoscopy	0.024 (0.154)	42	0.000 (0.000)	5	0.042 (0.200)	168	0.056 (0.236)	18	0.061 (0.242)	49	921
MRI enteroclysis	0.071 (0.261)	42	0.000 (0.000)	5	0.030 (0.170)	168	0.056 (0.236)	18	0.122 (0.331)	49	192

Variable	Multifocal/proximal SBCD				Isolated terminal ileal SBCD				No multifocal or isolated SBCD (N = 51)		Unit cost (£)
	Active disease (N = 42)		Inactive disease (N = 5)		Active disease (N = 168)		Inactive disease (N = 18)		Frequency (SD)	n	
	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	Frequency (SD)	n	
MRI pelvis	0.048 (0.309)	42	0.000 (0.000)	5	0.006 (0.077)	168	0.000 (0.000)	18	0.020 (0.143)	49	139
MRI small bowel	0.738 (0.544)	42	0.800 (0.447)	5	0.744 (0.489)	168	0.722 (0.461)	18	0.673 (0.516)	49	180
US small bowel	1.024 (0.643)	42	1.200 (0.837)	5	0.952 (0.646)	168	0.944 (0.539)	18	0.918 (0.640)	49	52
Other	0.071 (0.261)	42	0.000 (0.000)	5	0.083 (0.317)	168	0.111 (0.323)	18	0.143 (0.456)	49	33
Additional imaging/endoscopic investigations, months 4–6											
Barium enteroclysis	0.000 (0.000)	39	0.000 (0.000)	5	0.006 (0.079)	162	0.000 (0.000)	15	0.000 (0.000)	49	126
BaFT	0.000 (0.000)	39	0.000 (0.000)	5	0.025 (0.156)	162	0.000 (0.000)	15	0.000 (0.000)	49	126
CT abdomen and/or pelvis	0.000 (0.000)	39	0.000 (0.000)	5	0.037 (0.220)	162	0.000 (0.000)	15	0.041 (0.286)	49	112
CapE	0.051 (0.223)	39	0.000 (0.000)	5	0.006 (0.079)	162	0.000 (0.000)	15	0.000 (0.000)	49	1170
Colonoscopy	0.000 (0.000)	39	0.000 (0.000)	5	0.049 (0.217)	162	0.133 (0.352)	15	0.041 (0.200)	49	912
Flexible sigmoidoscopy	0.000 (0.000)	39	0.000 (0.000)	5	0.012 (0.111)	162	0.000 (0.000)	15	0.000 (0.000)	49	921
MRI pelvis	0.000 (0.000)	39	0.000 (0.000)	5	0.000 (0.000)	162	0.000 (0.000)	15	0.020 (0.143)	49	139
MRI small bowel	0.077 (0.270)	39	0.200 (0.447)	5	0.086 (0.282)	162	0.000 (0.000)	15	0.082 (0.277)	49	180
US small bowel	0.077 (0.480)	39	0.200 (0.447)	5	0.136 (0.361)	162	0.000 (0.000)	15	0.102 (0.306)	49	52
Other	0.103 (0.307)	39	0.000 (0.000)	5	0.123 (0.470)	162	0.267 (0.458)	15	0.204 (0.735)	49	22
Other outpatient visits											
Months 1–3	0.222 (0.548)	18	0.000	1	0.350 (1.202)	80	0.000 (0.000)	9	0.286 (0.854)	28	154
Months 4–6	0.182 (0.603)	11	0.000	1	0.532 (1.411)	62	0.167 (0.408)	6	0.143 (0.359)	21	25
5-ASA, aminosaliclylate; 6-MP, mercaptopurine; EUA, examination under anaesthetic. Results are presented for all participants (new diagnosis and suspected relapse cohorts) combined. Unit costs are in 2016–17 GBP (£) and rounded to the nearest whole pound. For CD medications, unit costs are taken from NICE guidance ^{88–91} and the <i>British National Formulary</i> . ⁹² For primary care contacts, unit costs were taken from Curtis and Burns. ⁸⁴ All other unit costs were taken from NHS Reference Costs. ⁸⁷											

Appendix 22 Mean Crohn's disease management costs and utilities per participant: complete-case analysis with no imputed data

TABLE 60 Mean CD management costs and utilities per participant: complete-case analysis with no imputed data

Category	Costs				Utilities						
	Months 1–3		Months 4–6		Baseline		3 months		6 months		
	Mean (SE) (£)	<i>n</i>	Mean (SE) (£)	<i>n</i>	Mean (SE)	<i>n</i>	Mean (SE)	<i>n</i>	Mean (SE)	<i>n</i>	
<i>New diagnosis</i>											
Multifocal/proximal SBCD											
Active disease	4290 (3359)	7	269	1	0.79 (0.13)	16	0.78 (0.20)	13	0.83 (0.17)	8	
Inactive disease	–	0	–	0	0.79	1	–	0	–	0	
Isolated terminal ileal SBCD											
Active disease	2055 (2529)	28	2597 (5977)	19	0.73 (0.24)	75	0.75 (0.25)	48	0.72 (0.31)	28	
Inactive disease	–	0	–	0	0.76 (0.04)	4	0.77	1	0.43 (0.47)	2	
No multifocal or isolated SBCD	2353 (2361)	9	1421 (1694)	4	0.82 (0.20)	21	0.78 (0.15)	13	0.81 (0.14)	9	
<i>Suspected relapse</i>											
Multifocal/proximal SBCD											
Active disease	5751 (428)	2	4637	1	0.79 (0.16)	20	0.73 (0.25)	7	0.65 (0.33)	4	
Inactive disease	–	0	–	0	0.78 (0.11)	4	0.67 (0.18)	2	0.66	1	
Isolated terminal ileal SBCD											
Active disease	3249 (3219)	25	2644 (2445)	15	0.75 (0.18)	81	0.80 (0.17)	41	0.77 (0.16)	32	
Inactive disease	1406 (2284)	5	1903 (2331)	3	0.82 (0.09)	13	0.87 (0.10)	8	0.81 (0.10)	8	
No multifocal or isolated SBCD	4234 (2832)	9	2562 (1974)	4	0.73 (0.14)	26	0.76 (0.20)	14	0.82 (0.13)	9	

Costs are in 2016–17 GBP (£). All data are raw data with no imputed values.

Appendix 23 Incremental cost-effectiveness of magnetic resonance enterography versus ultrasonography

TABLE 61 Incremental cost-effectiveness of MRE vs. US

Cohort	Mean (95% CI)		
	Total cost (£)	Total QALYs	NMB (£)
New diagnosis cohort			
MRE	4445 (3679 to 5466)	0.38 (0.36 to 0.39)	3161 (2017 to 3969)
US	4291 (3498 to 5334)	0.38 (0.36 to 0.39)	3310 (2109 to 4101)
Suspected relapse cohort			
MRE	6487 (5713 to 7238)	0.39 (0.37 to 0.39)	1214 (416 to 2029)
US	6276 (5515 to 7018)	0.38 (0.38 to 0.39)	1423 (603 to 2179)
Cohort	Incremental cost (£)	QALYs gained	INMB (£)
New diagnosis cohort			
MRE minus US	154 (–299 to 574)	0.0002 (–0.007 to 0.009)	–149 (–589 to 341)
Suspected relapse cohort			
MRE minus US	211 (–78 to 517)	0.00008 (–0.004 to 0.004)	–210 (–527 to 102)

NMB and INMB calculated at a cost-effectiveness threshold of £20,000 per QALY gained. Six-month time horizon.

Appendix 24 Mean Crohn's disease utilities per participant

TABLE 62 Mean CD utilities per participant: complete-case analysis with no imputed data using the EQ-5D-5L value set for England

Category	Utilities					
	Baseline		3 months		6 months	
	Mean (SD)	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	<i>n</i>
<i>New diagnosis cohort</i>						
Multifocal/proximal SBCD						
Active disease	0.86 (0.11)	16	0.84 (0.17)	13	0.88 (0.13)	8
Inactive disease	0.87	1	–	0	–	0
Isolated terminal ileal SBCD						
Active disease	0.80 (0.21)	75	0.81 (0.23)	48	0.78 (0.28)	28
Inactive disease	0.85 (0.05)	4	0.87	1	0.70 (0.22)	2
No multifocal or isolated SBCD	0.88 (0.15)	21	0.85 (0.13)	13	0.87 (0.12)	9
<i>Suspected relapse cohort</i>						
Multifocal/proximal SBCD						
Active disease	0.86 (0.13)	20	0.78 (0.27)	7	0.79 (0.20)	4
Inactive disease	0.85 (0.11)	4	0.77 (0.17)	2	0.73	1
Isolated terminal ileal SBCD						
Active disease	0.84 (0.13)	81	0.87 (0.13)	41	0.85 (0.14)	32
Inactive disease	0.90 (0.06)	13	0.92 (0.07)	8	0.88 (0.08)	8
No multifocal or isolated SBCD	0.82 (0.16)	26	0.85 (0.13)	14	0.89 (0.10)	9
All data are raw data with no imputed values. Utilities computed using <i>Syntax for EQ-5D-5L Value Set for England</i> . ¹¹⁸						

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library