NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Magnetic resonance enterography to predict disabling disease in newly diagnosed Crohn's disease: the METRIC-EF multivariable prediction model, multicentre diagnostic inception cohort study

Shankar Kumar,[1] Andrew Plumb,[1,2] Sue Mallett,[1] Caroline Clarke,[3] Tom Parry,[1] Jing Yi Jessica Weng,[3] Gauraang Bhatnagar,[1,4] Stuart Bloom,[5] John Hamlin,[6] Ailsa Hart,[7] Simon Travis,[8] Roser Vega,[5] Maira Hameed,[1] Anisha Bhagwanani,[4] Rebecca Greenhalgh,[9] Emma Helbren,[10] James A Stephenson,[2,11] Ian Zealley,[12] Vivienne Eze,[13] James Franklin,[14] Alison Corr,[9] Arun Gupta,[9] Elizabeth Isaac,[1] Damian Tolan,[15] William Hogg,[15] Antony Higginson,[16] Michela Cicchetti,[9] Sunita Gupta,[9] Miguel Serran,[9] Tim Raine,[17] Mohamed Ahmed,[17] Biljana Brezina,[17] Ilse Patterson,[2,17] Louise Lee,[11] Richard Pollok,[19] Jaymin Patel,[20] Abigail Seward,[20] Samantha Baillie,[19] Kashfia Chowdary,[21] Sue Philpott,[21] Anvi Wadke,[21] Steve Halligan[1] and Stuart A Taylor[1*]

[1]Centre for Medical Imaging, University College London, London, UK
[2]Department of Radiology, University College London Hospitals NHS Foundation Trust, London, UK
[3]The Research Department of Primary Care and Population Health, University College London, London, UK
[4]Department of Radiology, Frimley Health NHS Foundation Trust, Frimley Park Hospital, Surrey, UK
[5]Department of Gastroenterology, University College London Hospitals NHS Foundation Trust, London, UK
[6]Department of Gastroenterology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[7]Inflammatory Bowel Disease Unit, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK
[8]Translational Gastroenterology Unit, Kennedy Institute of Rheumatology Biomedical Research Centre, University of Oxford, Oxford, UK
[9]Department of Intestinal Imaging, St Mark's Hospital, London North West University Healthcare NHS Trust, Harrow, UK
[10]Department of Radiology, Hull University Teaching Hospitals NHS Trust, Hull, UK
[11]Department of Radiology, University Hospitals of Leicester NHS Foundation Trust, Leicester, UK
[12]Department of Radiology, Ninewells Hospital and Medical School, NHS Tayside, Dundee, UK
[13]Department of Radiology, Maidstone and Tunbridge Wells NHS Trust, Kent, UK
[14]Institute of Medical Imaging & Visualisation, Department of Medical Science & Public Health, Faculty of Health & Social Sciences, Bournemouth University, Bournemouth, UK
[15]Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[16]Department of Radiology, Queen Alexandra Hospital, Portsmouth Hospitals NHS Trust, Portsmouth, UK
[17]Department of Gastroenterology, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
[18]Department of Digestive Diseases, University Hospitals of Leicester NHS Trust, Leicester, UK
[19]Department of Gastroenterology, St George's University Hospitals NHS Foundation Trust, London, UK
[20]Department of Radiology, St George's University Hospitals NHS Foundation Trust, London, UK
[21]Comprehensive Clinical Trials Unit, University College London, London, UK

[*]**Corresponding author** stuart.taylor1@nhs.net

Published February 2026
DOI: 10.3310/THSN9956

Scientific summary

Magnetic resonance enterography to predict disabling disease in newly diagnosed Crohn's disease: the METRIC-EF multivariable prediction model, multicentre diagnostic inception cohort study

Health Technology Assessment 2026; Vol. 30: No. 18
DOI: 10.3310/THSN9956

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Crohn's disease (CD), a subtype of inflammatory bowel disease, is a chronic bowel disorder of uncertain aetiology that is progressive, can affect any part of the gastrointestinal (GI) tract from mouth to anus and typically has a relapsing and remitting disease course. Previously, the treatment of CD involved a stepwise approach guided by symptomatology, with the gradual escalation of drug therapy, including corticosteroids, immunomodulators through biologics, such as anti-tumour necrosis factor agents. However, there is a disconnect between clinical symptoms and the degree of underlying inflammatory activity, so this symptom-directed approach risks severe bowel damage if there is uncontrolled subclinical inflammation. Therefore, there is increasing interest in early treatment with biologic and immunomodulator therapy in a 'top-down' fashion to reduce the risk of progressive disease and subsequent complications. Although these agents are highly efficacious for improving symptoms and promoting bowel healing, they are associated with side effects and are relatively expensive, so administering these to all CD patients is inefficient. However, there are no validated or consensus definitions for which patients are most at risk of progressing to disabling CD. A current unmet major clinical need is the ability to accurately identify CD patients at initial diagnosis who are most at risk of developing future disabling CD. This would allow clinicians to administer early, aggressive treatment in those most likely to benefit and avoid unrequired over-treatment in others, so that side effects, complications, costs and clinical outcomes improve overall. A recent systematic review reported five clinical predictors (age, disease duration, disease location, smoking and Montreal behaviour) that demonstrated statistically significant prognostic potential to identify disabling CD. Magnetic resonance enterography (MRE) has become a first-line investigation for CD because it is highly accurate for identifying disease and activity, its extent and distribution, assessing treatment response as well as delineating complications. MRE is well placed to quantify both bowel damage and underlying inflammatory activity concurrently. Given its excellent performance characteristics for staging and monitoring CD, it has been postulated that MRE findings at diagnosis may also be able to predict clinical trajectory, but prognostic research evaluating cross-sectional imaging is lacking.

Objectives

The primary objective was to improve prediction of modified Beaugerie disabling CD (MBDD) within 5 years of diagnosis by developing and internally evaluating a multivariable prediction model comprising clinical predictors and adding predictors based on MRE. Secondary objectives were to improve prediction of disabling disease, defined by Montreal behaviour and Liège criteria, within 5 years of diagnosis, by developing and internally evaluating a multivariable model comprising clinical predictors and adding MRE severity scores. Additional secondary objectives focused on identifying predictive MRE parameters using principal component analysis (PCA) and studying the

healthcare costs of CD, specifically to estimate the healthcare costs incurred within 5 years of a new diagnosis of CD and to explore patient, imaging and disease characteristics driving higher health economic costs.

Methods

Magnetic resonance Enterography or ulTRasound In Crohn's disease Extended Follow-up (METRIC-EF) for predicting disabling disease was a multicentre, non-randomised, single-arm study of adult patients with newly diagnosed CD. We enrolled patients already recruited to the METRIC trial (ISRCTN03982913) and extended their follow-up for at least 4 years (average follow-up of 5.5 years). Magnetic resonance Enterography or ulTrasound In Crohn's Disease (METRIC) was a multicentre, prospective trial comparing the diagnostic accuracy of MRE and ultrasound (US) for the location and extent of CD. In METRIC-EF, we drew solely on those patients who were recruited into METRIC with a new diagnosis of CD and extended their follow-up for a minimum of 4 years. To achieve an adequate sample size, we supplemented these patients with a carefully matched, retrospectively identified group of patients, also newly diagnosed with CD and who had undergone MRE within 3 months of diagnosis. MRE was analysed by a pool of GI radiologists with at least 1 year of subspecialty GI imaging experience. Eleven radiologists interpreted MRE examinations from nine recruitment sites, deriving three activity/bowel damage scores: the Magnetic resonance Enterography Global Score (MEGS), Simplified Magnetic Resonance Index of Activity (sMARIA) and the Lémann Index (LI).

Consensus panels were convened at each recruitment site. These panels reviewed all available clinical information over the complete follow-up period, including biochemistry, endoscopy, imaging, surgery and overall clinical course. They also recorded the presence or absence of disabling disease. The primary definition of disabling disease was a modified version of that described by Beaugerie *et al.* We also used alternative definitions, including the Montreal behaviour and Liège criteria.

We assumed the prevalence of our MBDD definition of disabling disease would be around 55–60% at 5 years from diagnosis. The sample size was based on, including 207 participants of whom 114–124 we expected to develop MBDD. For the primary objective, we prespecified a multivariable prognostic model for predicting the development of MBDD within 5 years of diagnosis, using clinical predictors measured at diagnosis, unless otherwise stated. The clinical predictors included age, smoking status, sex, disease behaviour (stricturing or penetrating disease), perianal disease, developed MBDD ≤ 90 days from diagnosis, severe endoscopic disease, location of disease behaviour (ileal, colonic, ileocolonic, upper tract), initial need for steroid therapy and weight loss of at least 5 kg prior to diagnosis. We managed missing predictor values via multiple imputation. We evaluated whether adding MRE scores (MEGS, sMARIA and LI) improved the predictive ability of a model based on clinical predictors alone. To evaluate predictive ability, we predefined two risk group definitions (RDs) for classifying patients as high-risk or low-risk for developing MBDD. For RD1, the high-risk patients were the top 40% with the greatest predicted risk. For RD2, the high-risk patients had an absolute risk ≥ 10%. We calculated the absolute risk cut-off by sorting patients by predicted risk and using the risk of the eighth (10% of 81) patient who developed MBDD. For each RD, we estimated and compared the sensitivity, specificity and net benefit of the clinical predictor only model against the models adding MRE severity scores.

Due to a lack of patients developing disabling disease according to Montreal behaviour and Liège criteria, we could not fit statistically powered prognostic models.

We used PCA to identify the best combination of MRE features for predicting the development of MBDD. We predefined eleven features.

We estimated the healthcare costs incurred within 5 years of a new diagnosis of CD and investigated patient, imaging, treatment and other factors driving these costs. We collected NHS hospital resource use data for all patients over the 5-year follow-up period and applied unit costs to these data to calculate average (mean and median) 5-year costs per patient. We used regression analysis to identify the patient, imaging and disease characteristics driving higher NHS costs.

Results

We analysed 194 patients. The median age was 29 (interquartile range 22–44) years. Within 5 years of diagnosis, 42% (81/194) of participants developed MBDD. We found evidence of an unadjusted association between initial need for steroid therapy and developing MBDD {hazard ratio 2.11 [95% confidence interval (CI) 1.36 to 3.26]}. Using RD1, the model based on clinical predictors alone (model B) had a sensitivity of 49% (95% CI 39 to 60) and a specificity of 66% (57 to 74) for predicting the development of MBDD. There was no evidence of a significant difference in sensitivity nor specificity between model B and models adding MEGS, sMARIA and LI. Using RD2, model B had a sensitivity of 86% (77 to 92) and specificity of 35% (27 to 45) for predicting the development of MBDD. There was no evidence of a significant difference in sensitivity between model B and models adding MEGS, sMARIA and LI, but specificity was significantly lower for models adding MEGS [29% (22 to 38)] and LI [29% (22 to 38)].

The 6 principal components accounted for approximately 70% of the total variance in a PCA with 20 MRE features. However, we were unable to model the principal components to predict the development of MBDD due to collinearity issues.

The mean total 5-year per-patient costs of health care across the whole cohort were £24,267 [standard deviation (SD) £33,108]. For those developing disabling disease (modified Beaugerie criteria), mean 5-year costs were £29,763 (SD £38,278) compared to £20,327 (SD £28,368) for those without disabling disease. The largest contributor to costs was biologic use; a greater proportion of those patients developing disabling disease (modified Beaugerie criteria) received biologics [43/81 (53%)] than those without disabling disease [44/113 (39%)]. According to these unadjusted models, the following factors can be associated with driving higher costs over the first 5 years from diagnosis: age under 40 years at diagnosis, presence of perianal disease at diagnosis and presence of severe (ileocolonic) endoscopic disease at diagnosis. MRE scores at diagnosis were not associated with longer-term costs.

Conclusions

In an NHS setting, the addition of MRE activity/bowel damage scores to a multivariable model comprising existing standard clinical predictors did not improve prediction of disabling CD based on a modified Beaugerie criteria definition. Healthcare costs are increased in those aged under 40 years at diagnosis, those with perianal disease at diagnosis and those with severe (ileocolonic) endoscopic disease at diagnosis.

Recommendations for research

- Development and validation of an updated classification system for defining disabling disease.
- The predictive ability of MRE at diagnosis against alternative definitions for disabling CD.
- The predictive ability of intestinal US observations for disabling CD.

Trial registration

This trial is registered as ISRCTN76899103.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 15/59/17) and is published in full in *Health Technology Assessment*; Vol. 30, No. 18. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

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@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts 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

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

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.

This article

The research reported in this issue of the journal was funded by the HTA programme as award number 15/59/17. The contractual start date was in April 2017. The draft manuscript began editorial review in September 2023 and was accepted for publication in July 2025. 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' manuscript 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 article.

This article presents independent research funded by the National Institute for Health and Care 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, 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, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).