NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 30 • Issue 18 • February 2026

ISSN 2046-4924

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, Andrew Plumb, Sue Mallett, Caroline Clarke, Tom Parry, Jing Yi Jessica Weng, Gauraang Bhatnagar, Stuart Bloom, John Hamlin, Ailsa Hart, Simon Travis, Roser Vega, Maira Hameed, Anisha Bhagwanani, Rebecca Greenhalgh, Emma Helbren, James A Stephenson, Ian Zealley, Vivienne Eze, James Franklin, Alison Corr, Arun Gupta, Elizabeth Isaac, Damian Tolan, William Hogg, Antony Higginson, Michela Cicchetti, Sunita Gupta, Miguel Serran, Tim Raine, Mohamed Ahmed, Biljana Brezina, Ilse Patterson, Louise Lee, Richard Pollok, Jaymin Patel, Abigail Seward, Samantha Baillie, Kashfia Chowdary, Sue Philpott, Anvi Wadke, Steve Halligan and Stuart A Taylor

DOI 10.3310/THSN9956

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

This article should be referenced as follows:

Kumar S, Plumb A, Mallett S, Clarke C, Parry T, Weng JYJ, *et al*. 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 Technol Assess* 2026;**30**(18). https://doi.org/10.3310/THSN9956

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).

Abstract

Background: The ability to predict whether patients with a new diagnosis of Crohn's disease will develop disabling disease is an unmet clinical need. Magnetic resonance enterography is a first-line investigation for Crohn's disease, but its role in prognostication is unknown.

Objective(s): To improve prediction of disabling Crohn's disease within 5 years of diagnosis by developing and internally evaluating a multivariable prediction model comprising clinical predictors and adding magnetic resonance enterography scores (Magnetic resonance Enterography Global Score, Simplified Magnetic Resonance Index of Activity and Lémann Index). To estimate the healthcare costs incurred within 5 years of Crohn's disease diagnosis and to explore factors driving costs.

Design: A multicentre diagnostic inception cohort.

Setting: Nine National Health Service hospitals.

Participants: Aged ≥ 16 years with newly diagnosed Crohn's disease.

Main outcome measures: Comparative predictive ability of prognostic models, including magnetic resonance enterography scores (Magnetic resonance Enterography Global Score, Simplified Magnetic Resonance Index of Activity and Lémann Index) versus a model based on clinical predictors alone for the development of modified Beaugerie disabling Crohn's disease within 5 years of diagnosis.

Statistical analysis: We censored development of modified Beaugerie disabling disease ≤ 90 days from diagnosis, and utilised time-to-event models using Royston–Parmar flexible parametric models. Risk group definitions were prespecified; for risk group definition 1, the high-risk patients were the top 40% with the greatest predicted risk, and the high-risk patients had an absolute risk ≥ 10% for risk group definition 2. The absolute risk cut-off was calculated by sorting patients by predicted risk and using the risk of the eighth (10% of 81) patient who developed modified Beaugerie disabling disease.

Results: We studied 194 patients, median age 29, interquartile range 22–44 years. Within 5 years from diagnosis, 42% (81/194) developed modified Beaugerie disabling disease. There was a univariable association between initial need for steroid therapy and developing modified Beaugerie disabling disease [hazard ratio 2.11 (95% confidence interval 1.36 to 3.26)]. Using risk group definition 1, the baseline clinical model had 49% (95% confidence interval 39 to 60) sensitivity and 66% (95% confidence interval 57 to 74) specificity for predicting the development of modified Beaugerie disabling disease. There was no difference in sensitivity and specificity between models incorporating Magnetic resonance Enterography Global Score, Simplified Magnetic Resonance Index of Activity and Lémann Index compared to the baseline clinical model. Using risk group definition 2, the model, including magnetic resonance enterography predictors, had 86% (95% confidence interval 77 to 92) sensitivity and 35% (95% confidence interval 27 to 45) specificity for predicting the development of modified Beaugerie disabling disease. There was no difference in sensitivity between the clinical model and models incorporating Magnetic resonance Enterography Global Score, Simplified Magnetic Resonance Index of Activity and Lémann Index, but specificity was significantly lower for models incorporating Magnetic resonance Enterography Global Score [29% (95% confidence interval 22 to 38)] and Lémann Index [29% (95% confidence interval 22 to 38)].

The mean total 5-year per-patient cost of health care was £24,267 (standard deviation £33,108). Mean 5-year costs were £29,763 (standard deviation £38,278) compared to £20,327 (standard deviation £28,368) for those with and without disabling disease, respectively. The largest contributor to costs was biologic use. Age under 40 years, presence of perianal disease and presence of severe endoscopic disease were associated with higher costs.

Limitations: Liège and Montreal criteria for disabling disease could not be studied due to an insufficient event rate.

Conclusions: Addition of magnetic resonance enterography scores to a multivariable model comprising existing clinical predictors did not improve prediction of modified Beaugerie disabling disease. Healthcare costs were increased in those aged under 40 years and patients with perianal and severe endoscopic disease.

Future work: Testing the predictive ability of magnetic resonance enterography against alternative definitions for disabling Crohn's disease.

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.

Contents

List of tables

List of figures

List of abbreviations

A&E	accident and emergency
AIC	Akaike information criterion
BIC	Bayesian information criterion
CCTU	Comprehensive Clinical Trials Unit
CD	Crohn's disease
COI	cost of illness
CRF	case report form
CRP	C-reactive protein
CT	computed tomography
DWI	diffusion-weighted imaging
FC	faecal calprotectin
FP	false positive
GI	gastrointestinal
IBD	inflammatory bowel disease
LI	Lémann Index
MBDD	modified Beaugerie disabling disease
MEGS	Magnetic resonance Enterography Global Score
METRIC	Magnetic resonance Enterography or ulTRasound In Crohn's disease
METRIC-EF	Magnetic resonance Enterography or ulTRasound In Crohn's disease Extended Follow-up
MRCP	magnetic resonance cholangiopancreatography
MRE	magnetic resonance enterography
MRI	magnetic resonance imaging
NEL	non-elective long
NES	non-elective short stay
PC	principal component
PCA	principal component analysis
PET	positron emission tomography
PH	proportional hazards
PROFILE	Predicting Outcomes For Crohn's disease using a molecular biomarker
RD	risk group definition
ROC	receiver operating characteristic
SB	small bowel
sMARIA	Simplified Magnetic Resonance Index of Activity
TMG	Trial Management Group
TNF	tumour necrosis factor
TP	true positive
UCL	University College London
UCLH	University College London Hospital
US	ultrasound
WBC	white blood cell

Plain language summary

Crohn's disease is a chronic lifelong inflammatory bowel condition. Patients can have mild disease, but others develop 'disabling' disease, which often requires treatment with powerful medication and bowel surgery. At the time of diagnosis, there are currently no reliable ways to predict if a patient will have mild or disabling disease in the future, although some factors, such as young patient age and smoking, are associated with worse outcomes. Knowing which patients are most likely to develop disabling Crohn's disease would be very useful because they could be considered for medication earlier to try and prevent this. Typical medications include biologic drugs, which suppress the immune system. Magnetic resonance enterography is a bowel imaging test often performed at diagnosis which demonstrates the extent of bowel involvement, how much inflammation is present and if there are complications, such as abscesses or bowel narrowing. Using statistical modelling, we investigated whether adding detailed analysis of magnetic resonance enterography images to standard predictors, such as age and smoking, could improve the prediction of future disabling disease within 5 years. We also estimated the healthcare costs incurred within 5 years of a new diagnosis of Crohn's disease. We studied 194 newly diagnosed patients from 9 NHS hospitals, of whom 42% (81/194) developed disabling disease. Magnetic resonance enterography did not improve our predictive ability compared to standard clinical factors. In a hypothetical group of 1000 patients, we would predict 418 to develop disabling disease, of whom we would correctly predict 206 patients but incorrectly identify 212 patients who would actually not develop disabling disease. The average cost per patient over 5 years was £24,267 (£29,763 for those developing disabling disease and £20,327 for those who did not). The largest contributor to costs was biologic drugs. Magnetic resonance enterography remains vital for diagnosing and monitoring Crohn's disease but does not improve prediction of which patients will develop disabling disease.

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.

Chapter 1 Introduction

Background

Crohn's disease (CD), a subtype of inflammatory bowel disease (IBD), 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.[2–4] In nearly 50% of cases, disease involvement can also affect organs outwith the GI tract, including the joints, skin, eyes, liver, biliary tract and urinary tract.[5,6] The severity of bowel inflammation ranges from mild, superficial mucosal disease to advanced, transmural involvement.[7,8] CD is associated with considerable morbidity, especially when complications arise such as bowel strictures, fistulae and abscesses.[9,10] CD patients often require regular hospital care that may include surgery, posing substantial financial burden to the NHS, exceeding £10,000 per patient annually for those in clinical relapse.[11,12]

Current treatment strategies for Crohn's disease and prediction of 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 to biologics such as anti-tumour necrosis factor (TNF) agents.[13,14] 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.[15,16] The REACT (Randomised Evaluation of an Algorithm for Crohn's Treatment) trial compared early combined immunosuppression using anti-TNF agents and antimetabolites with conventional stepwise 'bottom-up' therapy.[17] At 2 years, the risk of major adverse outcomes was lower in the ECI group, despite no difference in symptom control between the two groups. Similarly, the CALM (Effect of tight control management on CD) trial showed that timely escalation with anti-TNF therapy based on both clinical symptoms and biomarkers resulted in superior clinical and endoscopic outcomes to treatment decisions driven solely by symptoms.[18] Most recently, the PROFILE (Predicting Outcomes For Crohn's disease using a molecular biomarker) trial reported that top-down treatment resulted in significantly better outcomes at 1 year compared to accelerated step-up treatment.[19]

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.[20] 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 not efficient.[3,6,21,22] However, there are no validated or consensus definitions for which patients are most at risk of progressing to severe/disabling CD.[23–26] Readily available biomarkers such as haemoglobin, C-reactive protein (CRP) and faecal calprotectin (FC) have a role in prognostication but have not been studied in newly diagnosed patients.[27]

A currently unmet major clinical need is the ability to robustly identify CD patients at initial diagnosis who are most at risk of developing future severe/disabling CD.[28] It would allow gastroenterologists to target early, aggressive treatment in those most likely to benefit, and avoid unrequired overtreatment in others, with better treatment targeting reducing associated side effects, complications and costs while improving outcomes.[29]

The role of magnetic resonance enterography and its potential to predict the development of disabling Crohn's disease

Diagnosis and follow-up of CD relies increasingly upon imaging, given the invasive nature of endoscopy and its inability to assess the entire small bowel (SB).[30–33] Magnetic resonance enterography (MRE) has become a first-line investigation for CD because it is highly accurate for identifying disease activity, its extent and distribution, assessing treatment response as well as delineating complications.[34–39] It has the additional benefits of conferring no exposure to ionising radiation and is generally well tolerated by patients.[40,41] As opposed to clinical biomarkers such as CRP and FC which only assess inflammation, MRE is well placed to quantify both bowel damage and underlying inflammatory activity concurrently.[42] Given its excellent performance characteristics for staging and monitoring CD, it has been postulated that MRE may also be able to predict clinical trajectory, but a recent systematic review found that primary research evaluating cross-sectional imaging is lacking.[43]

Jauregui-Amezaga *et al.* studied 112 patients with CD relapse from a single centre and compared the ability of MRE and colonoscopy to predict future disease course, defined as the need for subsequent major abdominal surgery.[44] The presence of perianal disease, strictures and/or fistulae on MRE were independent predictors of resection surgery, but severe endoscopic lesions were not, implying that underlying bowel damage rather than the degree of inflammation is more important for prognostication. Similar findings were reported by Fiorino *et al.* in a European study of 142 CD patients; bowel damage in the form of strictures, fistulae and abscesses was associated with significantly higher future risk of surgery.[45] In a single-centre longitudinal study, Fernández-Clotet and colleagues studied 72 CD patients who had at least 5 years' follow-up.[46] They found that stricturing disease, fistulae and ileal disease were associated with progressive CD. Finally, in a single centre of 52 patients, perienteric inflammation, restricted diffusion, penetrating disease and fibrofatty proliferation on outpatient MRE were associated with CD patients progressing to resective bowel surgery within 5 years. Longer length bowel involvement and more significant upstream bowel dilatation from strictures were also found in those patients progressing to surgery.[47]

The above studies suggest that MRE may have a predictive role to identify patients that are destined to develop severe/disabling CD. However, none have specifically studied newly diagnosed patients, which is fundamental since it is at this initial stage that gastroenterologists could administer early aggressive treatment while refraining from unwarranted treatment in others. To address this, we undertook the Magnetic resonance Enterography or ulTRasound In Crohn's disease Extended Follow-up (METRIC-EF) for predicting disabling disease trial to answer the question: 'Do MRE features at diagnosis improve prediction of disabling CD within 5 years of diagnosis?'

Chapter 2 Trial objectives

Primary objective

- To improve prediction of disabling CD within 5 years of diagnosis by developing and internally evaluating a multivariable prediction model comprising both existing standard predictors and those based on MRE.

Secondary objectives

- To improve the prediction of disabling CD within 5 years, defined by the Montreal and Liège behaviour criteria, by developing and internally evaluating a multivariable prediction model comprising clinical predictors and adding MRE scores.
- To identify specific combinations of individual MRE findings that best predict disabling CD within 5 years of diagnosis.
- 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.
- Assuming promising predictive potential, to then generate a research design for a subsequent prospective study to externally evaluate our MRE-based prediction model if appropriate.

Chapter 3 Methods

Study design

Magnetic resonance Enterography or ulTRasound In Crohn's disease Extended Follow-up for predicting disabling disease was a multivariable prediction model, multicentre diagnostic inception cohort of adult patients with newly diagnosed CD. We enrolled patients who had already been recruited to the Magnetic resonance Enterography or ulTRasound In Crohn's disease (METRIC) trial and extended their follow-up.[35,49,50] To achieve an adequate sample size, we also included a separate retrospective cohort.[1] We reported this trial according to TRIPOD reporting guidelines for prediction model development.[51]

Study population

Magnetic resonance Enterography or Ultrasound In Crohn's Disease was a multicentre, prospective trial performed in nine NHS centres across England and Scotland designed to compare the diagnostic accuracy of MRE and ultrasound (US) for the location and extent of CD.[35,50] Consenting adult patients presenting with either newly diagnosed CD or presenting with suspected relapse were recruited: all underwent both MRE and US. Patients were followed up for a minimum of 6 months. In METRIC-EF, we drew solely on the group of patients who were recruited into METRIC with a new diagnosis of CD (i.e. the 'relapse cohort' was excluded). We extended follow-up for the new diagnosis cohort to a minimum of 4 years.[35] To achieve an adequate sample size, we supplemented newly diagnosed patients from METRIC with a carefully matched retrospectively identified group of patients also newly diagnosed with CD. A total of 147 (76%) patients were from the retrospective group. With ethical approval, additional patients recruited to the METRIC trial could be included via the retrospective arm if they did not respond to two direct invitations to take part. Because data from the retrospective cohort was pseudonymised to the central trial team, we were unable to record how many patients originally recruited to the METRIC trial were included in the retrospective cohort. The median number of years from recruitment (diagnosis) until the last outpatient visit was 6.3 [interquartile range (IQR) 5.2–7.3] years. The date of recruitment (time of diagnosis) ranged from 10 June 2009 to 5 December 2017. The date of the last outpatient visit ranged from 28 February 2016 to 20 October 2022.

Eligibility criteria

The study focused on newly diagnosed CD patients either (a) enrolled in METRIC ('METRIC cohort') or (b) imaged using MRE as part of their routine care at diagnosis ('retrospective cohort') (*Figure 1*).

Magnetic resonance Enterography or ulTRasound In Crohn's disease cohort: inclusion criteria

All confirmed new diagnoses from METRIC were eligible for the present study; inclusion criteria were therefore equivalent to those of METRIC:

- Patients aged 16 years or more.
- New CD diagnoses (within 3 months of time of recruitment to METRIC), based on standard endoscopic, histological, clinical and radiological findings.

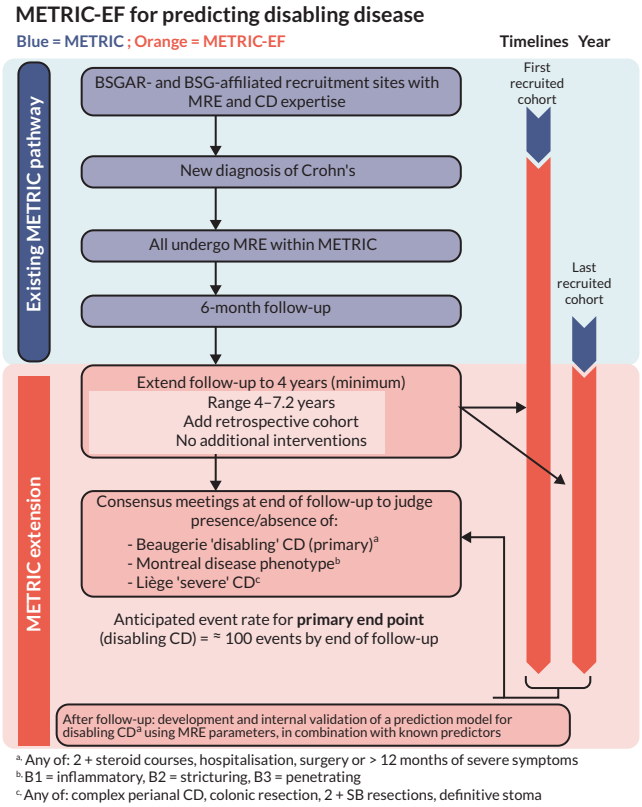

FIGURE 1 Flow diagram outlining the stages of the METRIC-EF trial. BSG, British Society of Gastroenterology; BSGAR, British Society of Gastrointestinal and Abdominal Radiology. This figure has been reproduced with permission from Kumar *et al.*[1] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The figure includes minor additions and formatting changes to the original text.

Additional retrospective cohort: Inclusion criteria

We added a retrospective cohort to the METRIC accruals to achieve the required sample size.

Inclusion criteria for the retrospective cohort were:

- Patients 16 year or more with newly diagnosed CD, based on endoscopic, histological, clinical and radiological findings
- MRE acquired according to METRIC standard minimum sequence data set, and performed either < 3 months before or after diagnosis
- Normal institutional practice is to perform MRE in all new diagnoses of CD.
- At least 4 years' clinical follow-up data available

Sites that were not part of the original METRIC trial were eligible to be recruitment sites for the retrospective cohort if they fulfilled all eligibility criteria.

Exclusion criteria

Exclusion criteria for METRIC (and so carried forward) were:

- Any psychiatric or other disorder likely to impact on informed consent
- Evidence of severe (non-Crohn's) comorbidities which makes it undesirable for the patient to participate in the study
- Pregnancy

- Contraindication to magnetic resonance imaging (MRI) (e.g. cardiac pacemaker, severe claustrophobia, inability to lie flat)
- Final diagnosis other than CD
- Enrolled in the METRIC study but not part of the final new diagnosis cohort

Ethics

Consent

The new diagnosis cohort patients recruited to METRIC were approached and consented (if willing) for participation in METRIC-EF at seven of the participating sites in the METRIC study. Patients declining participation were excluded. Two additional sites that were not originally METRIC recruitment sites were added for identification of patients to the retrospective cohort. We were granted permission to collate data from the retrospective cohort without direct patient consent (including patients recruited to the METRIC trial who did not respond to re-contact) as there was no direct patient intervention, and only pseudonymised data were collected by the Comprehensive Clinical Trials Unit (CCTU) at University College London (UCL).

Ethical permission

The METRIC-EF study achieved NHS Research Ethics Committee (NHS REC), London – Hampstead Research Ethics Committee approval on 26 October 2018 (IRAS 217422) and was conducted in accordance with the principles of good clinical practice. UCL's CCTU supervised the study.

Magnetic resonance imaging

Sequences

Magnetic resonance Enterography or UlTrasound In Crohn's disease patients underwent a standard minimum MRE sequence data set (1.5 T or 3 T) (*Table 1*). This was required for retrospective accruals, with the exception of diffusion-weighted imaging (DWI), which was not needed to calculate relevant activity and bowel damage indices.

We calculated established MRE activity indices as follows.

TABLE 1 Required and optional sequences for the MRE studies

Required	Optional
Coronal TrueFISP	Axial TrueFISP
Axial HASTE	Dynamic TrueFISP motility
Coronal HASTE	
Coronal HASTE with fat suppression	Axial HASTE with fat suppression
Axial DWI (b50 and b600)[a]	Additional b values
Coronal pre- and post-gadolinium VIBE (60–70 seconds)[a]	Axial post-gadolinium VIBE

HASTE, Half-Fourier Acquisition Single-shot Turbo spin Echo; TrueFISP, True Fast imaging with Steady State Precession; VIBE, volumetric interpolated breath-hold examination.
a Optional for retrospective cohort. Equivalent sequences from various MRI imaging vendor platforms were permitted.

Magnetic resonance Enterography Global Score

This score encompasses aspects of both inflammatory activity and bowel damage and has been tested against several reference standards, including a composite clinical reference,[52] FC,[53,54] CRP[54] and endoscopy (*Table 2*).[54,55]

Simplified Magnetic Resonance Index of Activity

The Simplified Magnetic Resonance Index of Activity (sMARIA) has been tested against endoscopic reference standards and is used increasingly to assess treatment response in clinical trials (*Table 3*).[36–38,56,57]

Lémann Index

The Lémann Index (LI) is based on comprehensive assessment of structural bowel damage, including stricturing, penetrating lesions (fistulae and abscesses), and surgical resection, and is applicable to different settings, such as early or advanced disease, patients with or without surgery, or with different CD locations and extension (*Table 4*).[58] The score comprises several factors that can be assessed either clinically, or using imaging, or via endoscopy. We used the imaging-derived score. Since the anal canal was not imaged specifically for METRIC, we omitted this score.

Interpretation and blinding

Magnetic resonance enterography examinations were sent by each recruitment site to a central facility at University College London Hospitals NHS Foundation Trust (UCLH) through pseudonymised compact disc or digital versatile discs (CD/DVDs). Here, they were uploaded to an online viewing platform (3Dnet®; Biotronics3D Ltd, London, UK) which offered the same functionality of a routinely used picture archiving and communication system (PACS) with the benefit of being accessible from any personal computer over the internet with appropriate password credentials.

TABLE 2 Calculation of the MEGS

Mural features	0	1	2	3	Score
Mural thickness	< 3 mm	> 3–5mm	> 5–7mm	> 7 mm	a
Mural T2 signal (oedema)	Normal	Minor increase	Moderate increase	Large increase	b
Perimural T2 signal	Normal	Increased signal but no fluid	Small (≤ 2 mm) fluid rim	Large (> 2 mm fluid rim)	c
Contrast enhancement: amount	Normal	Minor increase	Moderate increase	Large increase	d
Contrast enhancement: pattern	N/A or homogeneous	Mucosal	Layered		e
Haustral loss (colon only)	None	< 1/3 segment	1/3–2/3 segment	> 2/3 segment	f
Mural score for that segment					***a + b + c + d + e + f = g***
Multiplication factor	**1**	**1.5**	**2**	**Total segmental score g[a] multiplication factor**	
Length of disease in that segment	< 5 cm	5–15 cm	> 15 cm		

a Each enteric segment (jejunum; proximal ileum; terminal ileum; caecum; ascending colon; transverse colon; descending colon; sigmoid colon; rectum) is scored separately. The segmental score is then multiplied by a factor depending on the length of disease involvement in that segment. Finally, scores for extramural features are added, giving a total score (maximum possible = 296). Sum all segments, then add extramural score on a per-scan basis; 5 points for each of: (1) lymph nodes > 1 cm short axis, (2) comb sign (linear structures on the mesenteric border of an affected bowel segment), (3) abscess and (4) fistula.

Source

This table has been reproduced with permission from Kumar *et al.*[1] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The table includes minor additions and formatting changes to the original text.

TABLE 3 Derivation of the sMARIA

Feature	Description
Mural thickness	Binary: Measured in mm using software callipers, scored as abnormal if > 3 mm
Mural oedema	Binary: present if there is high signal intensity on T2 sequences with fat saturation, compared with normal-appearing loops
Fat stranding	Binary: present if there is loss of the normal sharp interface between the intestinal wall and mesentery, with oedema/fluid in the perienteric fat
Ulceration	Binary: present if mucosal surface has a deep depression, visible on 2 MRI sequences
sMARIA score for that segment	= 1 point for each of mural thickness, mural oedema, and fat stranding; 2 points for ulceration (maximum 5 points per segment)

Source
This table has been reproduced with permission from Kumar *et al.*[1] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The table includes minor additions and formatting changes to the original text.

TABLE 4 Derivation of the LI

Organ	Method of assessment	*n*[a]	Segment	Grade 1	Grade 2	Grade 3
Surgical interventions[b]						
Upper tract	History	3	Oesophagus, stomach, duodenum	–	Bypass diversion or strictureplasty	Resection
SB	History	20	Each 20 cm SB segment	–	Bypass diversion or strictureplasty	Resection
Colon/rectum	History	6	Each colonic segment	–	Stoma, bypass diversion or strictureplasty	Resection
Stricturing lesions						
Organ	Method of assessment	*n*	Segment	Grade 1	Grade 2	Grade 3
Upper tract	MRI	2	Stomach, duodenum	Wall < 3 mm; segmental enhancement without prestenotic dilatation	Wall thickening ≥ 3 mm or mural stratification with no prestenotic dilatation	Stricture with prestenotic dilatation
SB	MRI	20	Each 20 cm SB segment	Wall < 3 mm; segmental enhancement without prestenotic dilatation	Wall thickening ≥ 3 mm or mural stratification with no prestenotic dilatation	Stricture with prestenotic dilatation
Colon/rectum	MRI	6	Each colonic segment	Wall < 3 mm; segmental enhancement without prestenotic dilatation	Wall thickening ≥ 3 mm or mural stratification with no prestenotic dilatation	Stricture with prestenotic dilatation or > 50% of the lumen
Penetrating lesions						
Upper tract	MRI	2	Stomach, duodenum	–	Deep transmural ulceration	Phlegmon or fistula
SB	MRI	20	Each 20 cm SB segment	–	Deep transmural ulceration	Phlegmon or fistula

TABLE 4 Derivation of the LI (*continued*)

Organ	Method of assessment	*n*[a]	Segment	Grade 1	Grade 2	Grade 3
Colon/rectum	MRI	6	Each colonic segment	–	Transmural ulceration	Phlegmon or fistula

a *n* = number of segments within a particular organ.
b This information was collated from patient records, although a relevant past surgical history was very rare since included patients were, by definition, those with a new diagnosis of CD.

Magnetic resonance enterography scans were interpreted by one from a pool of radiologists; all were GI radiologists and experienced in MRE, in both clinical and research settings. They all had achieved Fellow of the Royal College of Radiologists and had at least 1 year of subspecialty GI imaging experience. Eleven radiologists interpreted the MRE examinations from nine centres (range 6–36 reads per radiologist).

Radiologists were allocated MRE scans for scoring by the CTU; readers did not interpret cases from their own hospital and were blinded to all clinical information other than that relevant for the calculation of the relevant index (e.g. surgical history for LI). We used Research Electronic Data Capture to record the radiological parameters.

Assessment of disabling disease at follow-up

Time point of follow-up

Follow-up was for a minimum of 4 years: since participants were recruited to METRIC over 30-months, this corresponded to an average follow-up of approximately 5.5 years. This was felt to be sufficient time for clinically relevant complications of CD to manifest.[24,59,60]

Collection of data were completed by the research team at each recruitment site using study case report forms (CRFs) and retrospective data available in the patient hospital record.

Consensus panel assessment of disease outcome

Consensus panels were convened at each recruitment site. Panels comprised, as a minimum, one gastroenterologist and one radiologist, aided by the site research nurse. The consensus panels reviewed all available clinical information over the complete follow-up period. This included investigations such as CRP, FC, endoscopy (conventional and capsule), imaging [MRI, US, computed tomography (CT), fluoroscopy], surgical and histopathological findings, clinical activity scores (e.g. Harvey-Bradshaw Index) and overall clinical course, including outpatient and inpatient clinical records.

Using all the available data, the consensus panels recorded the presence or absence of disabling disease (see definitions below), as well as the Montreal classification and the date at which this end point was reached. Participant data for the predefined clinical predictors for model inclusion (see below) were also collated when available.

Primary definition of disabling disease

The primary definition of disabling disease was a modified version of that described by Beaugerie *et al.*[25] The original definition was modified to clarify some of the symptoms and to permit use of disease-modifying therapy, since this has become common clinical practice. We also did not include events occurring within 90 days of diagnosis, to exclude events occurring within the initial diagnostic period. Disabling disease was therefore defined as any of any event reported at more than 90 days after diagnosis:

- Hospitalisation after CD diagnosis for flare or disease complication, as judged by the treating clinician.
- More than two corticosteroid courses required over 5 years and/or dependence on corticosteroids.

- Any intestinal resection > 50 cm, or surgical operation for perianal disease (examination under anaesthesia without seton placement did not meet this criterion; abscess drainage and/or seton placement did).
- Chronic disabling symptoms, defined as a cumulative time of over 12 months of one or more of:
  - Diarrhoea with nocturnal stool (getting up for a bowel movement after having gone to bed).
  - Urgency (defined as having to rush to the toilet for a bowel movement).
- Abdominal pain due to intestinal obstruction (requires imaging confirmation or surgical proof).
- Fever (documented tympanic temperature of > 38.0 °C or oral temperature of > 38.3 °C).
- Fatigue.
- Joint pain not due to an alternative cause.
- Uveitis.
- Pyoderma gangrenosum.

Alternative definitions of disabling disease

Since the Beaugerie criteria are imperfect, further definitions of adverse outcomes were also collected; specifically the Liège criteria,[24] and Montreal behaviour criteria.[61]

The Liège criteria were met if any of the following occurred:

- development of complex perianal disease
- any colonic resection
- two or more SB resections
- a single SB resection of > 50 cm
- definitive stoma.

Complex perianal disease was defined as per the American Gastroenterological Association.[62]

The Montreal behaviour criteria classify CD as either inflammatory (B1), stricturing (B2) or penetrating (B3). Stricturing disease was defined as a fixed luminal narrowing of > 50% relative to normal proximal bowel. Penetrating disease was defined as an intra-abdominal or enterocutaneous fistula, inflammatory mass, or abscess.

Outcomes

Primary outcome

Comparative predictive ability of prognostic models incorporating MRE scores (MEGS, sMARIA and LI) in combination with clinical predictors versus a model using standard clinical characteristics alone to predict disabling CD [modified Beaugerie disabling disease (MBDD)] within 5 years of diagnosis.

Secondary outcomes

1. Comparative predictive ability of prognostic models incorporating MRE scores (MEGS, sMARIA and LI) in combination with clinical predictors versus models using standard characteristics alone to predict the development of disabling CD within 5 years of diagnosis, defined by Montreal behaviour and Liège criteria.
2. Identification of the best combination of individual MRE features for the prediction of MBDD within 5 years of diagnosis.
3. Evaluation of the average costs per-participant for the NHS, incurred within 5 years of CD diagnosis.
4. Investigation of the participant, disease phenotype, and imaging characteristics associated with higher costs within 5 years of CD diagnosis.

Sample size and justification

Assumptions

We assumed that the prevalence of MBDD was approximately 55–60%; this was informed primarily by the external validation cohort of the Beaugerie descriptors, in which 57% of 361 participants had developed disabling disease within 5 years of diagnosis.[24] In support, a local audit of 33 newly diagnosed patients at 1 METRIC recruitment centre at the trial planning stage found 5 of 33 (15%) patients met the definition by mean 11.3 months, giving 16% at 1 year. Extrapolation to 5 years gave 58% prevalence, similar to that expected from the literature.[24]

The sample size was based on, including 207 participants newly diagnosed with CD; 207 participants provided 114–124 patients developed MBDD; the smaller proportion defines the minimum sample size for powering a modelling study. During the study, due to problems obtaining consent for additional follow-up due to the COVID-19 pandemic, the Trial Management Group (TMG) reduced the original target recruitment from 167 to 131 in the prospective METRIC cohort, with a corresponding increased target of 76 participants from the retrospective cohort. We anticipated that this sample size would provide between 114 and 124 patients developing MBDD. We would increase the number of the retrospective cohort to meet the 207-participant target if recruitment to the METRIC cohort was below 131.

Adequacy of this number of events/non-events

Calculating sample sizes for prognostic studies suffers from a relative lack of readily applied methods suitable for all study designs, since sample size for development depends on whether the primary aim is to select potential variables for a new model (via univariable significance within a data set) or to evaluate a model where the variables have been prespecified and are therefore fixed. In the present study, we fixed predictors since we were explicit that we would evaluate three MRE severity scores in the context of a model using fixed clinical predictors. Therefore, recommendations for sample sizes relevant to external validation were most appropriate. Accordingly, the literature suggested that we required 80–100 events for model evaluation where predictors were prespecified and fixed.[63] This also provided sufficient power to assess whether the addition of the 3 MRE severity scores enhanced prediction, under the hitherto widely used 'rule of thumb' of 10–20 events per predictor.[64] We are aware of recent methods to calculate model development and external validation sample size, but these were not reported in 2017, when the present study was powered.[65]

Power for secondary outcomes

Other definitions of adverse outcome

Development of Montreal severe disease was estimated to be 43% at 5 years.[66]

Development of Liège disabling disease was estimated to be 20% at 5 years.[24] This provided approximately 41 events for the present study which was likely insufficient to develop meaningful prognostic models. Accordingly, we planned that analysis for this end point would be descriptive only, unless our assumptions proved incorrect and sufficient events satisfying this definition had been accumulated.

Identification of the most important magnetic resonance enterography variables for model inclusion

We used principal component analysis (PCA) to reduce the number of individual MRE features to ideally two or three eigenvector variables for subsequent addition to the clinical predictor only model. This allowed us to determine how adding MRE features affected model performance.

Retention

Participants did not undergo additional testing to enter this study. Only data obtained during routine clinical care were necessary to both define disabling disease and provide variables for model inclusion. Where participants were lost to local follow-up, participants' general practitioner were contacted in an attempt to obtain routine clinical information, post consent (this was only applicable to the METRIC cohort and those patients on retrospective cohort who had provided consent).

Statistical methods: outcomes

Primary outcome

Comparative predictive ability of prognostic models incorporating MRE scores (MEGS, sMARIA and LI) versus a model based on clinical predictors alone for the development of MBDD within 5 years of diagnosis.

We developed a Royston–Parmar (RP) flexible parametric multivariable prognostic model using the following prespecified clinical predictors (based on a prior literature review and in consultation with the trial investigators):

- age at diagnosis (< 40, ≥ 40 years)
- smoking history
- sex
- disease status at diagnosis (stricturing disease, perianal disease, severe endoscopic disease)
- location of disease (ileal, colonic, ileocolonic, upper GI tract disease)
- initial need for steroid therapy
- weight loss of at least 5 kg prior to diagnosis
- CRP
- white blood cell (WBC) count
- FC
- haemoglobin
- platelet count.

There were five prespecified continuous predictors, including CRP level, WBC count, FC level, haemoglobin level and platelet count. We determined whether we should include the predictors as linear or to use fractional polynomials. Due to high levels of missing WBC count and FC levels, we could not investigate if fractional polynomial was appropriate, so assumed a linear relationship. For the remaining predictors, we calculated the best fractional polynomial models by searching through all power combinations. Then, we calculated *p*-values by comparing the deviance of the linear and fractional polynomial model 1 against the deviance of fractional polynomial model 2 (lowest deviance). We determined that retaining linear continuous predictors was the most efficient.

We retained categorical predictors as in the clinical report form, except for when modelling required us to combine specific levels of a predictor. For location of disease behaviour, we combined the ileal and upper tract levels due to small number of patients (*N* = 4) with disease in the upper tract.

Seventy-five per cent of participants had ≥ 1 predictor value missing. Because missing values were likely to be 'missing at random' based on other participant variables and to avoid loss in efficiency, we imputed values for smoking status, weight loss ≥ 5 kg prior to diagnosis, perianal disease, severe endoscopic disease, CRP level, WBC count, FC, haemoglobin level and platelet count using multiple imputation by chained equations [mi impute command in Stata 18 (StataCorp LP, College Station, TX, USA)].[67] We created 20 imputed data sets from a set of imputation models constructed from all predictors and outcomes (event indicator and Nelson–Aalen estimator for time to event).

We based an improvement in model performance on an increase in the number of patients correctly predicted to develop MBDD, relative to the clinical predictor only model. We used sensitivity, specificity and net benefit as measures of model performance. We conducted internal validation using 200 bootstrap samples (sampling with replacement) or until estimates remained stable. We describe the results from internal validation in *Appendix 8*, *Table 31*. We did not adjust for optimism. Statistical significance was based on Wilson's 95% CI. We calculated *p*-values for differences in sensitivity and specificity using McNemar's test.

Secondary outcomes

Secondary outcome 1

Comparative predictive ability of prognostic models, including MRE scores (MEGS, sMARIA and LI) versus a model based on clinical predictors alone to predict the development of disabling CD within 5 years of diagnosis, defined by

Montreal behaviour and Liège criteria. We conducted modelling using the same methods as in the primary outcome. Models were only developed if the number of patients developing disabling CD was adequate. Otherwise, we provided descriptive statistics.

Secondary outcome 2

Identification of the best combination of individual MRE features for predicting disabling CD (all definitions) within 5 years of new diagnosis. PCA was used to combine multiple MRE parameters into a small number of eigenscores variables. This allowed a larger number of features to be combined without compromising statistical power. The most influential imaging features were identified for further simplification of MRE variables included in modelling. Methods were as in the primary outcome, and the statistical significance of including MRE features were evaluated based on improvement of model fit [Bayesian information criterion (BIC)] in comparison to the standard model, with additional model performance reported as appropriate.

Model testing

To provide additional clinical relevance for potential model implementation, we formed a group from the trial group, including three gastroenterologists and two radiologists. The group in consensus defined a priori how the models could be best utilised in clinical practice. Specifically, following guidance from the study statisticians, we set two risk group definitions (RDs) for identifying patients at high and low risk of developing disabling disease which we felt would have clinical utility.

For RD1, the high-risk group included the top 40% of participants with the greatest predicted risk from the model. For RD2, the high-risk group included participants with an absolute risk ≥ 10%. The absolute risk threshold is determined by sorting the participants by predicted risk and then using the predicted risk of the eighth (10% of 81) participant who developed MBDD as the threshold.

Secondary outcomes 3 and 4; see Chapter 5.

Chapter 4 Results

Participants

The final cohort consisted of 194 patients, median age 29, IQR 22–44 years (*Figure 2*). Ninety-three (48%) of the final cohort were male.

All patients and patient events were followed up across the full 5-year period. Because it was possible to include non-contactable patients recruited to the METRIC trial in the retrospective group, it was not possible to define the exact split between the two cohorts (as no patient identifiable data could be sent to the clinical trial unit). Direct consent was obtained from 47 METRIC patients.

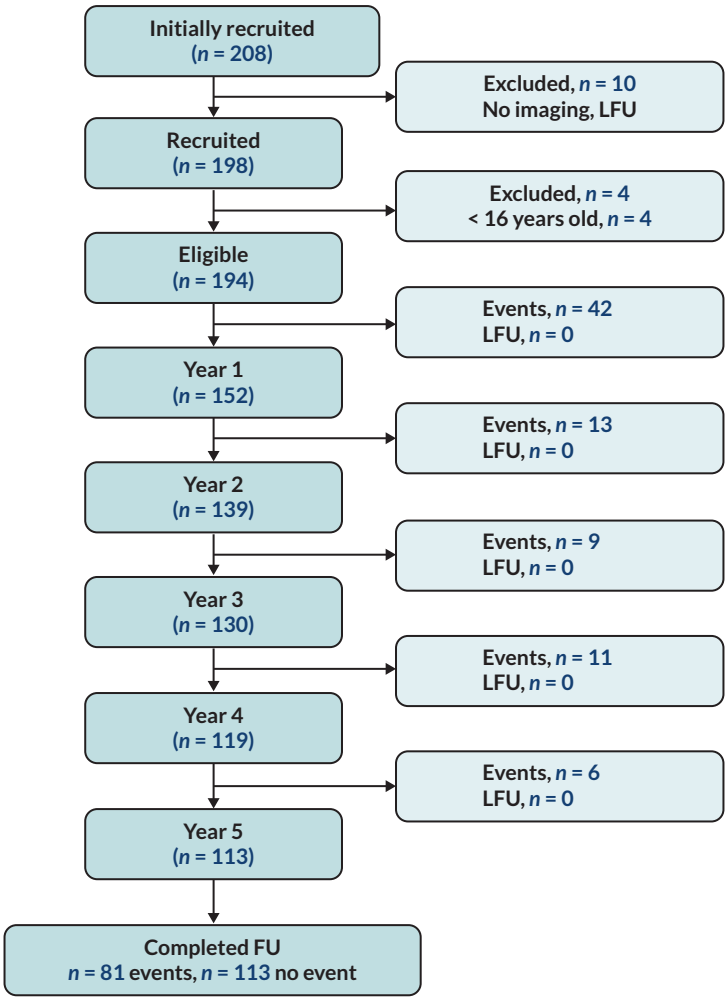

FIGURE 2 Participant flow through trial identifying participants by timing of first event during the follow-up period (> 90 days to 5 years). FU, follow-up; LFU, lost to follow-up. Reproduced with permission from Taylor *et al.*[48] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The figure includes minor additions and formatting changes to the original text.

According to the modified Beaugerie criteria, 42% (81/194) of participants developed disabling disease between 90 days post diagnosis to 5 years from diagnosis (*Table 5* and *6*). An additional 25 patients were hospitalised within the first 90 days and although they also fulfilled the modified Beaugerie criteria for disabling disease, after discussion by the TMG, these were not treated as events as these patients were being treated immediately at diagnosis, taking priority over consideration of instigating top-down therapy; including events within 90 days of diagnosis would impact on the clinical utility of any prognostic model.

TABLE 5 Number of participants who developed disabling disease within 5 years of diagnosis, according to the modified Beaugerie, Montreal B2 or B3, and Liège criteria

	Modified Beaugerie criteria	Montreal B2 or B3	Liège criteria
Years from diagnosis to developing disabling disease	***N* = 81**	***N* = 12**	***N* = 39**
1	43 (52)	2 (17)	28 (72)
2	13 (16)	1 (8)	3 (8)
3	9 (11)	5 (42)	4 (10)
4	11 (14)	3 (25)	3 (8)
5	6 (7)	1 (8)	1 (3)

TABLE 6 Demographic and clinical characteristics of participants who developed MBDD within 5 years of diagnosis

		MBDD negative	MBDD positive	Total
Demographic and clinical characteristics		***N* = 113**	***N* = 81**	***N* = 194**
Age (years)		31 (22–49)	27 (22–37)	29 (22–44)
Sex	Male	54 (48)	39 (48)	93 (48)
	Female	59 (52)	42 (52)	101 (52)
Recruitment site	Cambridge	2 (2)	7 (9)	9 (5)
	Leeds	24 (21)	18 (22)	42 (22)
	Leicester	5 (4)	0 (0)	5 (3)
	Ninewells	2 (2)	7 (9)	9 (5)
	Oxford	7 (6)	10 (12)	17 (9)
	Portsmouth	19 (17)	8 (10)	27 (14)
	St George's	2 (2)	3 (4)	5 (3)
	St Marks	4 (4)	6 (7)	10 (5)
	UCLH	48 (42)	22 (27)	70 (36)
Medication administered within 5 years from diagnosis[a]	Aminosalicylate	52	37	89
	Biologic	80	113	193
	Immunomodulator	143	117	260
	Other	11	27	38
	Steroid	104	156	260

a Participants could be administered more than one of the same medications within 5 years from diagnosis.

Note
Data are *n* (%) or median (IQR).

Only 20% (39/194) developed disabling disease according to the Liège criteria and 6% (12/194) developed B2 or B3 disease on the Montreal criteria (see *Table 5*).

The event rate using Liège and Montreal definitions of disabling disease was insufficient to perform survival analysis/predictive modelling and therefore all data were analysed using the primary outcome measure (modified Beaugerie criteria) (see *Table 6*).

Of those with an event that is, development of disabling disease according to the modified Beaugerie criteria, the median survival time prior to the event was 0.82 (IQR 0.42–2.75) years. A total of 81 events occurred in the 5-year follow-up period, 42 (52%) in year 1 and 6 (7%) in year 5 (*Table 7*). Most Beaugerie-defining events were hospitalisation due to a CD flare or complication (48 events; 59%) or due to use of corticosteroids (21 events; 26%; *Table 7*).

Clinical predictors of disabling disease

The number of participants with the predefined candidate clinical predictors of disabling disease is shown in *Tables 8–10*.

Prognostic models

A summary of the various models is shown in *Figure 3*.

Model B included the prespecified clinical predictors (excluding blood and stool markers). Models A1, A2 and A3 added the MEGS, sMARIA and LI scores, respectively. Model B1 added the blood and stool markers to the baseline clinical model (model B) (see *Figure 3*).

TABLE 7 Number of participants who developed MBDD over years from diagnosis, stratified by descriptors

	Years from diagnosis to developing MBDD					
Descriptors	**1**	**2**	**3**	**4**	**5**	**Total**
Hospitalisation due to a CD flare or complication	27	10	5	4	2	48
≥ 3 corticosteroid courses or dependence on corticosteroids	11	4	3	4	1	23
Intestinal resection > 50 cm or surgery for perianal disease[a]	1	0	0	0	0	1
Diarrhoea with nocturnal stools	0	0	0	1	0	1
Urgency	1	0	0	2	1	4
Abdominal pain due to intestinal obstruction	0	0	2	2	1	5
Fever	0	0	0	0	0	0
Fatigue	0	0	0	1	0	1
Joint pain not caused by another factor	1	1	0	0	1	3
Uveitis	2	0	0	0	0	2
Pyoderma gangrenosum	0	0	0	0	0	0
Total	43	15	10	14	6	88[b]

a Forty-five (23%) participants had an intestinal resection within 5 years of diagnosis.
b Participants could fulfil multiple descriptors on the same day.

Note
Data are *n*.

TABLE 8 Number of participants who developed MBDD within 5 years of diagnosis, stratified by prespecified clinical predictors

Prespecified clinical predictors		MBDD negative *N* = 113	MBDD positive *N* = 81	Total *N* = 194
Age category (years)	< 40	76 (67)	62 (77)	138 (71)
	≥ 40	37 (33)	19 (23)	56 (29)
Sex	Male	54 (48)	39 (48)	93 (48)
	Female	59 (52)	42 (52)	101 (52)
Smoking status	Non-smoker	81 (72)	49 (60)	130 (67)
	Smoker	22 (19)	25 (31)	47 (24)
	Missing	10 (9)	7 (9)	17 (9)
Weight loss ≥ 5 kg prior to diagnosis	Absent	71 (63)	51 (63)	122 (63)
	Present	28 (25)	18 (22)	46 (24)
	Missing	14 (12)	12 (15)	26 (13)
Initial need for steroid therapy	Absent	84 (74)	43 (53)	127 (65)
	Present	29 (26)	38 (47)	67 (35)
Developed MBDD ≤ 90 days from diagnosis	Absent	100 (88)	69 (85)	169 (87)
	Present	13 (12)	12 (15)	25 (13)
Perianal disease	Absent	100 (88)	70 (86)	170 (88)
	Present	12 (11)	11 (14)	23 (12)
	Missing	1 (1)	0 (0)	1 (0)
Severe endoscopic disease	Absent	74 (65)	56 (69)	130 (67)
	Present	27 (24)	20 (25)	47 (24)
	Missing	12 (11)	5 (6)	17 (9)
Disease behaviour	B1	80 (71)	50 (62)	130 (67)
	B2	17 (15)	17 (21)	34 (18)
	B3	16 (14)	14 (17)	30 (15)
Location of disease behaviour	Ileocolonic	52 (46)	42 (52)	94 (48)
	Ileal/upper tract	41 (36)	28 (35)	69 (36)
	Colonic	20 (18)	11 (14)	31 (16)

Note
Data are *n* (%).

Years to events by predictors

Table 11 and *Figure 4* show the median survival time to developing disabling disease according to the prespecified clinical predictors. The median MBDD free survival time of all participants was 0.82 (IQR 0.42–2.75) years.

Univariable hazard ratio by clinical predictors

The univariable hazard ratio (HR) for the prespecified clinical predictors and MRI scores is summarised in *Table 12*. Initial requirement for steroid therapy was associated with a higher risk of an event (development of disabling disease) in both original and imputed data.

TABLE 9 Number of participants who developed MBDD within 5 years of diagnosis, stratified by prespecified blood and stool predictors

Prespecified blood and stool predictors		MBDD negative *N* = 113	MBDD positive *N* = 81	Total *N* = 194
CRP level (mg/l)	*n* (%)	87 (77)	75 (93)	162 (84)
	Median (IQR)	12 (4–39)	16 (6–56)	14 (6–46)
WBC count (10^9/l)	*n* (%)	82 (73)	70 (86)	152 (78)
	Median (IQR)	9 (8–12)	9 (7–12)	9 (8–12)
FC level (µg/g)	*n* (%)	43 (38)	30 (37)	73 (38)
	Median (IQR)	527 (108–600)	521 (196–600)	527 (132–600)
Haemoglobin level (g/l)	*n* (%)	86 (76)	70 (86)	156 (80)
	Mean (SD)	126 (18)	125 (18)	126 (18)
Platelet count (10^9/l)	*n* (%)	78 (69)	69 (85)	147 (76)
	Mean (SD)	380 (127)	380 (127)	380 (127)

TABLE 10 Number of participants who developed MBDD within 5 years of diagnosis, stratified by prespecified MRE score predictors

Prespecified MRE score predictors	MBDD negative *N* = 113	MBDD positive *N* = 81	Total *N* = 194
Global MEGS	21 (8–34)	24 (8–37)	22 (8–35)
Normalised global MEGS (%)	14 (5–23)	17 (5–25)	15 (5–24)
Global sMARIA	5 (2–8)	5 (2–6)	5 (2–7)
Normalised global sMARIA (%)	18 (7–29)	18 (7–21)	18 (7–25)
LI	2 (1–4)	2 (1–3)	2 (1–3)
Normalised LI (%)	10 (4–21)	11 (6–19)	11 (4–20)

Note
Data are median (IQR). Scores were normalised to enable comparison of the scores on a standardised scale.

Source
Reproduced with permission from Taylor *et al.*[48] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The table includes minor additions and formatting changes to the original text.

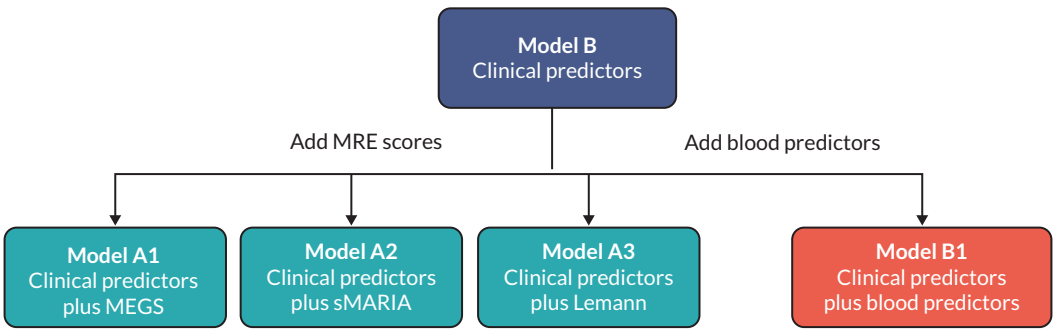

FIGURE 3 Summary of the derivation of prognostic models for developing disabling disease according to the modified Beaugerie criteria. Reproduced with permission from Taylor *et al.*[48] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The figure includes minor additions and formatting changes to the original text.

TABLE 11 Modified Beaugerie disabling disease event-free time, stratified by prespecified clinical predictors

Prespecified clinical predictors		MBDD positive	MBDD event-free time (years)
Age category (years)	< 40	62	0.77 (0.42–2.75)
	≥ 40	19	0.93 (0.39–2.82)
Sex	Male	39	1.25 (0.41–3.02)
	Female	42	0.68 (0.43–2.11)
Smoking status	Non-smoker	49	0.93 (0.39–2.69)
	Smoker	25	0.63 (0.49–3.41)
Weight loss ≥ 5 kg prior to diagnosis	Absent	51	0.62 (0.41–2.75)
	Present	18	1.75 (0.37–3.02)
Initial need for steroid therapy	Absent	43	1.25 (0.52–2.75)
	Present	38	0.50 (0.35–2.82)
Developed MBDD ≤ 90 days from diagnosis	Absent	69	1.24 (0.46–2.85)
	Present	12	0.51 (0.29–1.21)
Perianal disease	Absent	70	0.88 (0.39–2.55)
	Present	11	0.62 (0.46–3.92)
Severe endoscopic disease	Absent	56	0.73 (0.41–2.80)
	Present	20	0.88 (0.44–2.02)
Disease behaviour	B1	50	1.03 (0.44–2.85)
	B2	17	1.44 (0.47–2.69)
	B3	14	0.47 (0.36–1.77)
Location of disease behaviour	Ileocolonic	42	0.68 (0.43–2.92)
	Ileal/upper tract	28	1.51 (0.40–2.33)
	Colonic	11	0.62 (0.39–3.18)

Note
Data are median (IQR).

The multivariable models (models B, A1, A2, A3 and B1) are shown in *Appendix 1*, *Tables 21–25*. The corresponding receiver operating characteristic (ROC) curves are shown in *Appendix 2*, *Figure 8*.

Modified Beaugerie disabling disease event-free Kaplan–Meier for clinical risk groups

The predictive performance of all the models was assessed using the two a priori defined clinical risk groups and shown in *Figure 5*. For RD1, the high-risk group included the top 40% of participants with the greatest predicted risk from the model. For RD2, the high-risk group included participants with an absolute risk ≥ 10%.

Tables 13 and *14* show the sensitivity and specificity of each model for identifying patients in each of the two predefined risk groups.

Overall, there was no statistically significant difference in sensitivity or specificity between model B and each of the other models for patient risk group 1 (see *Table 13*). For patient risk group 2, models A1, A3 and B1 had significantly

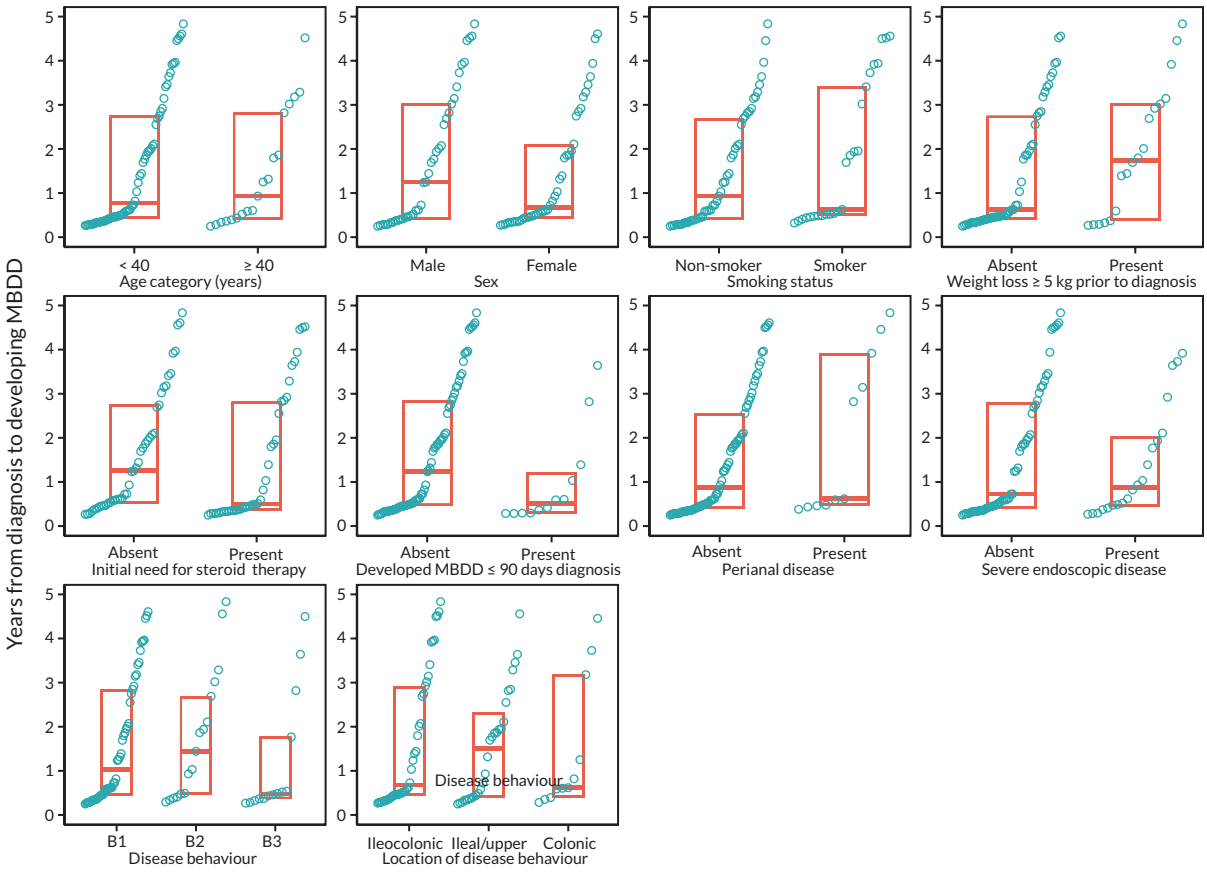

FIGURE 4 Scatter plots of years from diagnosis to developing MBDD, stratified by the prespecified clinical predictors. Markers represent patients who developed MBDD, and orange boxes represent median and IQR.

TABLE 12 Univariable HR of prespecified predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data

		Observed data			Imputed data		
Prespecified predictors		***N***	**HR (95% CI)**	***p*-value**	***N***	**HR (95% CI)**	***p*-value**
≥ 40 years of age		194	0.71 (0.42 to 1.18)	0.185			
Female		194	1.01 (0.65 to 1.57)	0.954			
Smoker		181	1.51 (0.94 to 2.45)	0.092	194	1.52 (0.95 to 2.44)	0.082
Weight loss ≥ 5 kg prior to diagnosis		171	0.89 (0.52 to 1.52)	0.659	194	0.83 (0.48 to 1.43)	0.509
Initial need for steroid therapy		194	2.11 (1.36 to 3.26)	0.001			
Developed MBDD ≤ 90 days from diagnosis		194	1.38 (0.75 to 2.54)	0.305			
Perianal disease		193	1.15 (0.61 to 2.16)	0.674	194	1.15 (0.61 to 2.18)	0.664
Severe endoscopic disease		181	1.00 (0.60 to 1.66)	0.995	194	1.04 (0.62 to 1.75)	0.869
Disease behaviour	B1	194	–	–			
	B2		1.38 (0.80 to 2.40)	0.247			
	B3		1.39 (0.77 to 2.51)	0.281			

TABLE 12 Univariable HR of prespecified predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data (*continued*)

Prespecified predictors		Observed data			Imputed data		
		N	HR (95% CI)	*p*-value	*N*	HR (95% CI)	*p*-value
Location of disease behaviour	Ileocolonic	194	–	–			
	Ileal/upper tract		0.89 (0.55 to 1.43)	0.619			
	Colonic		0.75 (0.39 to 1.45)	0.392			
Normalised global MEGS (%)		194	1.01 (0.99 to 1.02)	0.366			
Normalised global sMARIA (%)		194	1.00 (0.98 to 1.01)	0.544			
Normalised LI (%)		194	1.00 (0.99 to 1.01)	0.888			
CRP level (mg/l)		166	1.00 (1.00 to 1.01)	0.344	194	1.00 (1.00 to 1.01)	0.348
WBC count (10^9/l)		156	–	–	194	1.01 (0.94 to 1.07)	0.877
FC level (µg/g)		75	1.00 (1.00 to 1.00)	0.892	194	1.00 (1.00 to 1.00)	0.719
Haemoglobin level (g/l)		160	1.00 (0.98 to 1.01)	0.674	194	1.00 (0.98 to 1.01)	0.680
Platelet count (10^9/l)		151	–	–	194	1.00 (1.00 to 1.00)	0.762

Note
Scores were normalised to enable comparison of the scores on a standardised scale.

Source
Reproduced with permission from Taylor *et al.*[48] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The table includes minor additions and formatting changes to the original text.

lower specificity than model B (see *Table 14*). There was no benefit identified by including the MRE scores (MEGS, sMARIA, LI) or blood or stool parameters, demonstrating that inclusion of these measurements does not improve the prediction of disabling disease.

Model performance in a hypothetical 1000 participants

The performance of the various models in a hypothetical cohort of 1000 patients for RD1 (top 40% of patients with the greatest predicted risk of developing disabling disease) is shown in *Table 15*. For example, for model B, we would predict 402 to have an event within 5 years of diagnosis, of whom we would correctly predict 217, and falsely predict 185.

The performance of the various models in a hypothetical cohort of 1000 patients for RD2 (patients with an absolute risk of developing disabling disease of ≥ 10%) is shown in *Table 15*.

In *Table 16*, we present the participants who started biologic therapy within 180 days of their diagnosis of CD. There was no evidence that early biologics were preventative of MBDD.

Net benefit

An illustration of the net benefit of each model is shown in *Figures 6* and *7*.

For example, permitting one true positive (TP) to five false positives (FPs) means that to correctly identify one participant developing disabling disease, no more than five participants should be identified incorrectly (i.e. risking unnecessary top-down therapy). For this ratio to be clinically acceptable, the harm of delaying top-down therapy

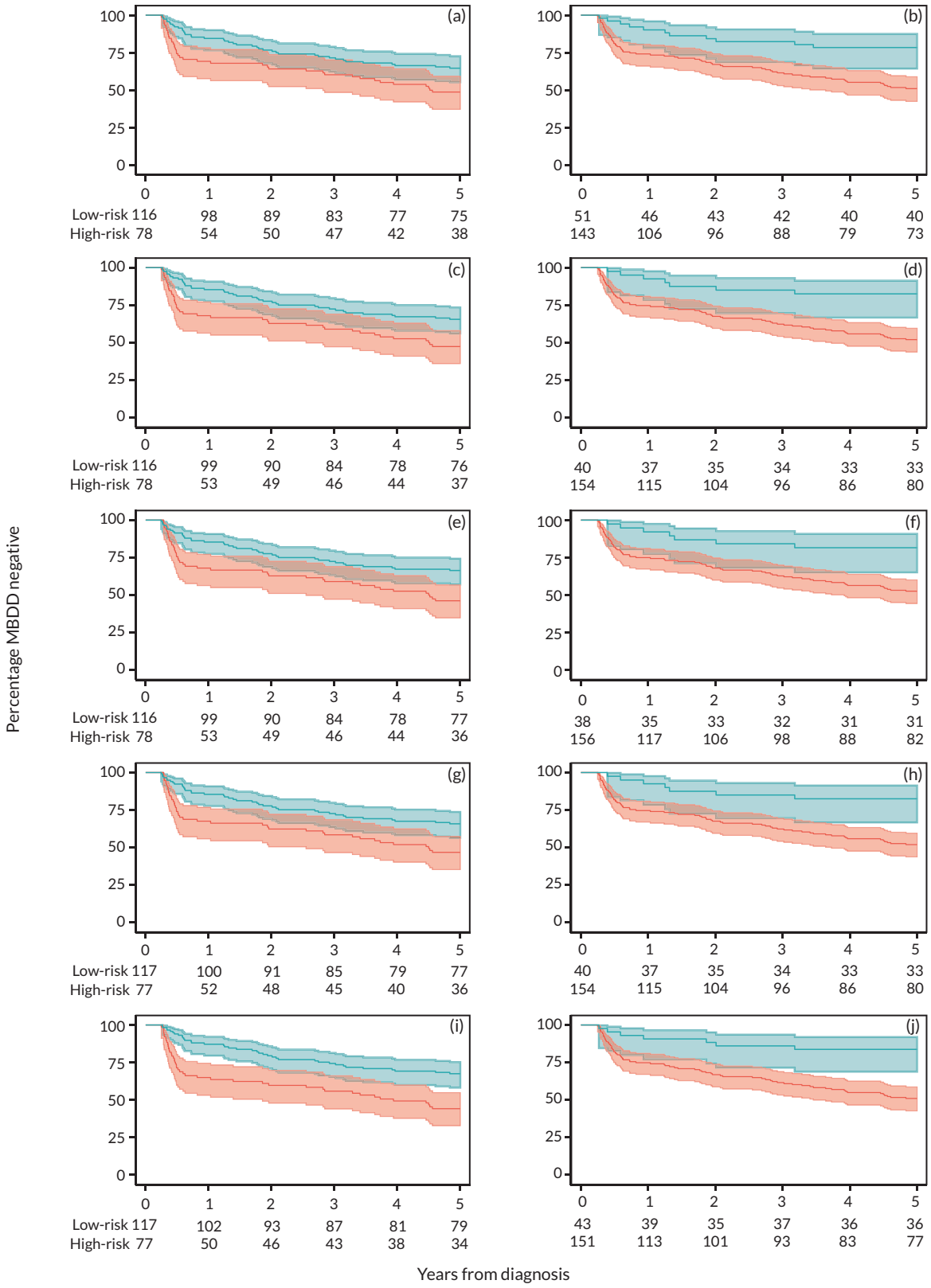

FIGURE 5 Kaplan–Meier plots of the percentage of MBDD negative participants in low-risk and high-risk groups over years from diagnosis, stratified by RD. (a) Model B and RD1; (b) Model B and RD2; (c) Model A1 and RD1; (d) Model A1 and RD2; (e) Model A2 and RD1; (f) Model A2 and RD2; (g) Model A3 and RD1; (h) Model A3 and RD2; (i) Model B1 and RD1; (j) Model B1 and RD2. Blue lines represent the low-risk group and red lines represent the high-risk group. Data are *n*.

TABLE 13 Sensitivity and specificity of prognostic models, stratified by RD

Prognostic model	RD	Risk group	Percentage MBDD negative (95% CI)	MBDD negative	MBDD positive	Sensitivity (95% CI)	Specificity (95% CI)
				N = 113	*N* = 81		
B	1	Low	65 (55 to 73)	75	41	49 (39 to 60)	66 (57 to 74)
		High	49 (38 to 60)	38	40		
	2	Low	79 (65 to 88)	40	11	86 (77 to 92)	35 (27 to 45)
		High	51 (43 to 59)	73	70		
A1	1	Low	66 (56 to 74)	76	40	51 (40 to 61)	67 (58 to 75)
		High	48 (36 to 58)	37	41		
	2	Low	83 (67 to 91)	33	7	91 (83 to 96)	29 (22 to 38)
		High	52 (44 to 60)	80	74		
A2	1	Low	67 (57 to 74)	77	39	52 (41 to 62)	68 (59 to 76)
		High	47 (35 to 57)	36	42		
	2	Low	82 (66 to 91)	31	7	91 (83 to 96)	27 (20 to 36)
		High	53 (45 to 60)	82	74		
A3	1	Low	66 (57 to 74)	77	40	51 (40 to 61)	68 (59 to 76)
		High	47 (36 to 58)	36	41		
	2	Low	83 (67 to 91)	33	7	91 (83 to 96)	29 (22 to 38)
		High	52 (44 to 60)	80	74		
B1	1	Low	68 (58 to 75)	79	38	53 (42 to 64)	70 (61 to 78)
		High	45 (33 to 55)	34	43		
	2	Low	84 (69 to 92)	36	7	91 (83 to 96)	32 (24 to 41)
		High	51 (43 to 59)	77	74		

TABLE 14 Difference in sensitivity and specificity of prognostic models using Model B as the reference, stratified by RD

Prognostic model	RD	Sensitivity (95% CI)	Sensitivity difference (95% CI)	*p*-value	Specificity (95% CI)	Specificity difference (95% CI)	*p*-value
B	1	49 (39 to 60)	–	–	66 (57 to 74)	–	–
A1		51 (40 to 61)	−1.2 (−4.9 to 2.4)	> 0.999	67 (58 to 75)	0.9 (−1.7 to 3.5)	> 0.999
A2		52 (41 to 62)	−2.5 (−9.6 to 4.7)	0.688	68 (59 to 76)	1.8 (−4.0 to 7.5)	0.727
A3		51 (40 to 61)	−1.2 (−6.7 to 4.2)	> 0.999	68 (59 to 76)	1.8 (−1.5 to 5.1)	0.500
B1		53 (42 to 64)	−3.7 (−10.3 to 2.9)	0.375	70 (61 to 78)	3.5 (−2.2 to 9.3)	0.289
B	2	86 (77 to 92)	–	–	35 (27 to 45)	–	–
A1		91 (83 to 96)	−4.9 (−10.9 to 1.0)	0.125	29 (22 to 38)	−6.2 (−11.5 to −0.9)	0.016
A2		91 (83 to 96)	−4.9 (−12.0 to 2.1)	0.219	27 (20 to 36)	−8.0 (−16.3 to 0.3)	0.064
A3		91 (83 to 96)	−4.9 (−10.9 to 1.0)	0.125	29 (22 to 38)	−6.2 (−12.2 to −0.2)	0.039
B1		91 (83 to 96)	−4.9 (−12.0 to 2.1)	0.219	32 (24 to 41)	−3.5 (−9.3 to 2.2)	0.289

TABLE 15 Number of participants correctly predicted to develop MBDD within 5 years of diagnosis in a hypothetical sample of 1000 participants, stratified by RD

Prognostic model	RD	High-risk and MBDD positive (TP)	High-risk and MBDD negative (FP)	Low-risk and MBDD positive (false-negative)	Low-risk and MBDD negative (true-negative)
B	1	206	212	196	386
A1		212	206	191	391
A2		217	201	185	397
A3		212	206	185	397
B1		222	196	175	407
B	2	361	57	376	206
A1		382	36	412	170
A2		382	36	422	160
A3		382	36	412	170
B1		382	36	397	185

TABLE 16 Number of participants who started biologic therapy < 180 days from diagnosis and developed MBDD ≥ 90 days later, stratified by maximum segmental sMARIA score

	Maximum segmental sMARIA				Global sMARIA		
	< 1	≥ 1	< 2	≥ 2	< 6	≥ 6	Total
	N = 19	*N* = 139	*N* = 31	*N* = 127	*N* = 98	*N* = 60	*N* = 158
Did not start biologic therapy and developed MBDD	4 (21)	19 (14)	5 (16)	18 (14)	17 (17)	6 (10)	23 (15)
Did not start biologic therapy and did not develop MBDD	12 (63)	51 (37)	22 (71)	41 (32)	40 (41)	23 (38)	63 (40)
Started biologic therapy < 180 days from diagnosis and developed MBDD ≥ 90 days later	1 (5)	15 (11)	1 (3)	15 (12)	10 (10)	6 (10)	16 (10)
Started biologic therapy < 180 days from diagnosis and did not develop MBDD	0 (0)	23 (17)	1 (3)	22 (17)	12 (12)	11 (18)	23 (15)
Started biologic therapy ≥ 180 days from diagnosis and developed MBDD ≥ 90 days later	0 (0)	6 (4)	0 (0)	6 (5)	5 (5)	1 (2)	6 (4)
Started biologic therapy ≥ 180 days from diagnosis and did not develop MBDD	2 (11)	25 (18)	2 (6)	25 (20)	14 (14)	13 (22)	27 (17)

for someone destined to develop disabling disease is five times greater than that of commencing unnecessary top-down therapy. In this example, the 'weight' of preventing one patient developing disabling disease is five times more important than administering top-down therapy to a patient who will not develop disabling disease.

A net benefit of 18% means that the model would lead to starting biologics in 180 participants per 1000 at risk, with all biologics truly preventing events (see *Figures 6* and *7*).

The net benefit of prognostic models for predicting development of MBDD within 5 years of diagnosis, using ratios of TP to FP predictions is provided in *Appendix 3*, *Table 26*.

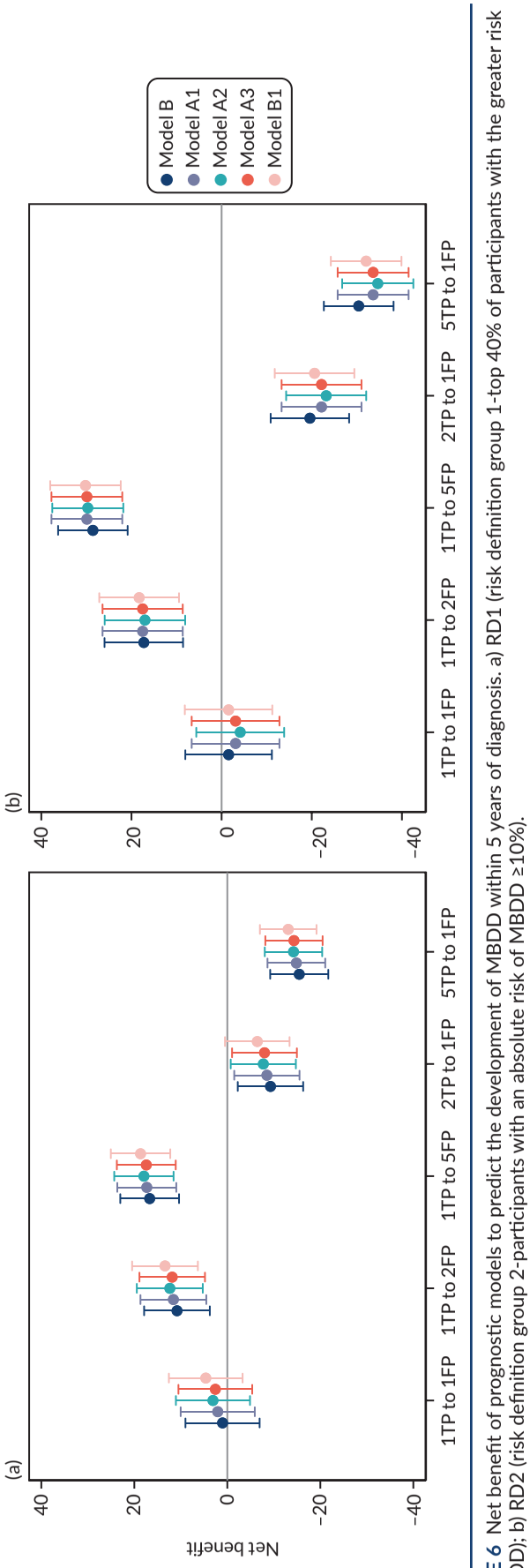

FIGURE 6 Net benefit of prognostic models to predict the development of MBDD within 5 years of diagnosis. a) RD1 (risk definition group 1-top 40% of participants with the greater risk of MBDD); b) RD2 (risk definition group 2-participants with an absolute risk of MBDD ≥10%).

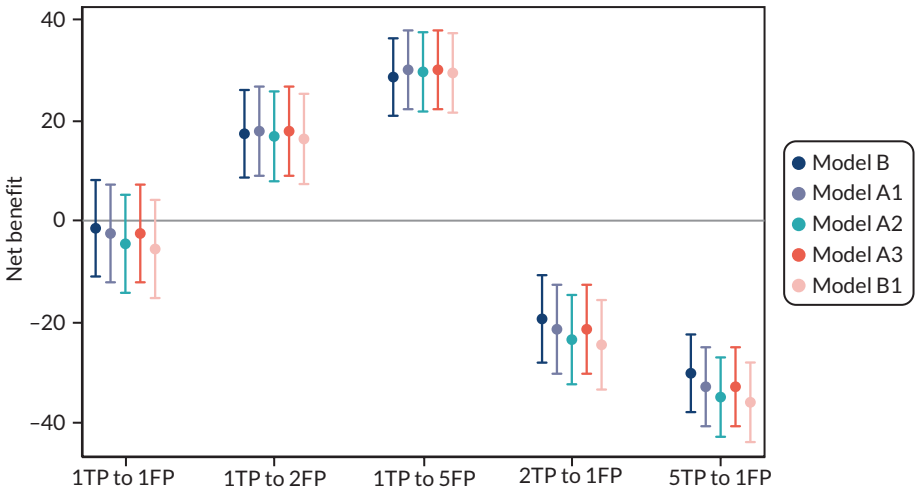

FIGURE 7 Net benefit of the various models according to the ratio of TP to FPs RD2 (patients with an absolute risk of developing disabling disease of ≥ 10%).

For secondary outcome 2, we had planned to use PCA to determine the optimal combination of individual MRE features (see *Appendix 4*, *Table 27*). These features would be used in conjunction with predefined clinical variables (Model B) to predict disabling CD. The MRE features selected (n = 20) were those used in the calculation of the sMARIA, MEGS and LI.

We computed scores for principal components (PCs) and found that the first six PCs accounted for approximately 70% of the total variance. We intended to include these PCs along with the clinical variables in a flexible parametric survival model. However, upon incorporating the PCs, the model failed to converge due to significant collinearity issues with the clinical variables. Consequently, we were unable to explore the predictive ability of individual MRE features using this approach.

Additionally, we performed PCA on a prespecified list of MRE features (n = 11) selected as of clinical interest by the study radiologists, and computed four PCs that accounted for more than 70% of the total variance. We present the scree plot (see *Appendix 5*, *Figure 9*) and loading table (see *Appendix 8*, *Table 31*). As with the full list of MRE features, the inclusion of these four PCs to the clinical variables led to collinearity problems, hindering the convergence of the model (see *Appendix 6*, *Figures 10* and *11*, *Tables 28* and *29*, *Appendix 7*, *Table 30*, *Appendix 8*, *Table 31*, *Appendix 9*, *Table 32*).

Chapter 5 Average per-patient costs incurred within 5 years of a new Crohn's disease diagnosis and baseline characteristics associated with higher costs

The secondary outcomes addressed in this chapter are:

Secondary outcome 3:

> Average per-patient national healthcare cumulative costs incurred within 5 years of newly diagnosed CD.

Secondary outcome 4:

> Patient, disease phenotype and imaging characteristics associated with higher health economic costs, within 5 years of diagnosis.

Methods

Overview of cost evaluation

All costs are calculated from an NHS perspective, specifically, including data on hospital resource use.

We estimated the healthcare costs incurred within 5 years of a new diagnosis of CD and investigated patient, imaging, treatment and other factors that drive 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. This provides valuable information regarding the costs of CD and the factors associated with these costs, which are data that are scarce currently for the NHS setting, particularly in an era where biologic therapy is administered earlier in the disease course. Mean costs per patient were also multiplied by the estimated number of UK patients, stratifying by presence or absence of disabling disease within 5 years, to estimate the UK cost of illness (COI) following diagnosis.

Cost analysis took a UK NHS and Personal Social Services perspective.[68] Costs were calculated in 2020–1 Great British pounds (£) and the time horizon was 5 years. Discounting was not applied to costs or outcomes, as this is a retrospective costing analysis and not a decision analytic model. Hence, no consideration of preferences for present benefits over future benefits is required.

Baseline patient, imaging, treatment and other factors

The baseline factors considered in the regression analyses are those from the primary modelling analysis (see *Chapter 2*) and include the nine prespecified clinical factors at diagnosis (unless otherwise specified): age, smoking status, presence of stricturing, penetrating disease or perianal disease, initial need for steroid therapy, sex, weight loss of at least 5 kg prior to diagnosis, severe endoscopic disease and location of disease. Other factors include the MRE imaging scores (LI, MEGS and sMARIA, each standardised to a 0–100 scale) and the development or not of severe/disabling disease within 5 years, using the modified Beaugerie criteria.

Cost components

Resource use data were obtained on the following cost components for every patient in the study:

1. Medications the patient received in secondary care for CD (name, dose, frequency, start date and either confirmation that the drug was ongoing at the follow-up date or the end date or the duration for which the drug was taken).
2. Imaging and endoscopic assessments related to CD, to include numbers of tests and dates of examination at diagnosis and during the follow-up period (e.g. colonoscopy, flexible sigmoidoscopy, and so on.).

3. Surgical treatments related to CD (procedure, whether day case or not, whether elective, date of procedure and date of discharge, and total length of stay).
4. Total number of outpatient visits and non-surgical day case visits related to CD over the follow-up period. Some patients had longer follow-up periods than 5 years (up to 12 years), but since we did not capture dates for these outpatient and day case visits, we instead included a pro rata proportion of the total number of visits to cover the 5-year period using the equation: number of visits over 5 years = number of visits over X years * 5/X, assuming a uniform distribution of events in accordance with discussions with clinical colleagues.
5. Inpatient stays related to CD, excluding those related to surgery recorded in item 3 above (reason for stay, specialty, dates admitted and discharged, and total length of stay).

Uniform distribution of outpatient and non-surgical day case visits

Regarding assumption 4, it was felt by the clinicians that some patients could have more appointments earlier on in their disease course and then fewer once their disease had stabilised. Conversely, other patients may have had stable CD to begin with but then subsequently have more complex disease requiring more healthcare resource input. Indeed, other patients could fall between these two extremes and have a fairly uniform distribution of appointments across their post-diagnosis period. After discussion with clinical members of the trial group, it was agreed that on average the distribution of Outpatient and day case events would be more or less uniform and there was no clear evidence towards these events being more frequent either early or late within the 5-year period to provide meaningful input parameters for sensitivity analysis. Any impact of modifying these relatively low-cost appointments would in any case have been minimal, as they accounted for only around 16% of the total cost.

Unit costs

The CRF requested information on specific scans and other assessments for imaging and endoscopy assessments. The NHS Reference Costs include costs for these assessments on various tabs (including Outpatient Procedures, Imaging, Nuclear Medicine), and unit costs were assigned on the basis of discussions with clinical colleagues according to what was the most relevant setting in this patient group in order to allow more accurate and appropriate costs.

Unit costs (*Table 17*) for imaging and endoscopic assessments were taken from the Outpatient Procedures tab (colonoscopy (FE32Z), endoscopy (FE50A), gastroscopy (FE22Z), sigmoidoscopy (FE35Z)), the Imaging tab (dexa scan (RD50Z), fluoroscopy (RD32Z), magnetic resonance cholangiopancreatography [(MRCP) RD01A], MRI (RD05Z), MRI enteroclysis (RD05Z + RD30Z), US (RD40Z), X-ray (PF), CT scans (RD24Z), barium follow-through (RD32Z) and contrast enema (RD30Z)), and the Nuclear Medicine tab [positron emission tomography (PET), RN07A] and PET-CT scans (RN01A)) of the 2020–1 NHS Reference Costs.[69]

Unit costs for outpatient visits (£168.02) used weighted means from General Surgery, General Medicine, Gastroenterology and Dietetics from the Total Outpatient Attendances tab. Unit costs for non-surgical day case visits (£632.98) used weighted means from the Day Case tab for Non-Malignant GI Tract Disorders (FD10D/F/G/H/J/K/L/M).

Unit costs for inpatient stays where accident and emergency (A&E) attendance was reported were costed as 'A&E not admitted' if length of stay was zero days, as 'A&E admitted plus non-elective short stay' if length of stay was more than zero days, and as 'Emergency Medicine, Category 1 Investigation with Category 1–2 Treatment' in either case. Unit costs for planned inpatient stays of shorter than 3 days were costed as the weighted mean of 'Day Case, IBD without interventions' if the specialty was indicated as General Medicine or Gastroenterology, and as the same but with a single intervention if the specialty was indicated as Surgery. Unit costs for planned inpatient stays of 3 days or longer were costed as the weighted mean of 'Elective Inpatient, IBD without interventions' if the specialty was indicated as General Medicine or Gastroenterology, and as the same but with a single intervention if the specialty was indicated as Surgery. Trim points and excess bed day costs from the 2023–4 NHS Reference Costs were included to allow additional cost for the longer stays (*Table 18*).

Unit costs for surgeries were costed as 'day case, elective' or 'Non-elective short' (length of stay < 3 days) or 'non-elective long stay' (NEL) (more than 3 days), according to those variables (see *Table 18*) using similar methods as those described above for inpatient stays.

TABLE 17 Unit costs for imaging and outpatient procedures from NHS Reference Costs 2020–1

Imaging and outpatient procedures	Unit costs (£)
X-ray	57.78
US	106.26
DEXA scan	131.18
CT scan	152.64
Contrast enema	187.52
Endoscopy	189.37
MRCP	211.24
MRI	245.37
Fluoroscopy	336.59
Barium follow-through	336.59
Colonoscopy	345.55
MRI enteroclysis	432.89
Sigmoidoscopy	499.32
Gastroscopy	615.22
PET scan	697.61
PET-CT scan	1127.81

TABLE 18 Unit costs for inpatient stays and surgeries from NHS Reference Costs 2020–1

NHS Reference Cost category and assumptions	Unit cost (£)	Specialty according to study data	Which page on CRF
AE not admitted	124.00	A&E	Page 8
AE admitted + NES IBD weighted mean no interventions	867.30	A&E plus NES stay	Page 8
DC IBD weighted mean no interventions	388.94	General medicine/gastroenterology	Page 8
DC IBD weighted mean single intervention	2493.69	Surgery	Page 6 and 8
EL IBD weighted mean no interventions (trim point 5 days, excess bed-day cost £298)	1780.48	General medicine/gastroenterology/ haematology	Page 8
EL IBD weighted mean single intervention (trim point 17 days, excess bed-day cost £298)	6025.45	Surgery	Page 6 and 8
NES IBD weighted mean single intervention	741.30	Surgery	Page 6
NEL IBD weighted mean single intervention (trim point 17 days, excess bed-day cost £298)	4991.87	Surgery	Page 6

AE, A&E tab; DC, Day Case tab; EL, elective inpatient stay tab; NES, non-elective short stay tab.

Unit costs were multiplied by resource use collected in the trial. For CD medication costs, we used the cheapest available Drug Tariff cost listed in the online *British National Formulary* unless a specific brand name was specified, in which case that brand was costed. If drug tariff prices were not available and a brand name was not specified, we used the lowest available NHS Indicative Price instead. Costs of whole packets were used. There were very little missing data, which were left as missing unless there was an obvious suggestion in another free text variable regarding reasonable or commonly indicated doses or routes of administration to be imputed.

Patient-level data capture and time horizon

No directly identifiable patient information was used for this analysis.

Data were included for costing from the date of diagnosis and for the following 5 years. Patients with shorter than 5 years of follow-up were omitted. Patients with longer than 5 years of follow-up had their resource use data truncated at the 5-year point. Information from before the diagnosis date was also removed.

Medications and other treatments, procedures and inpatient stays that began before the date of diagnosis were truncated at the date of diagnosis (i.e. exclude earlier period), and those that continued after the 5-year date were also truncated (i.e. exclude later period).

Regression analysis

Regression analysis methods were used to investigate patient, disease phenotype and imaging characteristics associated with higher health economic costs within 5 years of diagnosis. Unadjusted 5-year costs were calculated separately according to the presence versus absence of severe/disabling CD, treatments received and patient demographic characteristics.

Multivariable regression was used to identify factors (CD status, MRE parameters, treatments received, patient characteristics) associated with higher costs, in accordance with the choice of variables from the risk model.

To account for skewed cost data, a generalised linear model with gamma family and log-link was used, experimenting with other distributional assumptions (log-normal, Gaussian, inverse Gaussian and negative binomial distributions), and the best fit was selected as judged by residual plots and the Akaike information criterion (AIC). The BIC was also examined to assess agreement between AIC and BIC regarding the best-fit model and the most statistically important cost drivers, although the BIC can have a tendency to under-fit.

Calculating the cost of illness in recently diagnosed patients

Mean costs per patient were multiplied by the estimated number of patients in the UK with CD, stratifying by presence or absence of disabling disease according to the proportions of these two groups calculated during our analysis, to estimate the COI for the 5 years following diagnosis in the UK. COI studies can be based on prevalence or incidence, depending on the research question and data availability. Prevalence-based approaches allow calculation of costs over a fixed period of time, for example, 5 years, according to the number of people in the region or country known to currently be suffering from the disease. Incidence-based approaches instead aim to calculate lifetime costs for patients, using yearly incidence rates to provide numbers of new patients joining the population of interest each year.[70] In the context of this analysis, we have calculated the 5-year prevalence-based COI, and focus on those patients who are within the first 5 years of their diagnosis.

Results

Resource use and costs covering 5 years from diagnosis

Table 19 shows unadjusted mean (standard deviation) and median (IQR) total per-patient costs of health care for the categories of costs as collected in the study CRF. Mean (SD) total costs per patient were £24,267 (£33,108) for the

TABLE 19 Mean (SD) unadjusted and median (IQR) CD management costs for newly diagnosed patients, total costs per patient for the 5 years from diagnosis

	n (%)		Mean	SD	Median	IQR – lower	IQR – upper
All patients at 5 years							
Total unadjusted costs over 5 years from diagnosis, £							
Drug costs	146	(75)	21,592	25,741	8988	913	37,844
Drug costs – biologics	87	(45)	34,051	25,568	32,961	13,271	41,553
Drug costs – non biologics	143	(74)	1328	1905	562	225	1794
Inpatient stays	62	(32)	9288	27,765	4500	2078	7481
Surgeries	41	(21)	8999	10,238	6025	4992	8227
Non-surgical day cases	30	(15)	1860	3487	643	482	1130
Outpatient visits	150	(77)	2124	941	1997	1516	2666
Imaging	151	(78)	1564	1063	1427	736	2116
Total costs	**194**	(100)	**24,267**	**33,108**	**10,136**	**1856**	**41,092**
Patients with disabling disease at 5 years (modified Beaugerie)							
Total unadjusted costs over 5 years from diagnosis, £							
Drug costs	57	(70)	25,429	23,718	20,220	3323	38,007
Drug costs – biologics	43	(53)	31,566	22,657	32,093	12,677	40,842
Drug costs – non biologics	56	(69)	1645	2438	773	280	2028
Inpatient stays	43	(53)	10,994	33,138	4553	2078	7481
Surgeries	25	(31)	8896	9919	6025	6025	8227
Non-surgical day cases	9	(11)	950	695	581	547	1382
Outpatient visits	58	(72)	2490	918	2451	1886	2891
Imaging	59	(73)	1919	1112	1646	1095	2402
Total costs	**81**	(100)	**29,763**	**38,278**	**17,352**	**.0**	**45,969**
Patients without disabling disease at 5 years (modified Beaugerie)							
Total unadjusted costs over 5 years from diagnosis, £							
Drug costs	89	(79)	19,135	26,798	3620	555	33,307
Drug costs – biologics	44	(39)	36,480	28,175	33,640	19,015	42,728
Drug costs – not biologics	87	(77)	1125	1444	453	176	1292
Inpatient stays	19	(17)	5427	5236	2922	2078	7813
Surgeries	16	(14)	9160	11,046	6025	4992	8521
Non-surgical day cases	21	(19)	2250	4112	684	468	886
Outpatient visits	92	(81)	1893	884	1693	1209	2262
Imaging	92	(81)	1336	971	1118	571	1847
Total costs	**113**	(100)	**20,327**	**28,368**	**7293**	**2495**	**30,976**

Note
Costs are in 2020–1 Great British pounds (£).

whole group. When split according to whether they had developed severe/disabling disease by 5 years, according to the modified Beaugerie criteria, costs were £29,763 (SD £38,278) for those with disabling disease and £20,327 (SD £28,368) for those without, over the 5-year period.

When considering descriptive statistics of cost breakdowns (note, formal statistical comparisons were not made of the breakdowns), the largest cost contributor was biologics, which manifested as a greater proportion of patients receiving these more expensive therapies: 43/81 (53%) of those patients developing disabling disease within 5 years received biologics, compared with 44/113 (39%) of patients who did not develop disabling disease. This led to a higher mean drug cost in the disabling disease group. Another difference was inpatient stays: a greater proportion of patients developing severe/disabling disease had these (53% vs. 17%), resulting in a higher mean cost per patient (£10,994 vs. £5427). Of note, however, hospitalisation is part of the calculation of the modified Beaugerie criteria so will likely have influenced this observation. Similarly, patients who developed disabling disease underwent more surgeries than those who did not (31% vs. 14%). Numbers of non-surgical day case visits, outpatient visits and imaging events appeared slightly lower in the group without disabling disease, but the differences were small.

Appendix 10, *Table 33* shows unadjusted mean (SD) 5-year costs calculated separately for the nine clinical categories specified above, and for presence versus absence of severe/disabling CD (modified Beaugerie criteria).

Regression analysis

Results of separate unadjusted regressions using each of the variables listed above are given in *Table 20*. According to these unadjusted models, the following factors were associated with driving higher costs over the first 5 years from diagnosis: age under 40 years at diagnosis; perianal disease at diagnosis; severe (ileocolonic) endoscopic disease at diagnosis.

TABLE 20 Separate unadjusted regressions using the preselected covariates as independent variables associated with 5-year total costs

	n	Mean	SD	*p*-value	95% CI
sMARIA (continuous)					
Raw	194	0.035	0.023	0.125	−0.010 to 0.079
Standardised	194	0.010	0.006	0.125	−0.003 to 0.022
LI (continuous)					
Raw	194	0.030	0.041	0.454	−0.049 to 0.110
Standardised	194	0.005	0.007	0.454	−0.008 to 0.018
MEGS (continuous)					
Raw	194	0.006	0.005	0.204	−0.003 to 0.016
Standardised	194	0.009	0.007	0.204	−0.005 to 0.023
Beaugerie – no disabling disease at 5 years					
Beaugerie – yes disabling disease at 5 years	194	0.381	0.197	0.053	−0.005 to 0.767
Age – under 40 years at diagnosis					
Age – over 40 years at diagnosis	194	−0.566	0.214	0.008*	−0.986 to −0.146
Sex – male					
Sex – female	194	0.272	0.194	0.160	−0.107 to 0.652
Smoking status – not smoker at diagnosis					
Smoking status – smoker at diagnosis	177	0.210	0.212	0.320	−0.204 to 0.625

TABLE 20 Separate unadjusted regressions using the preselected covariates as independent variables associated with 5-year total costs (*continued*)

	n	Mean	SD	*p*-value	95% CI
Weight loss – no					
Weight loss – recent weight loss at diagnosis	168	0.248	0.216	0.249	−0.174 to 0.671
Disease behaviour – not stricturing or penetrating					
Disease behaviour – stricturing	194	0.417	0.272	0.125	−0.115 to 0.950
Disease behaviour – penetrating	194	0.425	0.286	0.137	−0.135 to 0.985
Perianal disease – absent					
Perianal disease – present	193	0.698	0.313	0.026*	0.085 to 1.310
Severe endoscopic disease – absent					
Severe endoscopic disease – present	177	0.510	0.240	0.034*	0.039 to 0.981
Location of endoscopic disease – ileocolonic					
Location of endoscopic disease – ileal	194	−0.674	0.207	0.001*	−1.079 to −0.269
Location of endoscopic disease – colonic	194	−0.346	0.265	0.192	−0.867 to 0.174
Location of endoscopic disease – upper	194	−0.489	0.654	0.454	−1.771 to 0.793
Need for steroid therapy – no					
Need for steroid therapy – yes	194	0.371	0.193	0.054	−0.007 to 0.749

*statistically significant.

The best-fit regression according to the AIC and residual plot for modelling the total 5-year healthcare costs in this patient group included the modified Beaugerie criteria for severe/disabling disease within 5 years (yes or no), age at diagnosis (split into two categories: above or below 40 years of age), sex (male or female), smoker at diagnosis (yes or no), recent weight loss prior to diagnosis (< 5 kg or >= 5 kg), behaviour of disease at diagnosis (stricturing, penetrating or neither), presence of perianal disease at diagnosis, severe endoscopic disease at diagnosis, disease location at diagnosis (ileocolonic, ileal, colonic or upper digestive tract), and need for initial steroid therapy as well as two of the three MRE indices (LI and sMARIA). Removal of the MEGS improved the fit. The model used gamma family with log-link. The coefficients and other details for this model are given in *Appendix 10*, *Table 34*.

A second model with fewer covariates gave the best fit according to the BIC, although use of this summary value can be associated with under-fitting, and indeed the residuals plot was less good with this model. However, we note it here as it corroborates the statistically significant covariates identified in the best-fit regression model (see *Appendix 10*, *Table 35*).

Prevalence-based cost-of-illness calculation

The prevalence of CD is 106 per million people in the UK.[71] The Office for National Statistics population estimates the UK population at mid-year 2021 to be 67.0 million.

Cost of illness for those developing disabling disease within 5 years:

Our analysis suggests 41.75% of newly diagnosed patients develop disabling disease by 5 years. Using the prevalence estimate of 106 per million patients, this equates to 106/100 * 41.75 = 44.26 people per million developing disabling disease within 5 years. This corresponds to 44.26 * 67.0 = 2965.1 people in total for the UK. The mean (SD) 5-year cost for those who have developed disabling disease is £29,763 (£38,278). So, over a 5-year period, the cost would be £29,763 * 2965.1 = £88,250,271, using the assumptions listed above.

Cost of illness for those not developing disabling disease within 5 years

Our analysis suggests 58.25% of newly diagnosed patients do not develop disabling disease within 5 years of diagnosis. Using the CD prevalence estimate of 106 per million people, this means that 106/100 * 58.25 = 61.75 * 58.25 = 61.75 people per million are CD patients who have not developed disabling disease within 5 years. This corresponds to 61.75 * 67.0 = 4137 people in total for the UK. The mean (SD) 5-year cost for those who have not developed severe/disabling disease is £20,327 (£28,368). Over a 5-year period, the total cost would therefore be £20,327 * 4137 = £84,090,766, when using the assumptions listed above.

It is important to note that performing a COI study does not provide information regarding whether resources are well spent. COI estimates are intended to illustrate which baseline patient and disease characteristics lead to higher costs as a marker for greater care needs. Future work could include this cost information in a health economic analysis that incorporates clinical outcomes and makes comparisons between different patient pathways.[72]

Chapter 6 Discussion

Main trial

Cross-sectional imaging, especially MRE, plays a pivotal role in assessing CD, including at the time of initial diagnosis, and is advocated by international guidelines.[3,73,74] This makes MRE a particularly attractive proposition as a prognostic tool, that is to predict which newly diagnosed CD patients are destined to develop subsequent severe/disabling disease. If accurate, this would facilitate timely biological therapy in these patients while avoiding unnecessary administration to others, mitigating potential side effects and definite costs. Notwithstanding, few studies have evaluated cross-sectional imaging as a prognostic tool.[43] Accordingly, the present trial exploited METRIC participants by extending follow-up to a minimum of 4 years for the newly diagnosed cohort, as well as identifying a matched retrospective group of patients who were also newly diagnosed with CD to increase our sample size.[1,35,49]

A priori, we hypothesised that either active inflammation or established bowel damage on baseline MRE, as measured by sMARIA, MEGS or the LI, would help prognosticate disabling CD as part of a multivariable model that also incorporated standard clinical predictors. However, we found that no MRE scoring system predicted disabling disease when added to the standard baseline model. In univariable analysis, initial requirement for steroid therapy was the only clinical predictor for the subsequent disabling CD that was statistically significant.

The sMARIA quantifies CD activity, disease severity, and treatment response, and has been externally validated against a range of reference standards.[36,38,56,75,76] It has excellent performance characteristics compared with endoscopy, although its specificity is lower when compared to histological reference standards.[38] The MEGS comprises both imaging markers of active inflammatory disease and established bowel damage, and has also been tested extensively against multiple reference standards.[52–54,77] The LI differs from both the sMARIA and MEGS by providing an assessment of structural bowel damage exclusively and omits the extent of disease activity and mucosal inflammation.[58] MRE identifies abnormalities that persist in intestinal segments even after endoscopic remission of CD has occurred, implying that intestinal damage is established.[78] Such MRE findings include persistent mural thickening, mural fat deposition, creeping fat, and strictures. The relevance of these findings, especially in the context of future disease outcomes, remains unclear. Our findings imply that neither active inflammation nor bowel damage measured by sMARIA, MEGS and LI from MRE parameters predict disabling CD as defined by the modified Beaugerie criteria.

That we did not find an association between MRE and subsequent development of disabling disease may reflect inherent limitations of the definition we used to define disabling disease, rather than a lack of predictive capability. There is no universally accepted disease severity classification or validated definition for severe CD, with substantial literature heterogeneity. Clarification represents a major clinical and research need.[6] Therefore, we were obliged to employ a range of definitions for the end point of disabling disease. We adopted a modified version of the relatively inclusive Beaugerie description,[25] removing the variable of commencing biologic therapy as a definition for disabling disease, a decision that we took a priori after much consideration. Our rationale for this modification was that a 'top-down' approach with early, aggressive treatment to minimise disease progression is being increasingly used and likely to increasingly become the standard of care following the results of the PROFILE trial; thus being prescribed biologic therapy is not necessarily an indicator of severe disease in itself, but rather a desire to prevent it.[19,20] We considered the possibility that patients in our cohort who were treated aggressively with biologics and immunomodulators at diagnosis may have diminished the proportion who developed severe CD subsequently but we found no evidence to support this. Furthermore, many of the Beaugerie events occurred in the first 90 days, largely driven by hospitalisation. Since patients who fulfil the criteria immediately may not benefit from model predictions as there is little time to initiate top-down treatment, we decided to remove the Beaugerie events that occurred ≤ 90 days from diagnosis. This decision, made post hoc, was made to ensure that the predictive model is clinically useful for those patients destined to develop disabling CD and relevant in clinical practice as top-down treatment becomes the standard of care.[19]

Other studies have considered the potential prognostic role of MRE for CD outcome, although not specifically in newly diagnosed cohorts. Fiorino *et al.* studied 142 patients in a dual-centre prospective study.[45] Using univariate analysis,

they found that bowel damage (defined as intestinal strictures, fistulae or abscesses) was associated with a significantly higher risk of hospitalisation and surgery during a median follow-up period of just under 5 years. Similarly, the LI was an independent predictor for disease progression and need for subsequent surgery. Patients were eligible for recruitment if MRE was performed within 2 years of diagnosis. In contrast, in our study, patients were imaged at initial diagnosis. We found MRE had no predictive potential when applying modified Beaugerie criteria. Similarly, and again using univariate analysis, a single-centre study comprising 112 patients with relapsed rather than newly diagnosed CD also found that established bowel damage on MRE predicted future surgical resection.[44] Another single-centre study of 52 patients with CD that did not distinguish between newly diagnosed and established CD reported that restricted diffusion, increased upstream dilatation from a stricture, complex fistula, perienteric inflammation, fibrofatty proliferation and increased length of disease involvement on outpatient MRE were significantly more common in patients that underwent subsequent surgery.[47] However, it is unclear if these findings were also statistically significant in univariate analysis or if this was a misinterpretation of a multivariable model. It is also possible that our results are at odds with these studies because our cohort was exclusively newly diagnosed patients, who are therefore earlier in the disease trajectory than later diagnoses.[44] Additionally, we did not investigate surgery as a standalone definition of severe disease nor initiation of biologic therapy.

Our work has key strengths. To the best of our knowledge, this is the first study to investigate MRI as a predictive biomarker for the development of severe/disabling CD that exclusively considers newly diagnosed patients without severe disease at baseline. Generalisability was enhanced by a large cohort recruited from multiple NHS hospitals with multiple radiologists scoring MRE examinations, thereby closely reflecting routine clinical practice.

There are also noteworthy limitations. Despite US being a component of the METRIC trial, we could not evaluate it presently as only static images rather than cine loops were available, and these were not of sufficient quality for post-hoc analysis.[79] A priori, if there were sufficient patients with events, we had intended to also employ the more stringent Liège[24] and Montreal behaviour criteria as end points for severe/disabling disease.[61] However, the event rate was too low for developing reliable predictive models, so this analysis was not undertaken. The lower-than-expected Beaugerie event rate (42% vs. 50% to 60% anticipated) may also reflect greater upfront use of biologics, thereby reducing the prevalence reported in older literature. We were unable to assess interobserver variation in the interpretation of MRE across the different sites, although this has been studied extensively,[80–83] and the multicentre nature of this study is an inherent strength as it more closely reflects clinical practice.[38] Owing to the challenges posed by the COVID-19 pandemic, we had to reduce the original target recruitment from 167 to 131 in the prospective METRIC cohort, with a corresponding increased target of 76 patients from the retrospective cohort. Finally, recruitment was mainly from 3 centres.

Average per-patient costs incurred within 5 years of a new Crohn's disease diagnosis and baseline characteristics associated with higher costs

We calculated mean and median 5-year costs for newly diagnosed patients. Mean (SD) 5-year per-patient costs were £24,267 (£33,108) for the entire cohort, £29,763 (SD £38,278) for those who went on to develop severe/disabling disease, and £20,327 (SD £28,368) for those who did not. The main cost drivers were development of disabling/severe CD within 5 years of diagnosis (as per the modified Beaugerie criteria), age at diagnosis (above or below 40 years of age), sex, presence of perianal disease at diagnosis, and disease location at diagnosis (ileocolonic, ileal, colonic or upper digestive tract). Mean costs were increased by developing disabling/severe CD within 5 years, being female, perianal disease at diagnosis and having ileocolonic disease instead of ileal, colonic or upper digestive tract. Of note, there was no association between MRE scores at diagnosis and 5-year patient costs.

Our analysis has some limitations. Firstly, resource use information came mostly from hospital records, which are not optimised for research, and so certain assumptions were required around which unit costs to apply to events. We did our best to mitigate this through internal discussion, and weighted means of a variety of costs were used where required. A further limitation of this type of observational data was that primary and community care health and social services are excluded. Prevalence values came originally from Thompson *et al.* using figures provided by primary care practices in England and Wales in 1991–2, as there do not appear to be any more recent estimates of UK CD prevalence.[84]

Patient and public involvement

Both the METRIC trial and the present study were developed in collaboration with Crohn's and Colitis UK. A patient representative played a pivotal role during the inception and undertaking of the METRIC trial and this has continued to the present work. Their input towards refining the research questions, protocol development and funding application have been gratefully received. They have sat on the Trial Management Committee and Trial Steering Committee throughout the trial period, advising on multiple aspects, including recruitment strategies, and we look forward to continuing this collaboration to maximise the study's findings through patient groups.

Equality, diversity and inclusion

We recruited patients from NHS hospitals across the country, which ensured a diverse and representative population. Our recruitment criteria were inclusive, although participants who could not tolerate MRE were not included. Future research could focus on US which is often better tolerated than MRE. Basic demographic data presented in this report shows good participant representation according to age and gender. We thus anticipate the results will be applicable to the wider patient population. We reduced the burden on participants by not requiring additional hospital attendance or investigations. The investigators, including the reading radiologists, represent diverse ethnic backgrounds. Junior researchers were included in the study team, including the primary write-up of the research. All the reading radiologists are affiliated to the British Society of Gastrointestinal and Abdominal Radiology. The executive committee of the Society has drawn up a manifesto on diversity, equity and inclusion emphasising its importance.[85]

Impact and learning

Our work suggests that MRE does not have a role in prognosticating CD at the time of initial diagnosis, thus other tools deserve investigation. Notwithstanding, further analyses using alternative end-points such as the need for surgery and different definitions for disabling disease should be undertaken.

Implications for decision makers

Despite the continued importance of MRE as a first-line investigation for diagnosing and monitoring CD, our research suggests that it does not prognosticate at the time of initial diagnosis. A prognostic tool at the time of initial diagnosis of CD remains an unmet clinical need and should be a research priority.

Research recommendations

- An updated classification system for defining disease severity/disabling disease is needed to reflect the current treatment paradigms in CD and to include the perspectives of patients.
- Although MRE does not aid prognostication of disabling disease according to the modified Beaugerie criteria, the predictive ability of MRE against alternative definitions for disabling CD should be investigated.
- How findings of MRE at the time of diagnosis influence management deserves investigation.
- The utility of US observations at diagnosis for prognostication should be explored as a better tolerated alternative to MRE.

Conclusions

In an NHS setting, the addition of MRE activity/bowel damage scores to a multivariable model comprising existing clinical predictors did not improve the prediction of disabling CD using modified Beaugerie criteria. We recognise that predicting outcomes based on imaging is likely to be impaired as the results of imaging alter treatment for the better with identification of severe disease leading to more effective treatment, and thereby better outcomes. Notwithstanding our findings, MRE remains an essential tool for diagnosing and monitoring CD.

Additional information

CRediT contribution statement

Shankar Kumar (https://orcid.org/0000-0002-5945-8791): Conceptualisation, Data curation, Investigation, Methodology, Visualisation, Writing – original draft, Writing – editing and reviewing.

Andrew Plumb (https://orcid.org/0000-0003-1322-5113): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing – editing and reviewing.

Sue Mallett (https://orcid.org/0000-0002-0596-8200): Conceptualisation, Formal analysis, Funding acquisition, Methodology, Methodology, Supervision, Writing – editing and reviewing.

Caroline Clarke (https://orcid.org/0000-0002-4676-1257): Formal analysis, Methodology, Supervision, Visualisation, Writing – original draft, Writing – editing and reviewing.

Tom Parry (https://orcid.org/0000-0001-8855-0661): Formal analysis, Visualisation, Writing – editing and reviewing.

Jing Yi Jessica Weng (https://orcid.org/0000-0002-2875-4029): Formal analysis, Visualisation, Writing – original draft, Writing – editing and reviewing.

Gauraang Bhatnagar (https://orcid.org/0000-0001-6607-1117): Data curation, Investigation, Writing – editing and reviewing.

Stuart Bloom (https://orcid.org/0009-0002-5027-0765): Conceptualisation, Data curation, Funding acquisition, Investigation, Writing – editing and reviewing.

John Hamlin (https://orcid.org/0000-0001-8080-1358): Conceptualisation, Data curation, Funding acquisition, Investigation, Writing – editing and reviewing.

Ailsa Hart (https://orcid.org/0000-0002-7141-6076): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Writing – editing and reviewing.

Simon Travis (https://orcid.org/0000-0002-2690-4361): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Writing – editing and reviewing.

Roser Vega (https://orcid.org/0000-0002-0116-7211): Data curation, Investigation, Writing – editing and reviewing.

Maira Hameed (https://orcid.org/0000-0002-6290-5695): Data curation, Investigation, Writing – editing and reviewing.

Anisha Bhagwanani (https://orcid.org/0000-0001-6112-5171): Data curation, Investigation, Writing – editing and reviewing.

Rebecca Greenhalgh (https://orcid.org/0009-0001-5345-3535): Data curation, Investigation, Writing – editing and reviewing.

Emma Helbren (https://orcid.org/0000-0001-9524-0055): Data curation, Investigation, Writing – editing and reviewing.

James A Stephenson (https://orcid.org/0000-0001-5649-6322): Data curation, Investigation, Writing – editing and reviewing.

Ian Zealley (https://orcid.org/0000-0001-8776-9071): Data curation, Investigation, Writing – editing and reviewing.

Vivienne Eze (https://orcid.org/0000-0001-8298-9614): Data curation, Investigation, Writing – editing and reviewing.

James Franklin (https://orcid.org/0000-0002-0316-3177): Data curation, Investigation, Writing – editing and reviewing.

Alison Corr (https://orcid.org/0000-0002-6902-7362): Data curation, Investigation, Writing – editing and reviewing.

Arun Gupta (https://orcid.org/0000-0003-0311-4284): Data curation, Investigation, Writing – editing and reviewing.

Elizabeth Isaac (https://orcid.org/0000-0002-4928-0870): Data curation, Investigation, Project administration, Writing – editing and reviewing.

Damian Tolan (https://orcid.org/0000-0001-9895-9874): Data curation, Investigation, Writing – editing and reviewing.

William Hogg (https://orcid.org/0009-0007-6548-0892): Data curation, Investigation, Writing – editing and reviewing.

Antony Higginson (https://orcid.org/0000-0002-9690-5394): Data curation, Investigation, Writing – editing and reviewing.

Michela Cicchetti (https://orcid.org/0009-0006-7206-3643): Data curation, Investigation, Writing – editing and reviewing.

Sunita Gupta (https://orcid.org/0000-0001-8657-7932): Data curation, Investigation, Writing – editing and reviewing.

Miguel Serran (https://orcid.org/0009-0008-8758-3291): Data curation, Investigation, Writing – editing and reviewing.

Tim Raine (https://orcid.org/0000-0002-5855-9873): Data curation, Investigation, Methodology, Writing – editing and reviewing.

Mohamed Ahmed (https://orcid.org/0000-0001-9329-6063): Data curation, Investigation, Writing – editing and reviewing.

Biljana Brezina (https://orcid.org/0000-0001-5060-5604): Data curation, Investigation, Writing – editing and reviewing.

Ilse Patterson (https://orcid.org/0000-0002-4092-4553): Data curation, Writing – editing and reviewing.

Louise Lee (https://orcid.org/0000-0002-6200-3069): Data curation, Investigation, Writing – editing and reviewing.

Richard Pollok (https://orcid.org/0000-0001-6452-6763): Data curation, Investigation, Writing – editing and reviewing.

Jaymin Patel (https://orcid.org/0009-0009-2886-8286): Data curation, Investigation, Writing – editing and reviewing.

Abigail Seward (https://orcid.org/0009-0004-6685-185X): Data curation, Writing – editing and reviewing.

Samantha Baillie (https://orcid.org/0000-0003-3280-0347): Data curation, Investigation, Writing – editing and reviewing.

Kashfia Chowdary (https://orcid.org/0000-0002-8185-5152): Data curation, Investigation, Writing – editing and reviewing.

Sue Philpott (https://orcid.org/0000-0002-1019-5871): Data curation, Investigation, Project administration, Writing – editing and reviewing.

Anvi Wadke (https://orcid.org/0009-0009-6846-1034): Data curation, Investigation, Project administration, Writing – editing and reviewing.

Steve Halligan (https://orcid.org/0000-0003-0632-5108): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft, Writing – editing and reviewing.

Stuart A Taylor (https://orcid.org/0000-0002-6765-8806): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft, Writing – editing and reviewing.

Acknowledgements

This project was supported by researchers at the NIHR UCLH Biomedical Research Centre. Thank you to Valentina Raspa and Kadri Sudarshan.

Clinical Trial Unit

Comprehensive Clinical Trials Unit at UCL

We are grateful to Nicola Gibbons (Radiographer), Bindiya Kerai (Research Radiographer), Mary Lucas (Research Coordinator), Jean Wilson (Research Nurse), Joanne Wormleighton (Research Radiographer), Katherine Coll (Research and Development Administrator), Ilan Jacobs (patient representative), Teresita Beeston (research nurse), Doris Quartey (research nurse), Claire Ward (research practitioner), Valentina Raspa (research nurse), Kadri Sudarshan (consultant gastroenterologist), Jacqueline Patterson (consultant gastroenterologist).

Trial Steering Committee

Vicky Goh, James Lindsay and Andrea Marshall.

Patient data statement

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data are 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 they are 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

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.

Ethics statement

The METRIC-EF study achieved NHS Research Ethics Committee (NHS REC), London – Hampstead Research Ethics Committee approval on 26 October 2018 (IRAS 217422) and was conducted in accordance with the principles of good clinical practice. UCL's CCTU supervised the study.

Information governance statement

University College London is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (EU GDPR) 2016/679. Under the Data Protection legislation, UCL is the Data Controller, and you can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for our Data Protection Officer here www.ucl.ac.uk/data-protection/data-protection-0.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/THSN9956.

Primary conflicts of interest: Andrew Plumb – NIHR HTA including grant support for the conduct of this study, grants from NIHR Fellowships programme; and personal fees from Acelity, Actavis, Dr Falk, Janssen-Cilag, and Takeda.

Sue Mallett – Grant support from the NIHR HTA for the present manuscript.

Stuart Bloom – Grant support from the NIHR HTA for the present manuscript.

John Hamlin – Grant support from the NIHR HTA for the present manuscript.

Ailsa Hart – Grant support from the NIHR HTA for the present manuscript.

Simon Travis – Grants/Research Support: Grant support from the NIHR HTA for the present manuscript, AbbVie, Buhlmann, Celgene, ECCO, Helmsley Trust, IOIBD, Janssen, Lilly, NIHR HTA including for the present manuscript. Pfizer, Takeda, UCB, UKIERI, Vifor, and Norman Collisson Foundation. Consulting Fees: Abacus; AbbVie; Actial; ai4gi; Alcimed; Allergan; Amgen; Apexian; Aptel; Arena; Asahi; Aspen; Astellas; Atlantic; AstraZeneca; Barco; Biocare; Biogen; BLPharma; Boehringer Ingelheim; BMS; Buhlmann; Calcico; Celgene; Cellerix; Cerimon; ChemoCentryx; Chiesi; CisBio; ComCast; Coronado; Cosmo; Ducentis; Dynavax; Elan; Enterome; EQrX; Equillium; Falk; Ferring; FPRT Bio; Galapagos; Genentech/Roche; Genzyme; Gilead; Glenmark; Grunenthal; GSK; GW Pharmaceuticals; Immunocore; Immunometabolism; Indigo; Janssen; Lexicon; Lilly; Medarex; Medtrix; Merck; Merrimack; Mestag; Millenium; Neovacs; Novartis; Novo Nordisk; NPS-Nycomed; Ocera; Optima; Origin; Otsuka; Palau; Pentax; Pfizer; Pharmaventure; Phesi; Phillips; P&G; Pronota; Protagonist; Proximagen; Resolute; Robarts; Sandoz; Santarus; Satisfai; Sensyne Health; Shire; SigmoidPharma; Sorriso; Souffinez; Syndermix; Synthon; Takeda; Theravance; Tigenix; Tillotts; Topivert; Trino Therapeutics with Wellcome Trust; TxCell; UCB Pharma; Vertex; VHsquared; Vifor; Warner Chilcott and Zeria

Speaker fees: AbbVie, Amgen, Biogen, Falk; Ferring, Janssen, Pfizer, Shire, Takeda, UCB.

No stocks or shares options.

Ian Zealley – Grant support from the NIHR HTA for the present manuscript.

Tim Raine – Grant: Abbvie

Personal Fees: Abbvie, Arena, Aslan, AstraZeneca, Boehringer-Ingelheim, BMS, Celgene, Ferring, Galapagos, Gilead, GSK, Heptares, LabGenius, Janssen, Mylan, MSD, Novartis, Pfizer, Roche, Sandoz, Takeda and UCB

Steve Halligan – Grant support from the NIHR HTA for the present manuscript. NIHR emeritus senior investigator

Stuart A Taylor – Grant support from the NIHR HTA for the present manuscript. Personal fees from Alimentiv, shareholding in Motilent and grants from the NIHR. Consultant for AstraZeneca. Board member EME programme committee. NIHR emeritus senior investigator.

Disclaimer

Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

Publications

Kumar S, Plumb A, Mallett S, Bhatnagar G, Bloom S, Clarke CS, *et al*. METRIC-EF: magnetic resonance enterography to predict disabling disease in newly diagnosed Crohn's disease-protocol for a multicentre, non-randomised, single-arm, prospective study. *BMJ Open* 2022;**12**:e067265. https://doi.org/10.1136/bmjopen-2022-067265

Taylor SA, Kumar S, Parry T, Mallett S, Travis S, Raine T, *et al*. Magnetic resonance enterography to predict subsequent disabling Crohn's disease in newly diagnosed patients (METRIC-EF)-multivariable prediction model, multicentre diagnostic inception cohort. *Eur Radiol* 2025;**35**(11):7333–45. https://doi.org/10.1007/s00330-025-11636-8

References

1. Kumar S, Plumb A, Mallett S, Bhatnagar G, Bloom S, Clarke CS, *et al*. METRIC-EF: magnetic resonance enterography to predict disabling disease in newly diagnosed Crohn's disease-protocol for a multicentre, non-randomised, single-arm, prospective study. *BMJ Open* 2022;**12**:e067265. https://doi.org/10.1136/bmjopen-2022-067265
2. Ananthakrishnan AN, Nguyen GC, Bernstein CN. AGA clinical practice update on management of inflammatory bowel disease in elderly patients: expert review. *Gastroenterology* 2021;**160**:445–51.
3. Lamb CA, Kennedy NA, Raine T, Hendy PA, Smith PJ, Limdi JK, *et al*.; IBD Guidelines eDelphi Consensus Group. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019;**68**:s1–106.
4. Lopes EW, Lochhead P, Burke KE, Richter JM, Ananthakrishnan AN, Chan AT, Khalili H. Risk factors for incident inflammatory bowel disease according to disease phenotype. *Clin Gastroenterol Hepatol* 2022;**20**:2347–57.e14.
5. Rogler G, Singh A, Kavanaugh A, Rubin DT. Extraintestinal manifestations of inflammatory bowel disease: current concepts, treatment, and implications for disease management. *Gastroenterology* 2021;**161**:1118–32.
6. Kumar S, Pollok R, Goldsmith D. Renal and urological disorders associated with inflammatory bowel disease. *Inflamm Bowel Dis* 2022;**29**:1306–16.
7. Gomollon F, Dignass A, Annese V, Tilg H, Van Assche G, Lindsay JO, *et al*.; ECCO. 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.
8. Peyrin-Biroulet L, Loftus EV Jr, Colombel JF, Sandborn WJ. The natural history of adult Crohn's disease in population-based cohorts. *Am J Gastroenterol* 2010;**105**:289–97.
9. Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. *Gastroenterology* 2010;**139**:1147–55.
10. GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020;**5**:17–30.
11. Ghosh N, Premchand P. A UK cost of care model for inflammatory bowel disease. *Frontline Gastroenterol* 2015;**6**:169–74.
12. Bassi A, Dodd S, Williamson P, Bodger K. Cost of illness of inflammatory bowel disease in the UK: a single centre retrospective study. *Gut* 2004;**53**:1471–8.
13. Dignass A, Van Assche G, Lindsay JO, Lémann M, Söderholm J, Colombel JF, *et al*.; European Crohn's and Colitis Organisation (ECCO). The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: current management. *J Crohns Colitis* 2010;**4**:28–62.
14. Lichtenstein GR, Hanauer SB, Sandborn WJ; Practice Parameters Committee of American College of Gastroenterology. Management of Crohn's disease in adults. *Am J Gastroenterol* 2009;**104**:465–83; quiz 464, 484.
15. Schoepfer AM, Beglinger C, Straumann A, Trummler M, Vavricka SR, Bruegger LE, Seibold F. Fecal calprotectin correlates more closely with the Simple Endoscopic Score for Crohn's disease (SES-CD) than CRP, blood leukocytes, and the CDAI. *Am J Gastroenterol* 2010;**105**:162–9.
16. Peyrin-Biroulet L, Reinisch W, Colombel JF, Mantzaris GJ, Kornbluth A, Diamond R, *et al*. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. *Gut* 2014;**63**:88–95.

17. Khanna R, Bressler B, Levesque BG, Zou G, Stitt LW, Greenberg GR, *et al.*; REACT Study Investigators. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015;**386**:1825–34.
18. Colombel JF, Panaccione R, Bossuyt P, Lukas M, Baert F, Vaňásek T, *et al*. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. *Lancet* 2017;**390**:2779–89.
19. Noor NM, Lee JC, Bond S, Dowling F, Brezina B, Patel KV, *et al*. A biomarker-stratified comparison of top-down versus accelerated step-up treatment: strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. *Lancet Gastroenterol Hepatol* 2024;**9**:e000347.
20. Cushing K, Higgins PDR. Management of Crohn disease: a review. *JAMA* 2021;**325**:69–80.
21. Lichtenstein GR, Yan S, Bala M, Blank M, Sands BE. Infliximab maintenance treatment reduces hospitalizations, surgeries, and procedures in fistulizing Crohn's disease. *Gastroenterology* 2005;**128**:862–9.
22. Peyrin-Biroulet L, Deltenre P, de Suray N, Branche J, Sandborn WJ, Colombel JF. Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: meta-analysis of placebo-controlled trials. *Clin Gastroenterol Hepatol* 2008;**6**:644–53.
23. Peyrin-Biroulet L, Panes J, Sandborn WJ, Vermeire S, Danese S, Feagan BG, *et al*. Defining disease severity in inflammatory bowel diseases: current and future directions. *Clin Gastroenterol Hepatol* 2016;**14**:348–54.e17.
24. Loly C, Belaiche J, Louis E. Predictors of severe Crohn's disease. *Scand J Gastroenterol* 2008;**43**:948–54.
25. Beaugerie L, Seksik P, Nion-Larmurier I, Gendre JP, Cosnes J. Predictors of Crohn's disease. *Gastroenterology* 2006;**130**:650–6.
26. Yang CH, Ding J, Gao Y, Chen X, Yang ZB, Xiao SD. Risk factors that predict the requirement of aggressive therapy among Chinese patients with Crohn's disease. *J Dig Dis* 2011;**12**:99–104.
27. Magro F, Estevinho MM, Catalano G, Patita M, Arroja B, Lago P, *et al.*; GEDII (Grupo de Estudos da Doença Inflamatória Intestinal). How many biomarker measurements are needed to predict prognosis in Crohn's disease patients under infliximab? A prospective study. *United European Gastroenterol J* 2023;**11**:531–41.
28. Noor NM, Verstockt B, Parkes M, Lee JC. Personalised medicine in Crohn's disease. *Lancet Gastroenterol Hepatol* 2020;**5**:80–92.
29. Vermeire S, Ferrante M, Rutgeerts P. Recent advances: personalised use of current Crohn's disease therapeutic options. *Gut* 2013;**62**:1511–5.
30. Rimola J, Torres J, Kumar S, Taylor SA, Kucharzik T. Recent advances in clinical practice: advances in cross-sectional imaging in inflammatory bowel disease. *Gut* 2022;**71**:2587–97.
31. 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.
32. Kumar S, De Kock I, Blad W, Hare R, Pollok R, Taylor SA. Magnetic resonance enterography and intestinal ultrasound for the assessment and monitoring of Crohn's disease. *J Crohns Colitis* 2024;**18**:1450–63.
33. Hameed M, Kumar S, Taylor SA. How I do it: cross-sectional imaging in small-bowel Crohn disease and ulcerative colitis. *Radiology* 2025;**314**:e241452.
34. Rimola J, Ordas 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.
35. Taylor SA, Mallett S, Bhatnagar G, Baldwin-Cleland R, Bloom S, Gupta A, *et al.*; METRIC Study Investigators. 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.

36. Ordas I, Rimola J, Alfaro I, Rodríguez S, Castro-Poceiro J, Ramírez-Morros A, *et al*. Development and validation of a simplified magnetic resonance index of activity for Crohn's disease. *Gastroenterology* 2019;**157**:432–9.e1.

37. Hanzel J, Jairath V, Ma C, Guizzetti L, Zou G, Santillan CS, *et al*. Responsiveness of magnetic resonance enterography indices for evaluation of luminal disease activity in Crohn's disease. *Clin Gastroenterol Hepatol* 2022;**20**:2598–606.

38. Kumar S, Parry T, Mallett S, Bhatnagar G, Plumb A, Walsh S, *et al*. Diagnostic performance of magnetic resonance enterography disease activity indices compared with a histological reference standard for adult terminal ileal Crohn's disease: experience from the METRIC trial. *J Crohns Colitis* 2022;**16**:1531–9.

39. Kumar S, Rao N, Bhagwanani A, Parry T, Hameed M, Rahman S, *et al*. Volumetric measurement of terminal ileal Crohn's disease by magnetic resonance enterography: a feasibility study. *Eur Radiol* 2024;**35**:117–26.

40. Rao N, Kumar S, Taylor S, Plumb A. Diagnostic pathways in Crohn's disease. *Clin Radiol* 2019;**74**:578–91.

41. Miles A, Bhatnagar G, Halligan S, Gupta A, Tolan D, Zealley I, Taylor SA; METRIC Investigators. 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.

42. Guglielmo FF, Anupindi SA, Fletcher JG, Al-Hawary MM, Dillman JR, Grand DJ, *et al*. Small bowel Crohn disease at CT and MR enterography: imaging Atlas and glossary of terms. *Radiographics* 2020;**40**:354–75.

43. Halligan S, Boone D, Archer L, Ahmad T, Bloom S, Rodriguez-Justo M, *et al*. Prognostic biomarkers to identify patients likely to develop severe Crohn's disease: a systematic review. *Health Technol Assess* 2021;**25**:1–66.

44. Jauregui-Amezaga A, Rimola J, Ordas I, Rodríguez S, Ramírez-Morros A, Gallego M, *et al*. Value of endoscopy and MRI for predicting intestinal surgery in patients with Crohn's disease in the era of biologics. *Gut* 2015;**64**:1397–402.

45. Fiorino G, Morin M, Bonovas S, Bonifacio C, Spinelli A, Germain A, *et al*. Prevalence of bowel damage assessed by cross-sectional imaging in early Crohn's disease and its impact on disease outcome. *J Crohns Colitis* 2017;**11**:274–80.

46. Fernandez-Clotet A, Panes J, Ricart E, Castro-Poceiro J, Masamunt MC, Rodríguez S, *et al*. Predictors of bowel damage in the long-term progression of Crohn's disease. *World J Clin Cases* 2022;**10**:12208–20.

47. Dane B, Qian K, Krieger R, Smereka P, Foster J, Huang C, *et al*. Correlation between imaging findings on outpatient MR enterography (MRE) in adult patients with Crohn disease and progression to surgery within 5 years. *Abdom Radiol* 2022;**47**:3424–35.

48. Taylor SA, Kumar S, Parry T, Mallett S, Travis S, Raine T, *et al*. Magnetic resonance enterography to predict subsequent disabling Crohn's disease in newly diagnosed patients (METRIC-EF)-multivariable prediction model, multicentre diagnostic inception cohort. *Eur Radiol* 2025;**35**(11):7333–45. https://doi.org/10.1007/s00330-025-11636-8

49. 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**:1–162.

50. 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.

51. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;**350**:g7594.

52. Prezzi D, Bhatnagar G, Vega R, Makanyanga J, Halligan S, Taylor SA. Monitoring Crohn's disease during anti-TNF-alpha therapy: validation of the magnetic resonance enterography global score (MEGS) against a combined clinical reference standard. *Eur Radiol* 2016;**26**:2107–17.

53. Makanyanga JC, Pendse D, Dikaios N, Bloom S, McCartney S, Helbren E, *et al*. Evaluation of Crohn's disease activity: initial validation of a magnetic resonance enterography global score (MEGS) against faecal calprotectin. *Eur Radiol* 2014;**24**:277–87.

54. Jose SK, Simon B, Simon EG, Eapen A, John RA, Putta T, *et al*. Comparison of magnetic resonance enterography global score (MEGS) with indices of Crohn's disease activity in South Asian population. *Abdom Radiol* 2022;**47**:547–53.

55. Klang E, Amitai MM, Lahat A, Yablecovitch D, Avidan B, Neuman S, *et al*.; Israeli IBD Research Nucleus (IIRN). Capsule endoscopy validation of the magnetic enterography global score in patients with established Crohn's disease. *J Crohns Colitis* 2017;**12**:313–20.

56. Capozzi N, Ordas I, Fernandez-Clotet A, Castro-Poceiro J, Rodríguez S, Alfaro I, *et al*. Validation of the Simplified Magnetic Resonance Index of Activity [sMARIA] without gadolinium-enhanced sequences for Crohn's disease. *J Crohns Colitis* 2020;**14**:1074–81.

57. Rimola J, Alvarez-Cofino A, Perez-Jeldres T, Ayuso C, Alfaro I, Rodríguez S, *et al*. Comparison of three magnetic resonance enterography indices for grading activity in Crohn's disease. *J Gastroenterol* 2017;**52**:585–93.

58. Pariente B, Mary JY, Danese S, Chowers Y, De Cruz P, D'Haens G, *et al*. Development of the Lemann index to assess digestive tract damage in patients with Crohn's disease. *Gastroenterology* 2015;**148**:52–63.e3.

59. Dias CC, Rodrigues PP, da Costa-Pereira A, Magro F. Clinical prognostic factors for disabling Crohn's disease: a systematic review and meta-analysis. *World J Gastroenterol* 2013;**19**:3866–71.

60. Rabilloud ML, Bajeux E, Siproudhis L, Hamonic S, Pagenault M, Brochard C, *et al*. Long-term outcomes and predictors of disabling disease in a population-based cohort of patients with incident Crohn's disease diagnosed between 1994 and 1997. *Clin Res Hepatol Gastroenterol* 2022;**46**:101974.

61. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;**55**:749–53.

62. Sandborn WJ, Fazio VW, Feagan BG, Hanauer SB; American Gastroenterological Association Clinical Practice Committee. AGA technical review on perianal Crohn's disease. *Gastroenterology* 2003;**125**:1508–30.

63. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;**58**:475–83.

64. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;**48**:1503–10.

65. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, Snell KIE. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med* 2021;**40**:4230–51.

66. Cosnes J, Cattan S, Blain A, Beaugerie L, Carbonnel F, Parc R, Gendre JP. Long-term evolution of disease behavior of Crohn's disease. *Inflamm Bowel Dis* 2002;**8**:244–50.

67. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–94.

68. National Institute for Health and Care Excellence. *NICE Health Technology Evaluations: The Manual. Process and Methods* (PMG36). URL: www.nice.org.uk/process/pmg36 (accessed 15 April 2025).

69. NHS. *2020/21 National Cost Collection Data Publication*. URL: www.england.nhs.uk/publication/2020-21-national-cost-collection-data-publication/ (accessed 15 April 2025).

70. Jo C. Cost-of-illness studies: concepts, scopes, and methods. *Clin Mol Hepatol* 2014;**20**:327–37.

71. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, *et al*. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012;**142**:46–54.e42; quiz e30.

72. Byford S, Torgerson DJ, Raftery J. Economic note: cost of illness studies. *BMJ* 2000;**320**:1335.
73. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG clinical guideline: management of Crohn's disease in adults. *Am J Gastroenterol* 2018;**113**:481–517.
74. Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, *et al*.; European Crohn's and Colitis Organisation [ECCO] and the European Society of Gastrointestinal and Abdominal Radiology [ESGAR]. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;**13**:144–64.
75. Roseira J, Ventosa AR, de Sousa HT, Brito J. The new simplified MARIA score applies beyond clinical trials: a suitable clinical practice tool for Crohn's disease that parallels a simple endoscopic index and fecal calprotectin. *United European Gastroenterol J* 2020;**8**:1208–16.
76. Tao Y, Li H, Xu H, Tang W, Fan G, Yang X. Can the simplified magnetic resonance index of activity be used to evaluate the degree of activity in Crohn's disease? *BMC Gastroenterol* 2021;**21**:409.
77. Klang E, Amitai MM, Lahat A, Yablecovitch D, Avidan B, Neuman S, *et al*.; Israeli IBD Research Nucleus (IIRN). Capsule endoscopy validation of the magnetic enterography global score in patients with established Crohn's disease. *J Crohns Colitis* 2018;**12**:313–20.
78. Rimola J, Rodriguez S, Garcia-Bosch O, Ordás I, Ayala E, Aceituno M, *et al*. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. *Gut* 2009;**58**:1113–20.
79. Kumar S, Parry T, Mallett S, Plumb A, Bhatnagar G, Beable R, *et al*.; METRIC Study Group. Diagnostic performance of sonographic activity scores for adult terminal ileal Crohn's disease compared to magnetic resonance and histological reference standards: experience from the METRIC trial. *Eur Radiol* 2024;**34**:455–64.
80. Bhatnagar G, Mallett S, Quinn L, Beable R, Bungay H, Betts M, *et al*. Interobserver variation in the interpretation of magnetic resonance enterography in Crohn's disease. *Br J Radiol* 2022;**95**:20210995.
81. 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.
82. 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 Fortschr Geb Rontgenstr Nuklearmed* 2013;**184**:992–7.
83. 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.
84. Thompson NP, Fleming DM, Charlton J, Pounder RE, Wakefield AJ. Patients consulting with Crohn's disease in primary care in England and Wales. *Eur J Gastroenterol Hepatol* 1998;**10**:1007–12.
85. Chew C, Albazaz R, Taylor SA; BSGAR Committee. Diversity and equity: a radiology society's effort. *Clin Radiol* 2021;**76**:475–6.

Appendix 1

TABLE 21 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model B

	Observed data (*N* = 146)		Imputed data (*N* = 194)	
Prespecified clinical predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
≥ 40 years of age	0.78 (0.40 to 1.52)	0.466	0.73 (0.42 to 1.27)	0.269
Female	0.87 (0.50 to 1.50)	0.608	0.85 (0.54 to 1.34)	0.485
Smoker	1.82 (1.07 to 3.11)	0.028	1.50 (0.93 to 2.42)	0.096
Weight loss ≥ 5 kg prior to diagnosis	0.79 (0.44 to 1.43)	0.437	0.70 (0.38 to 1.27)	0.240
Initial need for steroid therapy	2.42 (1.39 to 4.21)	0.002	2.05 (1.28 to 3.28)	0.003
Developed MBDD ≤ 90 days from diagnosis	1.18 (0.55 to 2.56)	0.670	1.16 (0.59 to 2.26)	0.664
Perianal disease	1.48 (0.65 to 3.36)	0.346	1.22 (0.60 to 2.47)	0.581
Severe endoscopic disease	0.73 (0.38 to 1.41)	0.351	0.81 (0.45 to 1.46)	0.492
Disease behaviour				
B1	–	–		–
B2	1.19 (0.62 to 2.29)	0.607	1.33 (0.73 to 2.43)	0.348
B3	1.80 (0.86 to 3.76)	0.119	1.40 (0.75 to 2.63)	0.297
Location of disease behaviour				
Ileocolonic	–	–		–
Ileal/upper tract	0.92 (0.51 to 1.66)	0.773	0.89 (0.53 to 1.49)	0.660
Colonic	0.77 (0.32 to 1.90)	0.575	0.98 (0.48 to 1.99)	0.957

TABLE 22 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model A1

	Observed data (*N* = 146)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
≥ 40 years of age	0.80 (0.41 to 1.57)	0.519	0.73 (0.42, 1.28)	0.275
Female	0.87 (0.50 to 1.51)	0.627	0.85 (0.54 to 1.34)	0.486
Smoker	1.77 (1.02 to 3.05)	0.041	1.49 (0.91 to 2.44)	0.113
Weight loss ≥ 5 kg prior to diagnosis	0.77 (0.43 to 1.41)	0.405	0.69 (0.37 to 1.27)	0.232
Initial need for steroid therapy	2.44 (1.40 to 4.27)	0.002	2.06 (1.28 to 3.29)	0.003
Developed MBDD ≤ 90 days from diagnosis	1.09 (0.47 to 2.53)	0.850	1.15 (0.58 to 2.30)	0.690
Perianal disease	1.54 (0.67 to 3.54)	0.307	1.23 (0.60 to 2.52)	0.565
Severe endoscopic disease	0.71 (0.36 to 1.37)	0.305	0.80 (0.43 to 1.46)	0.462

TABLE 22 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model A1 (*continued*)

	Observed data (*N* = 146)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
Disease behaviour				
B1	–	–		–
B2	1.17 (0.60 to 2.26)	0.644	1.33 (0.73 to 2.44)	0.349
B3	1.75 (0.83 to 3.68)	0.142	1.40 (0.74 to 2.64)	0.305
Location of disease behaviour				
Ileocolonic	–	–		–
Ileal/upper tract	0.93 (0.51 to 1.69)	0.818	0.89 (0.53 to 1.49)	0.660
Colonic	0.80 (0.32 to 1.98)	0.632	0.99 (0.48 to 2.01)	0.970
Normalised global MEGS (%)	1.01 (0.99 to 1.03)	0.598	1.00 (0.98 to 1.02)	0.918

Note
Scores were normalised to enable comparison of the scores on a standardised scale.

TABLE 23 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model A2

	Observed data (*N* = 146)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
≥ 40 years of age	0.74 (0.38 to 1.46)	0.384	0.69 (0.39 to 1.21)	0.193
Female	0.87 (0.50 to 1.51)	0.628	0.86 (0.54 to 1.37)	0.528
Smoker	1.95 (1.13 to 3.37)	0.017	1.66 (1.01 to 2.73)	0.046
Weight loss ≥ 5 kg prior to diagnosis	0.81 (0.45 to 1.46)	0.479	0.69 (0.38 to 1.26)	0.232
Initial need for steroid therapy	2.37 (1.37 to 4.13)	0.002	2.02 (1.27 to 3.23)	0.003
Developed MBDD ≤ 90 days from diagnosis	1.33 (0.60 to 2.93)	0.483	1.25 (0.64 to 2.44)	0.522
Perianal disease	1.30 (0.61 to 3.20)	0.430	1.16 (0.57 to 2.38)	0.676
Severe endoscopic disease	0.79 (0.40 to 1.55)	0.493	0.91 (0.49 to 1.68)	0.765
Disease behaviour				
B1	–	–	–	–
B2	1.18 (0.61 to 2.28)	0.631	1.33 (0.73 to 2.44)	0.349
B3	1.93 (0.91 to 4.09)	0.088	1.48 (0.78 to 2.82)	0.232
Location of disease behaviour				
Ileocolonic	–	–	–	–
Ileal/upper tract	0.86 (0.47 to 1.57)	0.621	0.85 (0.51 to 1.43)	0.547
Colonic	0.75 (0.30 to 1.84)	0.525	0.97 (0.48 to 1.98)	0.934
Normalised global sMARIA (%)	0.99 (0.97 to 1.01)	0.291	0.99 (0.97 to 1.00)	0.153

Note
Scores were normalised to enable comparison of the scores on a standardised scale.

TABLE 24 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model A3

	Observed data (*N* = 146)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
≥ 40 years of age	0.77 (0.39 to 1.50)	0.443	0.73 (0.42 to 1.26)	0.257
Female	0.86 (0.50 to 1.50)	0.601	0.85 (0.54 to 1.34)	0.487
Smoker	1.87 (1.09 to 3.22)	0.023	1.54 (0.94 to 2.51)	0.085
Weight loss ≥ 5 kg prior to diagnosis	0.81 (0.45 to 1.47)	0.496	0.70 (0.38 to 1.28)	0.248
Initial need for steroid therapy	2.40 (1.38 to 4.18)	0.002	2.05 (1.28 to 3.27)	0.003
Developed MBDD ≤ 90 days from diagnosis	1.28 (0.57 to 2.86)	0.555	1.19 (0.60 to 2.32)	0.621
Perianal disease	1.42 (0.62 to 3.27)	0.410	1.22 (0.60 to 2.47)	0.588
Severe endoscopic disease	0.77 (0.39 to 1.52)	0.451	0.84 (0.45 to 1.55)	0.568
Disease behaviour				
B1	–	–		–
B2	1.18 (0.61 to 2.28)	0.628	1.33 (0.73 to 2.43)	0.349
B3	1.88 (0.88 to 3.98)	0.101	1.42 (0.75 to 2.67)	0.281
Location of disease behaviour				
Ileocolonic	–	–		–
Ileal/upper tract	0.89 (0.48 to 1.62)	0.696	0.88 (0.53 to 1.48)	0.635
Colonic	0.76 (0.31 to 1.86)	0.546	0.98 (0.48 to 1.98)	0.945
Normalised LI (%)	1.00 (0.98 to 1.01)	0.567	0.98 (0.90 to 1.07)	0.695

Note
Scores were normalised to enable comparison of the scores on a standardised scale.

TABLE 25 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model B1

	Observed data (*N* = 46)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
≥ 40 years of age	0.62 (0.15 to 2.49)	0.502	0.70 (0.38 to 1.28)	0.248
Female	2.01 (0.55 to 7.30)	0.290	0.92 (0.52 to 1.62)	0.767
Smoker	3.81 (0.95 to 15.31)	0.059	1.67 (0.91 to 3.07)	0.099
Weight loss ≥ 5 kg prior to diagnosis	2.23 (0.69 to 7.22)	0.181	0.72 (0.36 to 1.45)	0.350
Initial need for steroid therapy	4.13 (1.24 to 13.73)	0.021	2.00 (1.22 to 3.31)	0.006
Developed MBDD ≤ 90 days from diagnosis	1.15 (0.20 to 6.53)	0.875	1.22 (0.54 to 2.75)	0.630
Perianal disease	4.73 (0.77 to 29.19)	0.094	1.21 (0.56 to 2.63)	0.630
Severe endoscopic disease	1.31 (0.30 to 5.64)	0.721	0.83 (0.45 to 1.55)	0.561
Disease behaviour				
B1	–	–		–
B2	1.18 (0.33 to 4.18)	0.796	1.36 (0.71 to 2.58)	0.353
B3	3.26 (0.70 to 15.20)	0.133	1.30 (0.61 to 2.78)	0.499

TABLE 25 Multivariable HRs of prespecified clinical predictors for predicting development of MBDD within 5 years of diagnosis, using observed and imputed data for Model B1 (*continued*)

	Observed data (*N* = 46)		Imputed data (*N* = 194)	
Prespecified predictors	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
Location of disease behaviour				
Ileocolonic	–	–		–
Ileal/upper tract	1.38 (0.32 to 5.99)	0.664	0.85 (0.49 to 1.48)	0.555
Colonic	1.96 (0.29 to 13.31)	0.490	0.97 (0.46 to 2.03)	0.933
CRP level (mg/l)	1.00 (0.98 to 1.02)	0.791	1.00 (1.00 to 1.01)	0.563
WBC count (10^9/l)	0.83 (0.62 to 1.10)	0.188	0.98 (0.90 to 1.06)	0.571
FC level (µg/g)	1.00 (1.00 to 1.00)	0.674	1.00 (1.00 to 1.00)	0.929
Haemoglobin level (g/l)	1.03 (0.99 to 1.08)	0.119	1.00 (0.98 to 1.02)	0.991
Platelet count (10^9/l)	1.00 (1.00 to 1.01)	0.101	1.00 (1.00 to 1.00)	0.741

Appendix 2

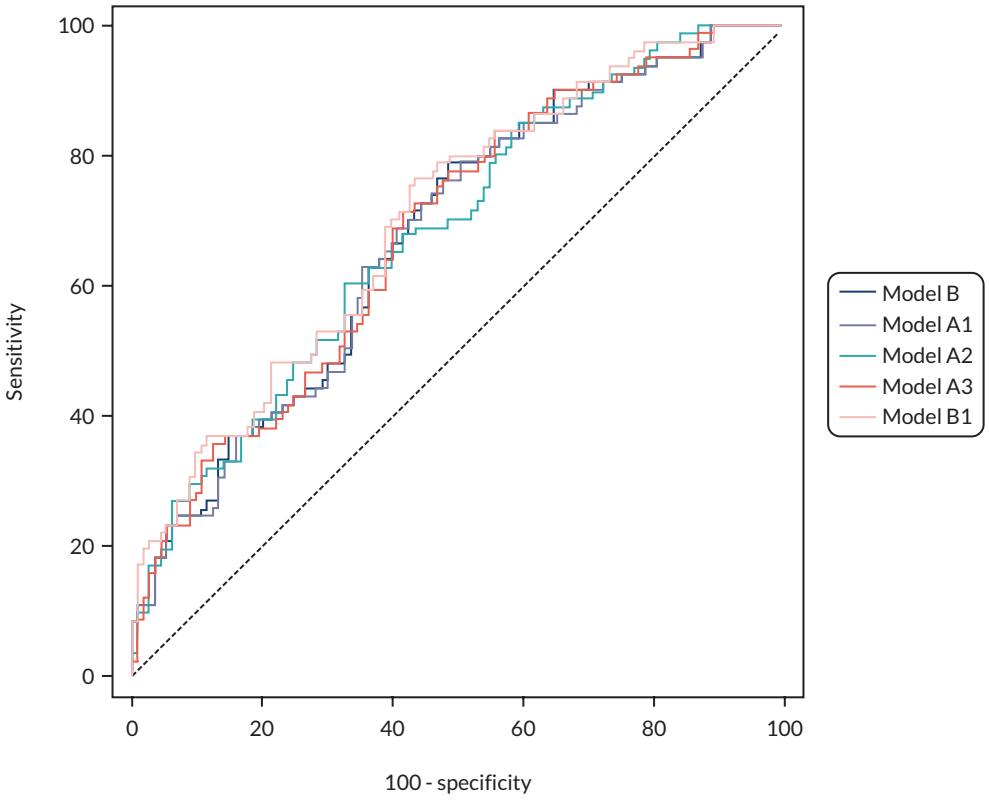

FIGURE 8 The ROC plot of prognostic models for predicting development of MBDD within 5 years of diagnosis, and the area under the curve of prognostic models for predicting development of MBDD within 5 years of diagnosis. The ROC plot demonstrates that the sensitivity and specificity of the models is similar across all risk thresholds. Reproduced with permission from Taylor *et al.*[48] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The figure includes minor additions and formatting changes to the original text.

Prognostic model	Area under the curve (95% CI)
B	0.68 (0.60 to 0.75)
A1	0.68 (0.60 to 0.75)
A2	0.68 (0.61 to 0.76)
A3	0.68 (0.61 to 0.76)
B1	0.70 (0.63 to 0.78)

Appendix 3

TABLE 26 Net benefit of prognostic models for predicting development of MBDD within 5 years of diagnosis, using ratios of TP to FP predictions

Prognostic model	RD	1TP to 1FP	1TP to 2FP	1TP to 5FP	2TP to 1FP	5TP to 1FP
B	1	1 (−7, 9)	11 (4, 18)	17 (10, 23)	−9 (−16, −2)	−15 (−22, −9)
A1		2 (−6, 10)	12 (5, 19)	17 (11, 24)	−9 (−16, −1)	−15 (−21, −9)
A2		3 (−5, 11)	12 (5, 19)	18 (12, 24)	−8 (−15, −1)	−14 (−20, −8)
A3		3 (−5, 11)	12 (5, 19)	17 (11, 24)	−8 (−15, −1)	−14 (−20, −8)
B1		5 (−3, 13)	13 (6, 20)	19 (12, 25)	−6 (−13, 0)	−13 (−19, −7)
B	2	−2 (−11, 8)	17 (9, 26)	29 (21, 36)	−20 (−28, −11)	−30 (−38, −23)
A1		−3 (−13, 7)	18 (9, 26)	30 (22, 38)	−22 (−31, −13)	−34 (−41, −26)
A2		−4 (−14, 6)	17 (8, 26)	30 (22, 38)	−23 (−32, −14)	−35 (−43, −27)
A3		−3 (−13, 7)	18 (9, 26)	30 (22, 38)	−22 (−31, −13)	−34 (−41, −26)
B1		−2 (−11, 8)	18 (9, 27)	30 (22, 38)	−21 (−29, −12)	−32 (−40, −24)

Appendix 4

TABLE 27 Variable loadings for PCs of prespecified predictors

Prespecified predictors	Component 1	Component 2	Component 3	Component 4
Maximum mural thickness (MEGS)	0.36	−0.24	0.1	−0.13
Maximum mural T2 signal (oedema) (MEGS)	0.4	−0.14	0.05	−0.03
Maximum contrast enhancement pattern	0.31	−0.15	0.07	0.38
Maximum length of disease (MEGS)	0.38	−0.22	−0.01	−0.12
Abscess (MEGS)	0.25	0.56	0.19	0.09
Maximum fat stranding (sMARIA)	0.36	−0.13	0.1	0
Number of abnormal segments	0.33	−0.2	−0.11	−0.39
Maximum upper tract and SB stricturing	0.2	−0.05	−0.2	0.67
Maximum colon stricturing	0.03	0.12	0.89	−0.03
Maximum upper tract and SB penetrating	0.29	0.52	−0.15	0.18
Maximum colon penetrating	0.22	0.44	−0.27	−0.43

Appendix 5

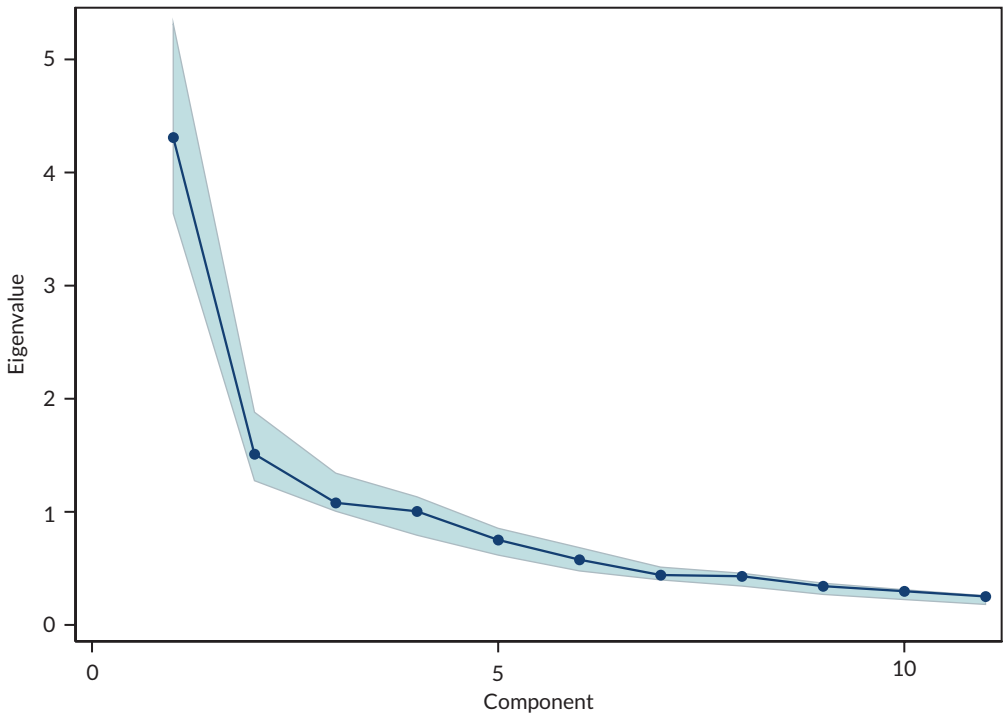

FIGURE 9 Scree plot of variance explained by PCs.

Appendix 6

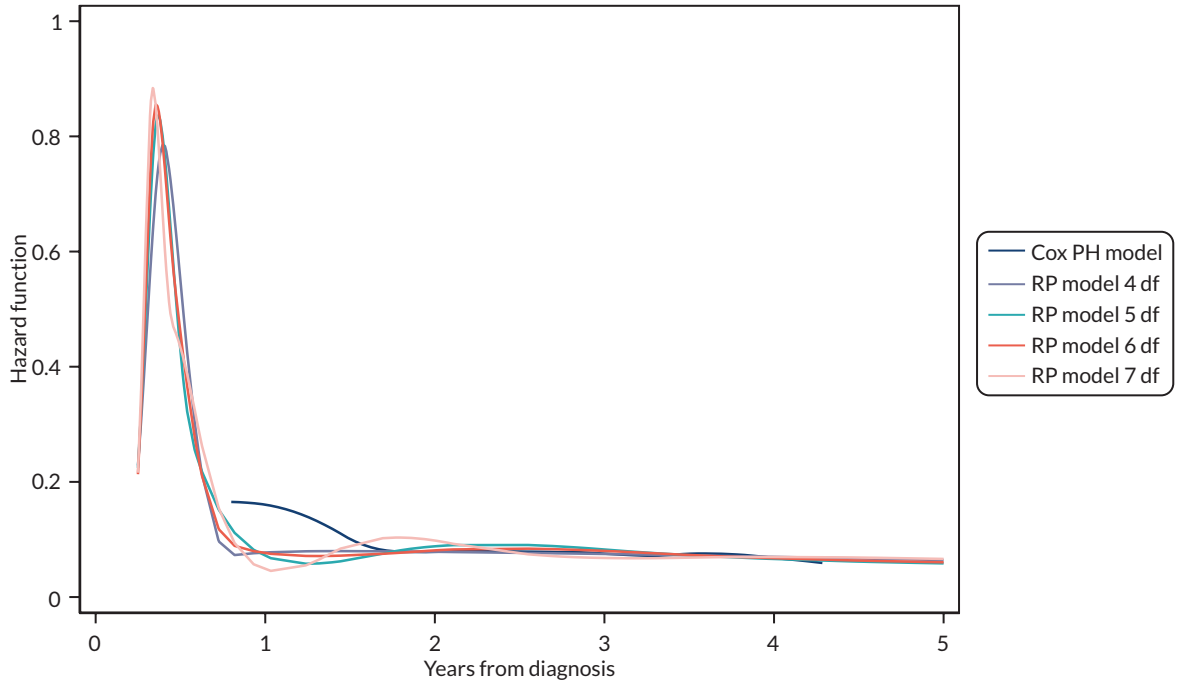

FIGURE 10 Baseline hazard functions of prognostic modelling methods. df, degrees of freedom; PH, proportional hazards.

- The rationale for choosing a RP model is that the Cox proportional hazards (PH) model is unable to account for participants who developed MBDD within a year from diagnosis.
- The rationale for choosing a RP model with 4 df is that the initial values were not feasible for models with fewer df.
- The model with 4 df also had the lowest AIC and BIC.

TABLE 28 The AIC and BIC of modelling methods

Prognostic modelling method	df	AIC	BIC
Cox PH model	–	814	814
RP flexible parametric model	4	445	461
	5	446	466
	6	447	470
	7	447	473

There were five prespecified continuous predictors, including CRP, WBC count, FC, haemoglobin and platelet count. We determined whether we should include the predictors as linear or to use fractional polynomials. We were unable to model WBC count and FC, so assumed a linear relationship. For the remaining predictors, we calculated the best fractional polynomial models by searching through all possible power combinations. Then, we calculated *p*-values by comparing the deviance of the linear and fractional polynomial model 1 against the deviance of fractional polynomial model 2 (lowest deviance). For each continuous predictor, we found that the linear is most efficient.

TABLE 29 Cubic splines and fractional polynomials of prespecified continuous predictors

	Power		Deviance difference from fractional polynomial model 2		*p*-value of deviance difference	
Prespecified continuous predictor	Fractional polynomial model 1	Fractional polynomial model 2	Linear	Fractional polynomial model 1	Linear	Fractional polynomial model 1
CRP level (mg/l)	−0.5	−2	4.754	2.007	0.191	0.367
WBC count (10^9/l)	3	0, 3	1.963	1.194	0.580	0.550
Haemoglobin level (g/l)	−2	−2, −2	0.209	0.109	0.947	0.947

Seventy-five per cent of participants had at least one predictor missing. Because missing values were likely to be missing at random, and to avoid loss in efficiency, missing values for smoking status, weight loss ≥ 5 kg prior to diagnosis, perianal disease, severe endoscopic disease, CRP, WBC count, FC, haemoglobin and platelet count were imputed using multiple imputation by chained equations. Twenty imputed data sets were created by replacing missing values with simulated values from a set of imputation models constructed from all predictors and outcomes (event indicator and Nelson-Aalen estimator for time to event).

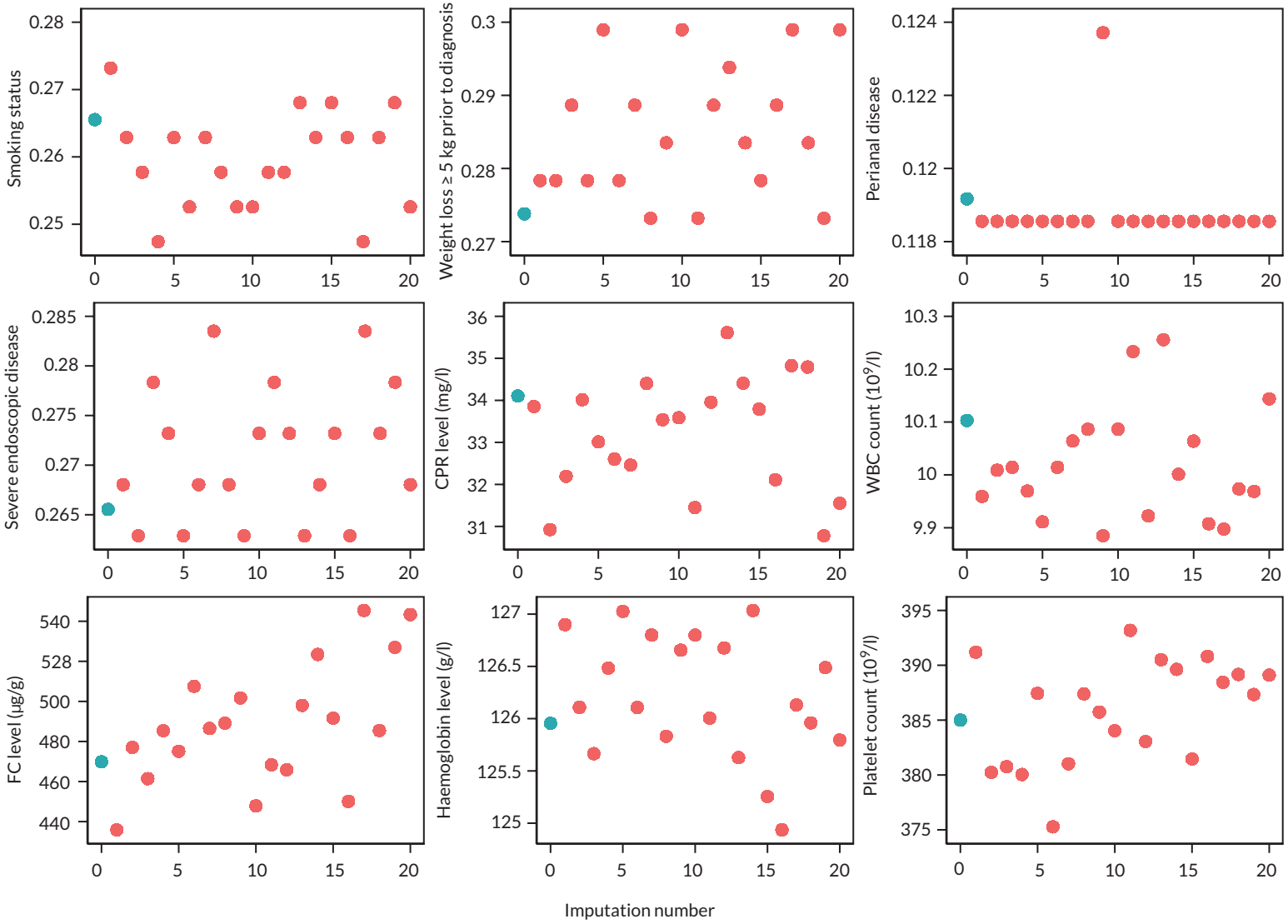

FIGURE 11 Mean observed and imputed by prespecified predictor. Distributions of imputed values were visually checked for comparability with observed data.

Appendix 7

TABLE 30 Internal validation of prognostic models

Prognostic model	Data	Mean linear predictor (SD)	Harrell's *c*-statistic	R^2 (95% CI)	*D*-statistic	*C*-slope (95% CI)	Heuristic shrinkage factor
B	Observed	0.30 (0.61)	0.66	0.19 (0.07 to 0.33)	1.00	1.00 (0.56 to 1.44)	0.977
	Imputed	0.12 (0.48)	0.67	0.14 (0.05 to 0.25)	0.83	1.04 (0.71 to 1.37)	0.980
A1	Observed	0.32 (0.61)	0.66	0.19 (0.07 to 0.33)	1.00	1.00 (0.56 to 1.44)	0.976
	Imputed	0.13 (0.48)	0.67	0.14 (0.04 to 0.23)	0.83	1.04 (0.81 to 1.27)	0.979
A2	Observed	0.13 (0.63)	0.67	0.22 (0.09 to 0.36)	1.08	1.00 (0.57 to 1.43)	0.976
	Imputed	−0.06 (0.51)	0.68	0.15 (0.06 to 0.26)	0.86	1.04 (0.72 to 1.35)	0.979
A3	Observed	0.24 (0.61)	0.66	0.20 (0.08 to 0.34)	1.02	1.00 (0.56 to 1.44)	0.976
	Imputed	0.09 (0.48)	0.67	0.14 (0.05 to 0.24)	0.83	1.04 (0.77 to 1.31)	0.979
B1	Observed	5.72 (1.22)	0.76	0.47 (0.22 to 0.64)	1.91	1.00 (0.57 to 1.43)	0.965
	Imputed	−0.24 (0.50)	0.68	0.16 (0.06 to 0.28	0.90	1.18 (0.86 to 1.50)	0.974

Appendix 8

TABLE 31 Variable loadings for PCs of prespecified predictors

Prespecified predictors	Component 1	Component 2	Component 3	Component 4
Maximum mural thickness (MEGS)	0.36	−0.24	0.1	−0.13
Maximum mural T2 signal (oedema) (MEGS)	0.4	−0.14	0.05	−0.03
Maximum contrast enhancement pattern	0.31	−0.15	0.07	0.38
Maximum length of disease (MEGS)	0.38	−0.22	−0.01	−0.12
Abscess (MEGS)	0.25	0.56	0.19	0.09
Maximum fat stranding (sMARIA)	0.36	−0.13	0.1	0
Number of abnormal segments	0.33	−0.2	−0.11	−0.39
Maximum upper tract and SB stricturing	0.2	−0.05	−0.2	0.67
Maximum colon stricturing	0.03	0.12	0.89	−0.03
Maximum upper tract and SB penetrating	0.29	0.52	−0.15	0.18
Maximum colon penetrating	0.22	0.44	−0.27	−0.43

Appendix 9

TABLE 32 Area under the curve of prognostic models for predicting development of MBDD within 5 years of diagnosis

Prognostic model	Area under the curve (95% CI)
B	0.68 (0.60 to 0.75)
A1	0.68 (0.60 to 0.75)
A2	0.68 (0.61 to 0.76)
A3	0.68 (0.61 to 0.76)
B1	0.70 (0.63 to 0.78)

Appendix 10

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories

	n (%)	Mean (SD)	Median (IQR)
All patients at 5 years			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	146 (75)	21,592 (25,741)	8988 (913–37,844)
Drug costs – biologics	87 (45)	34,051 (25,568)	32,961 (13,271–41,553)
Drug costs – not biologics	143 (74)	1328 (1905)	562 (225–1794)
Inpatient stays	62 (32)	9288 (27,765)	4500 (2078–7481)
Surgeries	41 (21)	8999 (10,238)	6025 (4992–8227)
Non-surgical day cases	30 (15)	1860 (3487)	643 (482–1130)
Outpatient visits	150 (77)	2124 (941)	1997 (1516–2666)
Imaging	151 (78)	1564 (1063)	1427 (736–2116)
Total costs	**194 (100)**	**24,267 (33,108)**	**10,136 (1856–41,092)**
Patients with disabling disease at 5 years (Beaugerie)			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	57 (70)	25,429 (23,718)	20,220 (3323–38,007)
Drug costs – biologics	43 (53)	31,566 (22,657)	32,093 (12,677–40,842)
Drug costs – not biologics	56 (69)	1645 (2438)	773 (280–2028)
Inpatient stays	43 (53)	10,994 (33,138)	4553 (2078–7481)
Surgeries	25 (31)	8896 (9919)	6025 (6025–8227)
Non-surgical day cases	9 (11)	950 (695)	581 (547–1382)
Outpatient visits	58 (72)	2490 (918)	2451 (1886–2891)
Imaging	59 (73)	1919 (1112)	1646 (1095–2402)
Total costs	**81 (100)**	**29,763 (38,278)**	**17,352 (00–45,969)**
Patients without disabling disease at 5 years (Beaugerie)			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	89 (79)	19,135 (26,798)	3620 (555–33,307)
Drug costs – biologics	44 (39)	36,480 (28,175)	33,640 (19,015–42,728)
Drug costs – not biologics	87 (77)	1125 (1444)	453 (176–1292)
Inpatient stays	19 (17)	5427 (5236)	2922 (2078–7813)
Surgeries	16 (14)	9160 (11,046)	6025 (4992–8521)
Non-surgical day cases	21 (19)	2250 (4112)	684 (468–886)
Outpatient visits	92 (81)	1893 (884)	1693 (1209–2262)
Imaging	92 (81)	1336 (971)	1118 (571–1847)
Total costs	**113 (100)**	**20,327 (28,368)**	**7293 (2495–30,976)**

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Male patients			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	63 (68)	22,313 (26,125)	9211 (802–39,449)
Drug costs – biologics	38 (41)	35,240 (25,436)	32,234 (15,029–46,748)
Drug costs – not biologics	61 (66)	1092 (1192)	660 (227–1333)
Inpatient stays	28 (30)	5182 (4359)	3267 (2078–6921)
Surgeries	15 (16)	7422 (3720)	6025 (4992–11,013)
Non-surgical day cases	15 (16)	1987 (3214)	684 (482–1462)
Outpatient visits	67 (72)	2095 (925)	1968 (1527–2633)
Imaging	68 (73)	1593 (1183)	1348 (746–2148)
Total costs	**93 (100)**	**20,867 (27,914)**	**6311 (00–36,457)**
Female patients			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	83 (82)	21,045 (25,591)	8087 (943–37,409)
Drug costs – biologics	49 (49)	33,129 (25,894)	33,278 (12,677–40,842)
Drug costs – not biologics	82 (81)	1505 (2289)	466 (225–1938)
Inpatient stays	34 (34)	12,670 (37,193)	4695 (2376–7481)
Surgeries	26 (26)	9910 (12,553)	6025 (4992–6323)
Non-surgical day cases	15 (15)	1732 (3850)	626 (468–840)
Outpatient visits	83 (82)	2147 (958)	2012 (1370–2797)
Imaging	83 (82)	1540 (961)	1427 (703–2094)
Total costs	**101 (100)**	**27,397 (37,122)**	**11,440 (4345–41,485)**
Patients diagnosed at age under 40 years			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	105 (76)	24,573 (25,755)	17,479 (992–39,622)
Drug costs – biologics	70 (51)	34,957 (24,415)	33,640 (16,917–47,860)
Drug costs – not biologics	103 (75)	1293 (1986)	526 (200–1451)
Inpatient stays	51 (37)	10,135 (30,527)	4637 (2078–7481)
Surgeries	30 (22)	9134 (11,784)	6025 (4992–7486)
Non-surgical day cases	22 (16)	2198 (4025)	643 (495–1130)
Outpatient visits	108 (78)	2115 (855)	2038 (1540–2676)
Imaging	109 (79)	1637 (1114)	1545 (749–2133)
Total costs	**138 (100)**	**27,727 (35,784)**	**12,031 (2572–42,604)**

continued

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients diagnosed at age over 40 years			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	*n* (%)	13,957 (24,374)	2676 (669–25,409)
Drug costs – biologics	41 (73)	30,320 (30,411)	29,046 (8671–37,679)
Drug costs – not biologics	17 (30)	1420 (1698)	729 (226–2425)
Inpatient stays	40 (71)	5365 (4710)	4455 (2376–7806)
Surgeries	11 (20)	8632 (4038)	6025 (6025–12,051)
Non-surgical day cases	11 (20)	931 (664)	685 (475–1166)
Outpatient visits	8 (14)	2146 (1144)	1781 (1279–2658)
Imaging	42 (75)	1374 (905)	1171 (697–1686)
Total costs	42 (75)	**15,740 (23,512)**	**6441 (693–17,368)**
Patients non-smokers at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	97 (75)	20,406 (24,735)	6435 (717–37,409)
Drug costs – biologics	52 (40)	35,513 (23,995)	34,343 (20,480–50,571)
Drug costs – not biologics	95 (73)	1397 (2109)	562 (227–1934)
Inpatient stays	34 (26)	5248 (4631)	4500 (2494–6025)
Surgeries	26 (20)	8043 (8708)	6025 (6025–7481)
Non-surgical day cases	24 (18)	1752 (3468)	643 (488–1008)
Outpatient visits	100 (77)	2145 (980)	2039 (1521–2676)
Imaging	101 (78)	1616 (1092)	1493 (777–2154)
Total costs	**130 (100)**	**21,436 (27,726)**	**8227 (1764–37,418)**
Patients smokers at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	37 (79)	21,566 (24,119)	14,465 (917–37,089)
Drug costs – biologics	24 (51)	31,506 (23,167)	30,700 (12,974–41,197)
Drug costs – not biologics	37 (79)	1130 (1364)	432 (172–1320)
Inpatient stays	21 (45)	7324 (8150)	3561 (2078–9297)
Surgeries	12 (26)	11,987 (13,898)	6756 (4992–12,875)
Non-surgical day cases	5 (11)	2654 (4201)	767 (581–1382)
Outpatient visits	38 (81)	2123 (901)	1911 (1509–2891)
Imaging	37 (79)	1455 (1076)	1095 (643–1840)
Total costs	**47 (100)**	**26,455 (28,671)**	**16,184 (2647–45,935)**

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients smoking status missing at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	12 (71)	31,261 (37,094)	33,174 (2576–42,181)
Drug costs – biologics	11 (65)	32,694 (37,717)	33,278 (1268–40,751)
Drug costs – not biologics	11 (65)	1409 (1685)	584 (325–3338)
Inpatient stays	7 (41)	34,806 (81,439)	5051 (1780–7813)
Surgeries	3 (18)	5336 (597)	4992 (4992–6025)
Non-surgical day cases	1 (6)	468 (0)	468 (468–468)
Outpatient visits	12 (71%)	1954 (766)	1885 (1421–2469)
Imaging	13 (76)	1465 (808)	1646 (736–2180)
Total costs	**17 (100)**	**39,867 (65,749)**	**10,298 (3966–47,064)**
Patients no recent weight loss at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	94 (77)	19,295 (26,704)	4581 (717–33,104)
Drug costs – biologics	48 (39)	35,182 (28,354)	32,527 (8483–48,110)
Drug costs – not biologics	93 (76)	1344 (1980)	584 (282–1597)
Inpatient stays	35 (29)	6688 (7069)	4987 (2078–8998)
Surgeries	28 (23)	8480 (9469)	6025 (4992–9249)
Non-surgical day cases	16 (13)	1920 (3139)	622 (481–1422)
Outpatient visits	98 (80)	2050 (949)	1954 (1368–2655)
Imaging	97 (80)	1469 (1022)	1236 (643–2009)
Total costs	**122 (100%)**	**21,797 (28,759)**	**8444 (2072–34,326)**
Patients recent weight loss at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	36 (78)	25,876 (23,613)	25,685 (2725–40,598)
Drug costs – biologics	26 (57)	34,177 (20,747)	35,739 (22,425–40,751)
Drug costs – not biologics	35 (76)	1227 (1577)	471 (148–1934)
Inpatient stays	17 (37)	5000 (4525)	3561 (2169–4753)
Surgeries	8 (17)	12,181 (15,288)	6025 (4992–12,118)
Non-surgical day cases	13 (28)	1861 (4121)	684 (482–871)
Outpatient visits	36 (78)	2306 (973)	2176 (1600–2911)
Imaging	37 (80)	1729 (1028)	1623 (1024–2459)
Total costs	**46 (100)**	**27,938 (28,393)**	**18,815 (2658–47,110)**

continued

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients recent weight loss status missing at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	16 (62)	25,449 (24,282)	22,460 (4275–36,372)
Drug costs – biologics	13 (50)	29,625 (24,762)	32,422 (12,677–37,897)
Drug costs – not biologics	15 (58)	1471 (2230)	627 (80–2119)
Inpatient stays	10 (38)	25,681 (68,118)	4699 (1780–7813)
Surgeries	5 (19)	6817 (2390)	6025 (6025–6025)
Non-surgical day cases	1 (4)	886 (0)	886 (886–886)
Outpatient visits	16 (62)	2166 (803)	2091 (1647–2520)
Imaging	17 (65)	1747 (1344)	1646 (936–2193)
Total costs	**26 (100)**	**29,359 (54,098)**	**9752 (00–35,578)**
Patients no stricturing or penetrating disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	98 (75)	19,012 (22,536)	8928 (802–33,307)
Drug costs – biologics	57 (44)	30,318 (22,306)	31,291 (13,271–40,751)
Drug costs – not biologics	96 (74)	1407 (2034)	544 (231–2165)
Inpatient stays	38 (29)	9731 (35,158)	2844 (1780–5051)
Surgeries	16 (12)	5664 (2536)	6025 (3743–6902)
Non-surgical day cases	17 (13)	640 (338)	547 (468–661)
Outpatient visits	101 (78)	2082 (963)	1936 (1370–2655)
Imaging	101 (78)	1440 (899)	1294 (697–1978)
Total costs	**130 (100)**	**20,694 (31,642)**	**6950 (1764–34,326)**
Patients stricturing disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	27 (79)	27,180 (31,928)	8887 (917–40,207)
Drug costs – biologics	16 (47)	43,631 (30,880)	36,182 (22,620–67,800)
Drug costs – not biologics	27 (79)	1325 (1873)	614 (189–1597)
Inpatient stays	13 (38)	8640 (7140)	6025 (2376–12,647)
Surgeries	14 (41)	7747 (3896)	6025 (6025–11,013)
Non-surgical day cases	7 (21)	1091 (614)	840 (727–1462)
Outpatient visits	26 (76)	2316 (928)	2327 (1679–2891)
Imaging	27 (79)	1679 (1136)	1623 (703–2387)
Total costs	**34 (100)**	**31,406 (34,272)**	**16,609 (7274–43,488)**

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients penetrating disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	21 (70)	26,446 (30,262)	9614 (943–40,272)
Drug costs – biologics	14 (47)	38,304 (29,789)	37,798 (8294–57,047)
Drug costs – not biologics	20 (67)	956 (1207)	624 (195–1062)
Inpatient stays	11 (37)	8526 (8112)	6025 (4851–8329)
Surgeries	11 (37)	15,444 (18,005)	6025 (4992–12,051)
Non-surgical day cases	6 (20)	6211 (6411)	5462 (356–10,140)
Outpatient visits	23 (77)	2090 (862)	2080 (1640–2666)
Imaging	23 (77)	1970 (1500)	1808 (1037–2409)
Total costs	**30 (100)**	**31,656 (36,485)**	**13,195 (1856–47,183)**
Patients no perianal disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	128 (75)	19,138 (24,819)	6201 (775–33,093)
Drug costs – biologics	71 (42)	32,058 (25,818)	31,734 (8671–40,751)
Drug costs – not biologics	126 (74)	1377 (1959)	588 (235–1934)
Inpatient stays	49 (29)	9881 (31,120)	3561 (2078–6326)
Surgeries	31 (18)	8309 (9126)	6025 (4992–6323)
Non-surgical day cases	25 (15)	1121 (1932)	626 (482–871)
Outpatient visits	132 (78)	2107 (972)	1954 (1385–2672)
Imaging	132 (78)	1523 (1074)	1294 (670–2105)
Total costs	**170 (100)**	**21,757 (31,843)**	**9049 (1821–34,243)**
Patients perianal disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	18 (78)	39,042 (26,112)	39,122 (20,220–55,167)
Drug costs – biologics	16 (70)	42,896 (23,129)	38,982 (30,725–56,173)
Drug costs – not biologics	17 (74)	966 (1443)	225 (142–1194)
Inpatient stays	13 (57)	7056 (6172)	4753 (2078–10,303)
Surgeries	10 (43)	11,138 (13,471)	7483 (4992–11,013)
Non-surgical day cases	5 (22)	5555 (6730)	1708 (547–9541)
Outpatient visits	17 (74)	2259 (696)	2177 (1640–2666)
Imaging	18 (78)	1841 (997)	1706 (1076–2409)
Total costs	**23 (100)**	**43,703 (36,937)**	**46,753 (8881–73,224)**

continued

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients no severe endoscopic disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	99 (76)	17,595 (22,973)	4824 (741–33,040)
Drug costs – biologics	54 (42)	29,764 (23,937)	29,577 (7917–39,299)
Drug costs – not biologics	97 (75)	1388 (2014)	564 (303–1938)
Inpatient stays	41 (32)	11,074 (33,896)	4553 (2376–7481)
Surgeries	25 (19)	6734 (3359)	6025 (4992–7486)
Non-surgical day cases	19 (15)	1213 (2083)	547 (468–840)
Outpatient visits	102 (78)	2099 (949)	1941 (1509–2655)
Imaging	102 (78)	1460 (972)	1311 (643–1978)
Total costs	**130 (100)**	**21,156 (32,264)**	**7954 (1821–34,326)**
Patients ileocolonic disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	69 (73)	27,690 (26,604)	27,017 (3323–41,241)
Drug costs – biologics	48 (51)	37,735 (24,303)	36,913 (23,098–48,110)
Drug costs – not biologics	67 (71)	1483 (2135)	614 (227–2194)
Inpatient stays	34 (36)	13,209 (37,173)	4695 (2078–8998)
Surgeries	19 (20)	11,523 (14,472)	6025 (4992–11,017)
Non-surgical day cases	16 (17)	2748 (4625)	651 (513–1256)
Outpatient visits	72 (77)	2314 (937)	2229 (1679–2959)
Imaging	73 (78)	1805 (1114)	1702 (936–2409)
Total costs	**94 (100)**	**31,075 (39,998)**	**13,503 (2495–47,947)**
Patients ileal disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	49 (75)	13,271 (18,791)	1877 (555–21,435)
Drug costs – biologics	24 (37)	24,702 (20,471)	23,186 (5705–37,891)
Drug costs – not biologics	48 (74)	1196 (1829)	535 (328–983)
Inpatient stays	18 (28)	4753 (3582)	4008 (2376–6025)
Surgeries	15 (23)	7320 (3422)	6025 (4992–11,013)
Non-surgical day cases	11 (17)	844 (592)	661 (452–886)
Outpatient visits	51 (78)	2042 (1022)	1945 (1122–2633)
Imaging	50 (77)	1405 (1032)	1199 (578–1915)
Total costs	**65 (100)**	**15,836 (19,261)**	**8881 (1764–21,802)**

TABLE 33 Total unadjusted costs over 5 years from diagnosis, broken down according to all planned categories (*continued*)

	n (%)	Mean (SD)	Median (IQR)
Patients colonic disease at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	25 (81)	21,304 (31,783)	4824 (1309–33,040)
Drug costs – biologics	13 (42)	38,521 (35,955)	32,093 (19,847–39,616)
Drug costs – not biologics	25 (81)	1274 (1473)	592 (151–2347)
Inpatient stays	9 (29)	4350 (2291)	4164 (2494–5051)
Surgeries	6 (19)	5874 (2018)	6025 (4992–7481)
Non-surgical day cases	3 (10)	847 (764)	581 (252–1708)
Outpatient visits	24 (77)	1739 (673)	1619 (1213–2161)
Imaging	25 (81)	1199 (852)	1024 (591–1442)
Total costs	**31 (100)**	**21,975 (30,793)**	**6351 (1821–35,578)**
Patients upper digestive tract disease at diagnosis			
Omitted due to small patient numbers.			
Patients no steroid required at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	83 (73)	19,163 (23,663)	5147 (719–33,104)
Drug costs – biologics	46 (40)	32,180 (23,263)	31,890 (12,064–43,903)
Drug costs – not biologics	80 (70)	1379 (1858)	523 (169–2321)
Inpatient stays	31 (27)	6871 (7153)	3561 (2078–10,303)
Surgeries	24 (21)	8100 (9234)	6025 (4992–6756)
Non-surgical day cases	17 (15)	2762 (4466)	840 (468–1462)
Outpatient visits	87 (76)	1965 (898)	1816 (1279–2556)
Imaging	88 (77)	1346 (908)	1154 (591–1837)
Total costs	**114 (100)**	**20,476 (26,498)**	**7954 (1463–34,243)**
Patients steroid required at diagnosis			
Total unadjusted costs over 5 years from diagnosis, £			
Drug costs	63 (79)	24,791 (28,120)	14,796 (1309–40,108)
Drug costs – biologics	41 (51)	36,151 (28,074)	36,795 (16,917–40,842)
Drug costs – not biologics	63 (79)	1265 (1977)	564 (308–1320)
Inpatient stays	31 (39)	11,706 (38,784)	4553 (2078–6025)
Surgeries	17 (21)	10,269 (11,685)	6025 (6025–11,013)
Non-surgical day cases	13 (16)	680 (364)	547 (495–661)
Outpatient visits	63 (79)	2343 (962)	2177 (1687–2813)
Imaging	63 (79)	1868 (1191)	1646 (1039–2387)
Total costs	**80 (100)**	**29,668 (40,290)**	**12,126 (2940–44,438)**

TABLE 34 Statistical significance of covariates in best-fit model for 5-year total costs

	p-value	Coefficient	(95% CI)
Beaugerie category (no)			
Disabling disease at 5 year	0.002	0.6979	0.2566 to 1.1392
Age category (< 40 years)			
Over 40 years at diagnosis	0.237	−0.2939	−0.7807 to 0.1929
Sex (male)			
Female	0.081	0.3627	−0.0445 to 0.7700
Smoker (non)			
Smoker at diagnosis	0.683	0.0961	−0.3653 to 0.5575
Weight loss (no)			
Recent weight loss at diagnosis	0.559	0.1385	−0.3263 to 0.6033
Behaviour of disease (neither)			
Stricturing	0.106	0.4482	−0.0953 to 0.9917
Penetrating	0.383	0.2678	−0.3344 to 0.8701
Perianal (no)			
Perianal disease at diagnosis	0.009	0.8667	0.2132 to 1.5201
Endoscopic disease (no)			
Severe endoscopic disease at diagnosis	0.143	0.4452	−0.1500 to 1.0403
Location of disease (ileocolonic)			
Ileal	0.056	−0.4345	−0.8806 to 0.0116
Colonic	0.071	−0.5768	−1.2035 to 0.0500
Upper digestive tract	0.842	0.1659	−1.4694 to 1.8011
Steroid (no)			
Steroid required at diagnosis	0.636	−0.1086	−0.5577 to 0.3406
Lemann (standardised 0–100)	0.381	−0.0850	−0.2751 to 0.1050
sMARIA (standardised 0–100)	0.353	0.0451	−0.0502 to 0.1405
_constant term	0.000	9.2694	8.7345 to 9.8042

TABLE 35 Statistical significance of covariates in alternative good fit model for 5-year total costs

	p-value	Coefficient	(95% CI)
Beaugerie category (no)			
Disabling disease at 5 year	0.049	0.4090	0.0011 to 0.8169
Sex (male)			
Female	0.019	0.5112	0.0837 to 0.9388
Behaviour of disease (neither)			
Stricturing	0.153	0.4501	−0.1668 to 1.0670
Penetrating	0.278	0.3376	−0.2727 to 0.9479
Perianal (no)			
Perianal disease at diagnosis	0.058	0.6667	−0.0214 to 1.3548
Endoscopic disease (no)			
Severe endoscopic disease at diagnosis	0.214	0.3228	−0.1866 to 0.8321
Location of disease (ileocolonic)			
Ileal	0.010	−0.6163	−1.0828 to −0.1497
Colonic	0.131	−0.4504	−1.0348 to 0.1339
Upper digestive tract	0.453	−0.5477	−1.9773 to 0.8819
Constant term	0.000	9.5432	9.0842 to 10.0022

EME
HSDR
HTA
PGfAR
PHR

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

Published by the NIHR Journals Library