Prognostic biomarkers to identify patients likely to develop severe Crohn's disease: a systematic review

Steve Halligan,[1]* Darren Boone,[1]
Lucinda Archer,[2] Tariq Ahmad,[3] Stuart Bloom,[4]
Manuel Rodriguez-Justo,[5] Stuart A Taylor[1]
and Sue Mallett[1]

[1]Centre for Medical Imaging, University College London, London, UK
[2]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, UK
[3]Department of Gastroenterology, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[4]Department of Gastroenterology, University College Hospital, London, UK
[5]Department of Histopathology, University College Hospital, London, UK

*Corresponding author s.halligan@ucl.ac.uk

Scientific summary

Prognostic biomarkers to identify severe Crohn's disease
Health Technology Assessment 2021; Vol. 25: No. 45
DOI: 10.3310/hta25450

Scientific summary

Background and objectives

The research detailed in this monograph arose from the National Institute for Health Research Health Technology Assessment programme call HTA 14/210: 'Prognostic markers in early Crohn's disease'. The need for this research turns on the understanding that early administration of biological therapies to patients with Crohn's disease may prevent disease progression. Indiscriminate administration of biologicals to all patients with a new diagnosis of Crohn's disease would not, however, be sensible because these drugs are both expensive and associated with serious side effects in a small but significant number of patients. Administering biologicals to patients whose disease is destined to be indolent would necessitate extensive over-treatment. Therefore, the unmet need is for a tool that accurately predicts those patients with a new diagnosis of Crohn's disease whose disease is destined to become severe at some point in the future. Biological therapy could be targeted to this group of patients and over-treatment avoided elsewhere. This approach would also probably be cost-effective.

A 'prognostic biomarker' is a measurement made on a patient that predicts a future event, for example a biological process and/or health state. Biomarkers may also be 'diagnostic', that is their level represents a current health state, for example the presence or absence of active disease. Some biomarkers may be both diagnostic and prognostic. The primary objective of this research was to perform a comprehensive systematic review across all biomarker types used for Crohn's disease (e.g. clinical, serological, genetic, radiological, endoscopic and histological), with the aim of identifying those that display potential prognostic capability for prediction of subsequent severe disease.

Review methods

The review was registered (PROSPERO CRD42016029363) and the protocol published. Following a scoping search, a clinical researcher experienced in systematic review and multidisciplinary Crohn's disease management designed and performed the search, supervised by co-researchers, including statisticians with extensive prior experience of prognostic systematic reviews, methodology experts and disease experts. Using an inclusive strategy, our search string combined five clusters of terms to identify (1) Crohn's disease, (2) severe disease, (3) biomarkers, (4) terms to identify prediction/prognosis and (5) exclusion criteria (e.g. animal research and narrative reviews). Hand-searching was also adopted. The PubMed and EMBASE databases were searched from inception to 1 January 2016, with an update carried out up to 1 January 2018 for serological biomarkers. The target condition was human Crohn's disease, confirmed by standard criteria. There was no age restriction, with paediatric subgroups identified where possible, and no language restriction. We anticipated finding a plethora of potential biomarkers and set a priori quality/quantity thresholds to prevent including biomarkers with insufficient primary evidence for sensible meta-analysis, stipulating that predictors should be reported in at least five individual studies. We excepted new but exceptionally promising predictors, which were chosen by an expert panel from a list of all candidates identified by the review. The panel also indicated predictors that were already widely used. Given that no generally accepted definition of 'severe' disease exists, we used a broad range of criteria to avoid discarding potentially valuable research. We stipulated a minimum 3-month duration between biomarker measurement and outcome to ensure that the data were prognostic rather than diagnostic.

Full-text articles were examined by Darren Boone and queries were raised with other collaborators. Ultimately, Sue Mallett examined all full-text articles selected for the review and verified data extraction. Data were extracted into a data sheet that included author, journal, design (e.g. cohort,

randomised controlled trial and retrospective database), methods, setting/context (organisation/service type and country), participants (including age range and gender), time since diagnosis (symptom duration and/or time since diagnosis for established disease), markers of severe disease, symptom severity, disease location and burden, disease complications, Crohn's activity indices, surgical details, perianal disease and continence outcomes. In addition, we extracted data specific to the prognostic marker investigated, including measurement (methods, frequency), adverse events, reliability and reproducibility, and costs, where available. We recorded study interventions and outcomes (including definitions, thresholds for severity/remission and whether or not these thresholds were prespecified in the publication) and median follow-up time with interquartile range. For models, we documented study type (development, internal/external validation), included predictors (including measurement methods, categorisation of continuous outcomes, blinding to outcome assessment and predictor variables), sample size (number of participants with events and included in modelling), statistical modelling methods (including fitting, missing data treatment and methods to adjust for overfitting), model performance (discrimination, calibration, sensitivity, specificity, net benefit and reclassification), model estimates and 95% confidence intervals (e.g. univariable unadjusted or adjusted estimates for predictors, adjusted coefficients for predictors in multivariable models). Where available, we extracted estimates with 95% confidence intervals including odds ratios, risk ratios and hazard ratios. The risk of bias for each included study was assessed via PROBAST (Prediction study Risk Of Bias ASsessment Tool). Domains were combined to give an overall risk of bias.

Where possible, participants with and without severe disease were extracted into 2 × 2 contingency tables. Studies were grouped for meta-analysis by effect estimates (e.g. odds ratios and hazard ratios) and meta-analysis was considered when there were three or more individual studies of a biomarker from which data using the same effect estimate could be extracted. We searched for predictor association with subsequent severe disease rather than exact estimates of strength or interpredictor comparisons. Studies were grouped for meta-analysis by prediction estimate type and by definition of severe disease, where possible. We grouped studies that were 'adequately adjusted', defined by clearly reported adjustment factors. Each study was included once per meta-analysis; where studies reported more than one estimate from the same participants using different definitions of severe disease, we selected the highest ranking. For example, estimates reporting surgery are preferentially included for meta-analysis as a higher-ranked estimate. Random-effects inverse variance meta-analysis was used to pool odds ratios. Meta-analysis was performed using 'metan' (Stata 14, StataCorp LP, TX, USA) because 2 × 2 contingency tables were not always available. Results were analysed across all ages, with groupings indicated on forest plots.

Results

A total of 29,947 abstracts were identified: 15,923 duplicates were removed and 14,024 abstracts were screened. In total, 247 full-text articles were assessed. Ultimately, 71 articles were included in the review, describing 56 non-overlapping patient cohorts. This included the addition of C-reactive protein, which was deemed important by the expert panel (other predictors identified by the panel were either already included or yielded insufficient data for meta-analysis). Most studies were European (37/71, 52%), with 14 (20%) studies from the USA/Canada. Forty (56%) studies were multicentre. In total, 36 (51%) studies were prospective, 33 (46%) were retrospective and two (3%) were unclear. Of the 71 studies, 11 (16%) were paediatric only, 23 (32%) were adult only and 37 (52%) were mixed. Recruitment dates varied, with 22 (31%), studies not reporting dates; follow-up also varied (median 8 years, interquartile range 5–10 years, range 0.8–18 years). Studies frequently presented several different definitions of severe disease. Most of the studies were rated as being at high risk of bias in at least one domain. Accordingly, 'Overall risk of bias' was rated as being 'high' in 65 (92%) studies. Only three studies were rated as being at a 'low' risk of bias. Overall, we identified 12 individual predictors that were eligible for meta-analysis: three serological, one genetic and eight clinical. There were no radiological, endoscopic or histological predictors that met a priori criteria for inclusion.

We were able to meta-analyse seven clinical markers from 58 studies. These were Montreal disease behaviour, age, disease duration, disease location, smoking, sex and family history. Meta-analysis of 12 studies (4376 participants, 1551 events) found that B2 stricturing, B3 fistulating, and either severe or disabling disease predicted severe disease more powerfully than B1 inflammation alone. B2 and B3 Montreal disease behaviours were the strongest predictors of subsequent severe disease in our review (B2: odds ratio 4.09, 95% confidence interval 2.59 to 6.48; B3: odds ratio 6.25, 95% confidence interval 3.68 to 10.63). Age at diagnosis was examined in 43 studies (19,623 participants, 7010 events). Most studies categorised age using Montreal/Vienna thresholds, so data were greatest for three groups: < 17 years, 17–40 years and > 40 years. Overall, diagnosis at < 17 years was associated with lower risk of subsequent severe disease than other ages. Disease duration was examined in 14 studies (8690 participants, 1714 events). Meta-analysis found that increased disease duration was associated with significant risk of subsequent severe disease at all durations examined. Disease location was examined in 32 studies (10,877 participants, 4193 events). Studies were diverse regarding the segments and/or segment combinations described. We were able to analyse 'colonic disease alone' and 'any colonic disease' versus other locations. Overall, colonic disease alone conferred significantly lower surgery risk (odds ratio 0.42, 95% confidence interval 0.31 to 0.58). Similarly, any colonic disease (23 studies, 7373 participants, 3086 events) predicted lower surgical or B2/B3 risk than no colonic disease. Perianal disease (24 studies, 13,483 participants, 5510 events) was associated with significantly increased risk of subsequent severe disease overall (1.84 odds ratio, 95% confidence interval 1.29 to 2.62). A total of 34 studies reported smoking associations (11,475 participants, 5097 events); meta-analysis of 26 studies found that current smoking increased risk significantly (odds ratio 1.53, 95% confidence interval 1.30 to 1.79). Sex was examined in 35 studies (14,489 participants, 5350 events) and was not a significant predictor overall. Family history was examined in 18 studies (5687 participants, 1413 events). Meta-analysis found no consistent direction and no significant association (odds ratio 1.05, 95% confidence interval 0.81 to 1.36).

Three serological markers (anti-*Saccharomyces cerevisiae* antibodies, anti-flagellin antibodies and C-reactive protein) from 13 studies were meta-analysed. Anti-*Saccharomyces cerevisiae* antibodies and anti-flagellin antibodies showed potential prognostic association. Anti-*Saccharomyces cerevisiae* antibody-positive patients had significantly increased odds of developing severe disease (odds ratio 2.29, 95% confidence interval 1.31 to 3.99, six studies). Anti-flagellin antibody data were identified in five studies; meta-analysis from odds ratios did not reach significance, but additional evidence from hazard ratios indicated significant association. For C-reactive protein, data were identified from only three studies (the expert panel requested C-reactive protein to be included) and the results found no significant predictive association. Of the genetic markers, results for nucleotide-binding oligomerisation domain-containing protein 2 (*NOD2*) in at least one variant were reported in 17 studies, although one study reported only *p*-values and was excluded from meta-analysis. Meta-analysis of 16 studies (5407 participants, 2645 events) suggested higher risk with a *NOD2* variant gene overall (odds ratio 1.69, 95% confidence interval 1.43 to 2.00).

Conclusions

During this review, we identified seven previous systematic reviews investigating the association of mostly serological and some genetic biomarkers with severe Crohn's disease. Three of these did not separate primary studies of diagnosis and prognosis. One review identified two studies of predictive biomarkers, one of which was also identified as the only prognostic study by a second review. The remaining systematic review included three biomarker prediction studies, only one of which met our inclusion criteria.

In summary, despite prognostic research being declared the greatest unmet need in Crohn's disease research and being faced with a vast amount of potential research, we could identify relatively few prognostic studies; just 56 non-overlapping patient cohorts were identified. Although a multitude of biomarkers were investigated, relatively few were examined individually in sufficient depth and

breadth for meta-analysis. Although genetic and serological biomarkers are perceived by clinicians as 'cutting edge', the reality is that we could identify only three that appeared genuinely prognostic: two serological (increased risk of developing severe disease was associated with anti-*Saccharomyces cerevisiae* antibodies and anti-flagellin antibodies positivity) and one genetic (*NOD2* variant gene positivity conveyed increased risk of developing severe disease). At the time of writing, this should help reassure clinicians that lack of access to novel biomarkers is not a major clinical disadvantage. We were able to identify five clinical biomarkers that appeared prognostic; the following were associated with increased risk of developing severe disease: Montreal B2/B3 disease behaviour versus B1, diagnosis at > 17 years of age, increased disease duration, disease beyond the colon and smoking.

Limitations

Our study does have limitations, predicated primarily by the quality of existing data. As noted already, despite being faced with a vast amount of potential research, we could identify relatively few prognostic studies. This paucity of genuine prognostic studies raises the possibility that our research is missing some clinically useful predictors simply because they have not been sufficiently studied (and so were discarded by our a priori quality criteria). This was the case for novel genetic and serological markers, the large majority of which were not studied by a sufficient number of primary studies to be eligible. Moreover, we were surprised that many apparently common predictors were not examined in sufficient depth; faecal calprotectin is one example. Furthermore, of those studies we were able to include, the majority were at high risk of bias, which will exert an effect on their results and, therefore, on the findings from our meta-analysis. A major frustration was that studies frequently confounded diagnosis and prognosis, which meant that our extraction was problematic.

There is also considerable existing excitement around multigene assays [e.g. PredictImmune Ltd (Cambridge, UK)] to predict severe disease, but these had been studied insufficiently for inclusion in our review and so could not be meta-analysed. Their recommendation will require either reporting in sufficient primary studies to allow meaningful meta-analysis or the publication of a prospective study of methodological quality such that meta-analysis of smaller, less rigorous studies becomes unnecessary.

Future work

The issue now turns to how to apply this review to the benefit of individual patients. Using existing data sets to which we have access, we are currently developing and validating a multivariable prognostic model that will combine the predictive factors identified so as to provide an estimate of the risk of subsequent severe disease for patients presenting with a new diagnosis of Crohn's disease. For obvious reasons, the model will not be applicable to those patients who present with severe disease. Contingent on the results of this validation, it may be clinically useful to perform an external validation in UK centres. In the meantime, we suggest that considerably more prognostic research in this area is required. The quality of this research and any multivariable models that arise should be improved by enhanced methodological rigour and adherence to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) reporting guidelines.

Study registration

This study is registered as PROSPERO CRD42016029363.

Funding

Health Technology Assessment

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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 report