Journal of Pediatric Gastroenterology & Nutrition, Publish Ahead of Print

DOI: 10.1097/MPG.0000000000003793

The paediatric Crohn's disease morbidity index (PCD-MI); development of a tool to assess long-term disease burden using a data driven approach

James J Ashton[1,2], Abhilasha Gurung[2], Cai Davis[3], Eleanor G Seaby[1], Tracy Coelho[2], Akshay Batra[2], Nadeem Afzal[2], Sarah Ennis[1], and R Mark Beattie[2]

1. Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton, UK
2. Department of Paediatric Gastroenterology, Southampton Children's Hospital, Southampton, UK
3. NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, UK

Correspondence to- Dr James J Ashton Human Genetics & Genomic Medicine University of Southampton Duthie Building (Mailpoint 808) Southampton General Hospital Southampton SO16 6YD UK j.ashton@soton.ac.uk

Conflicts of interest The authors declare no conflicts of interest

Data availability All data are available as supplementary data

Funding JJA is funded by an NIHR clinical lectureship, by a GUTS UK trainee award and by an ESPR post-doctoral award. The study is supported by the Southampton Biomedical Research Centre Data Science theme. There was no specific funding for this study.

Abstract

Background/Objective-Heterogeneity and chronicity of Crohn's disease (CD) make prediction of outcomes difficult. To date, no longitudinal measure can quantify burden over a patient's disease course, preventing assessment and integration into predictive modelling. Here, we aimed to demonstrate the feasibility of constructing a data driven, longitudinal disease burden score.

Methods-Literature was reviewed for tools used in assessment of CD activity. Themes were identified to construct a paediatric CD morbidity index (PCD-MI). Scores were assigned to variables. Data were extracted automatically from the electronic patient records at Southampton Children's Hospital, diagnosed from 2012 to 2019 (inclusive). PCD-MI scores were calculated, adjusted for duration of follow up and assessed for variation (ANOVA) and distribution (Kolmogorov-Smirnov).

Results-Nineteen clinical/biological features across five themes were included in the PCD-MI including blood/faecal/radiological/endoscopic results, medication usage, surgery, growth parameters and extraintestinal manifestations. Maximal score was 100 after accounting for follow-up duration.

PCD-MI was assessed in 66 patients, mean age 12.5 years. Following quality filtering, 9528 blood/faecal test results and 1309 growth measures were included. Mean PCD-MI score was 14.95 (range 2.2-32.5), data were normally distributed (p=0.2) with 25% of patients having a PCD-MI <10. There was no difference in the mean PCD-MI when split by year of diagnosis, F-statistic 1.625, p=0.147.

Conclusions-PCD-MI is a calculatable measure for a cohort of patients diagnosed over an 8-year period, integrating a wide-range of data with potential to determine high or low disease burden. Future iterations of the PCD-MI require refinement of included features, optimised scores and validation on external cohorts.

Key words Prediction; Crohn's disease; Paediatric

What is known

- Disease burden is difficult to quantify in Crohn's disease (CD), partly due to considerable heterogeneity. Current tools capture single timepoint disease activity rather than longitudinal disease burden.

What is new

- In this study we use a data-driven approach to create the first iteration of a pragmatic longitudinal disease burden score, the paediatric CD morbidity index (PCD-MI).
- We envisage future iterations of the PCD-MI becoming a useful tool for incorporation into disease prediction research, with the next steps to refine the included features, optimise scores and validate on external cohorts

Introduction

Crohn's disease (CD) is a relapsing and remitting, lifelong condition primarily affecting the intestine. It is a highly heterogenous condition and the clinical disease course is hugely variable between patients. Paediatric onset patients will have disease for their entire adult lives, with significant variation in disease burden. Many effective therapies exist, however targeted therapy and prediction of outcomes remains mostly theoretical. To date, trials and prospective predictive models have focused on single timepoint outcomes, commonly steroid-free remission at 52 weeks, the need to step-up therapy or occurrence of complicated disease during follow-up[1–4]. Recently, there have been renewed calls to prioritise longitudinal assessment of disease to facilitate and promote precision medicine for patients with inflammatory bowel disease[5]. Real world data has the potential to guide predictive models, including application of artificial intelligence algorithms to predict long-term outcomes and enable optimisation of treatment with a realistic and evidenced discussion of potential risks[6]. Single time point disease activity scores are well established as clinical and research tools, both to define remission and to guide escalation of therapy[7]. However, these only reflect a snapshot of disease burden. Longitudinal cohort studies aiming to determine predictors of disease activity are currently restricted to pragmatic proxies of long-term disease burden, with longitudinal disease burden scores currently lacking for CD. Well-defined and reproducible long-term assessment of disease burden has the potential to improve patient stratification and aid with discovery of clinical and molecular predictors of disease[8]. Recently it has been suggested that long-term disease activity and severity assessment should include three domains: impact of disease on a patient, measurable inflammatory burden and disease course (including complications)[9].

There is no current benchmark for the assessment of longitudinal disease. In this study we aimed to demonstrate what is feasible with a data driven approach to creating a pragmatic longitudinal disease burden score. We term this developmental tool, the paediatric CD morbidity index (PCD-MI), and include evidence from current activity scores and contemporary guidelines. We provide preliminary testing of the PCD-MI on a cohort of paediatric patients with variable follow-up times and assessed the ability of the tool to discriminate between high and low disease burden.

Methods

Review of existing evidence

Tools utilised for assessing disease activity in paediatric and adult-onset CD were retrieved through structured review of the literature using the search strategy-

(Crohn's disease) AND (score) OR (tool) OR (index) AND (activity) in the Medline database (Accessed 26/07/2022)

Clinical parameters from these tools assessing objective long-term measures of disease activity were extracted and utilised to develop the disease burden tool. Consensus ECCO-ESPGHAN paediatric CD guidelines were reviewed for treatment and management strategies reflecting differing disease courses[10]. Long-term treatments and measures used to monitor response were retrieved and incorporated into the disease burden assessment tool.

Clinical data extraction

Patients were identified from Southampton Children's hospital (SCH) paediatric IBD database. SCH cares for patients aged <18 years with inflammatory bowel disease (IBD) referred from 12 district general hospitals in the South of England. To be included, patients had to have a diagnosis of CD, live locally to Southampton, and have a Southampton (SO) postcode (to ensure

blood and stool monitoring were performed at SCH rather than in a referring/shared care unit), and have a minimum follow-up time of 2 years.

Clinical data were automatically extracted from electronic health records (EHR) through partnership with the University Hospital Southampton Digital and Southampton Biomedical Research Centre Data Science. Duration of follow-up from diagnosis until last clinical contact was calculated for all patients. Based on features implicated through review of the existing evidence the following results were retrieved from the EHR.

Longitudinal blood and stool results

All results underwent quality control measures. Measures that were biologically implausible, reported an insufficient sample, failed test or sample not received, or that were haemolysed were excluded from downstream analyses as previously described[11]. All measures were retrieved with age- and sex-specific normal ranges.

Measures, including C-reactive protein, albumin, haemoglobin, platelets and faecal calprotectin were individually plotted overtime using the ggplot2 package in 'RStudio' (build 492). For each measure, patients were extracted and plotted, and those with less than 5 data points for that measure were excluded. Using the abnormality flags for each test result (based on each test's normal range), individual patients were classified into three disease activity groups for each measure; 1) Initial abnormal (and normalises by 1 year), 2) Persistently normal (including at diagnosis) and 3) Relapsing and remitting abnormality (over disease course). No patients had <1 year of follow-up. We would recommend that any future inclusion of patients with <1 year of blood/faecal results should be assigned to group 1 if their results had normalised, or group 3 if their results showed ongoing inflammation.

Patients were annotated with their disease activity group and each group was plotted using ggplot2 to illustrate the trajectory of results over time. This was conducted for each blood/stool measure. Blood results included from our cohort represent the data available for our cohort, as results are included as longitudinal patterns to determine disease trajectory it would be possible to calculate a score using additional or alternative blood results (such as erythrocyte sedimentation rate) that could be tailored for a cohort, so long as these results represented inflammation or disease activity.

Disease behaviours

Endoscopy, small bowel magnetic resonance imaging (MRI), abdominal ultrasound, clinic letters and CT abdomen scan reports were retrieved from the University Hospital Southampton electronic health record (EHR). These records, including clinic letters, imaging reports and endoscopy reports, were electronically searched for stricturing (Fibrosis, Fibrotic, Stricture, Stricturing, Narrowing, Narrowed, Pre-stenotic dilatation, Stenotic, Reduced diameter) and fistulating (fistulating, fistula, fistulae, fistulising, fistulizing, penetrating, penetrate, penetrative, connection, connect, connecting) disease keywords to reduce the number of reports requiring clinical curation. Records without key words were recorded as a not-stricturing or non-penetrating phenotype, and the remaining were manually checked by a clinician to assign patients as having the correct disease behaviour.

Medications and surgery

Electronically held clinic letter and infusion records were manually checked for each patient to determine which long-term medications they had been prescribed during the follow-up period. Long-term medication was coded as a binary outcome for individuals and does not reflect starting or stopping therapies.

Surgical procedures were extracted from the EHR. All operation notes are stored in a standardised format. The name of the operation was retrieved and classified by a clinician (JJA) as an intestinal resection, perianal procedure (including seton placement, abscess drainage but excluding examination under anaesthesia) or an operation unrelated to CD.

Growth measures

Weight and height standard deviation scores (SDS) were retrieved from the EHR from diagnosis to the most recent follow-up time. SDS underwent quality control as previously described[12]. Change in SDS from diagnosis to most recent measure was calculated. Patients with an SDS at most recent follow-up <2.0 were flagged.

Extra intestinal manifestations (EIMs)

Clinic letters and discharge summaries from the EHR were manually checked for each patient to determine doctor-diagnosed extraintestinal manifestation of CD and concurrent autoimmune disease. Specifically, the presence of liver (concurrent autoimmune hepatitis, autoimmune sclerosing cholangitis and primary sclerosing cholangitis), skin (pyoderma gangrenosum, erythema nodosum),eye (uveitis, iritis) and arthritis (peripheral and axial) were assessed.

Calculation of PCD-MI

Extracted data were summarised for each patient. Based on longitudinal outcome domains identified from previous activity indexes and published guidance, scores were assigned to outcomes within each of the domains. The relative importance of clinical factors within the score was reflected in the assigned value for that clinical variable. These data were derived from previous activity index scores which attributed values to clinical parameters, where available.

Modification of scores based on follow-up duration

Based on the presumption that patients with longer disease duration are more likely to develop complications, escalation of therapy, occurrence of surgery and extraintestinal manifestations (EIMs)[13–15], a follow-up duration modification co-efficient was calculated to upweight high disease burden occurring within shorter follow-up times. The mean PCD-MI score was calculated, per year of follow-up and this value was divided by the mean overall PCD-MI score. Scores were banded by follow-up duration and the factor needed to normalise the PCD-MI score for that follow-up duration was calculated. This value was used as the follow-up duration modification co-efficient. As no patients with <1 year of follow-up were included the co-efficient for this group was extrapolated from the included data. Supplementary data 1, http://links.lww.com/MPG/D130.

Assessment of the PCD-MI

Through inclusion of additional features in the PCD-MI such as long-term inflammation (blood results, faecal calprotectin), complications and the need for surgery, we aimed to make scores from different standard clinical practice timepoints comparable. To assess this, we compared the scores of patients diagnosed in each year of study, accounting for variable follow-up time using the follow-up coefficient, and assessed for differences using ANOVA.

Ethics

This study was deemed to be a service evaluation and registered at University Hospital Southampton.

Results

Construction of PCD-MI

Tools which assess disease burden

The paediatric CD activity index score (PCDAI), and modifications of the PCDAI, wPCDAI, shPCDAI, abbrPCDAI, the Harvey-Bradshaw index (HBI) and the CD activity index (CDAI) were identified from the literature[7,16–18]. From these tools features that would be applicable over time were extracted, table 1. Measures reflecting transient proxies of disease activity, such as number of stools per day, abdominal pain on a specific day and general wellbeing on a specific day, were excluded as they do not reflect long-term disease burden[9]. Tools and scores requiring contemporaneous interpretation of endoscopic or radiological data were not included.

Themes assessing long-term outcomes

Features assessing longitudinal clinical features from each tool (PCDAI, HBI and CDAI) were be classified by theme; ‘medication and treatments, ‘disease complications’, ‘extraintestinal manifestations’, ‘investigation results’, and ‘growth measures’. Table 1.

All themes were then cross-referenced with the ECCO-ESPGHAN guidelines to include additional data pertinent to long-term disease burden, this included specific information on treatment algorithms, disease behaviour, investigations and growth. This resulted in a list of disease burden domains to include in the PCD-MI which included blood results, faecal calprotectin results, medication, complications, surgery, growth and extra-intestinal manifestations[10].

Features included in the PCD-MI

Clinical data from disease burden domain were used to populate the PCD-MI. Features within domains were assigned values reflecting proxies of long-term disease burden. These values were

then scaled to reflect an overall score with a minimal value of 0 and a maximal value of 50. Table 2. It is possible to achieve a score above 50 if more than 4 EIMs of IBD are present.

For longitudinal data (blood results, faecal calprotectin) each patient's data were modelled over their follow-up duration and classified into one of three groups, as described above- 1) Initial abnormal (and normalises), 2) Persistently normal (including at diagnosis) and 3) Relapsing and remitting abnormality (over disease course).

Growth measures were converted to standard deviation scores. For growth we defined a significant disease burden as a fall in weight or height of >1.0 SDS from baseline, or no improvement from SDS <-2.0 at baseline. This allowed the score to account for patients whose natural growth potential was lower, whilst identifying patients lying outside 2 standard deviations or with long-term growth deficits following diagnosis.

Use of specific medications, occurrence of complicated disease (stricturing or fistulating disease), occurrence of surgery and extra-intestinal manifestations, were included as binary outcomes, i.e. if they had occurred for a patient they were scored for that patient.

Modification of scores based on follow-up duration

We implemented a follow-up duration modification to the basic PCD-MI score as described. Follow-up duration was based on normalised mean PCD-MI for each year of follow-up. Comparing this value to the mean PCD-MI score for the whole cohort allowed a per year normalisation factor to be calculated. After normalising this to 10 years of follow-up we then banded the years of follow-up together to give the following coefficients- <1 year (PCD-MI score x 2), 1-3 years (PCD-MI score x 1.5), 5-10 years (PCD-MI score x 1.1), and >10 years (PCD-MI score x 1). Supplementary data 1, http://links.lww.com/MPG/D130.

Calculation of PCD-MI

We calculated the PCD-MI scores for a cohort of patients with automatically extracted follow-up data.

Testing cohort- patient identification and data extraction

All patients diagnosed between 01/01/2012 and 31/12/2019 and living within Southampton were included. Seventy patients had available data. Following exclusion of patients with less than 5 blood result data points across their follow-up period, a total of 66 patients were included in the testing cohort, mean age 12.5 years, 24 female (36.4%). Following exclusion of physiologically implausible values, haemolysed and clotted samples this data set included 2078 CRP measures, 2437 haemoglobin measures, 2434 platelet measures and 2367 albumin measures. Each patient was assigned to one of the three longitudinal groups (groups 1, 2 and 3) for each blood result measurement. Figure 1A-H.

There were 212 faecal calprotectin measures available. All patients had at least one faecal calprotectin result available and were assigned to one of the three longitudinal groups. As expected, all patients were assigned to normalisation after diagnosis or persistent relapsing and remitting abnormality, with no patients having a normal calprotectin at diagnosis. There was significant sparsity of data for faecal calprotectin, reflecting the initial diagnosis for some patients being >10 years ago. Despite this, any abnormal results since diagnosis allowed classification into the persistent relapsing and remitting abnormality group, with all other patients falling into the normalised or normal throughout groups, figures 1I-J.

Growth measures

A total of 1309 individual growth measures were recorded, consisting of 604 height measures, 705 weight measures. Following exclusion of physiologically implausible values, 601 height measures and 702 weight measures were included in the analysis.

No patients had a SDS <-2.0 at diagnosis and showed no growth improvement, 3 patients had a height SDS <-2.0 and 10 patients had a weight SDS score of <-2.0. Three patients dropped height SDS by >1.0 SDS during follow-up, but no patients dropped weight by >1.0 SDS. Supplementary data 1, http://links.lww.com/MPG/D130.

PCD-MI calculation for testing cohort

Using the values assigned to each clinical data measure we calculated the PCD-MI score for each patient. The mean PCD-MI score prior to application of the follow-up co-efficient was 13.06. After application the mean PCD-MI score was 14.95. Scores were then plotted on a histogram to demonstrate the distribution within this cohort. Figure 2A. The distribution of the PCD-MI without follow-up coefficient data was normal according to the Kolmogorov-Smirnov statistic 0.106, p=0.062. Each score was then multiplied by the follow-up coefficient for that patient, and replotted as a histogram, figure 2B. The distribution of the PCD-MI with follow-up coefficient data was normal according to the Kolmogorov-Smirnov statistic 0.069, p=0.2.

Assessment of the tool

Patients were split by year of diagnosis and the mean PCD-MI score, with follow-up coefficient, were calculated for each year, supplementary table 1, http://links.lww.com/MPG/D130. Following analysis with ANOVA, no significant differences were shown between the year groups, demonstrating comparability of PCD-MI scores across the whole cohort, regardless of year of diagnosis.

The highest total possible PCD-MI score for our test cohort was 75, as the minimum follow-up time for an individual was 2.75 years. The highest score we observed was 32.5, with the lowest 2.2. A total of 17 patients (25.8%) had a PCD-MI score of ≤10, mean follow-up 6.2 years (range 2.7-10.5 years), demonstrating long-term quiescent disease. Similarly, 15 patients (22.7%) had scores ≥20, mean follow-up 5.9 years (range 2.9-9.7 years), demonstrating a higher burden of longitudinal disease. Importantly for a representative score across a cohort, most patients (51.5%) fell between a PCD-MI score of 10 and 20, reflected by the normal distribution of the scores.

Discussion

Using a data driven approach and incorporating key longitudinal disease variables from previously validated indices and guidelines we have created the first iteration of a disease burden tool for paediatric CD, capable of assigning a long-term score based on readily available clinical data. This score provides a potential mechanism for prediction of long-term disease outcomes for incorporation into disease prediction models, and represents a staging post in the development of a collaborative and evidence-driven burden score. Further development, rationalisation and testing will provide the potential for the PCD-MI to compare patients with different follow-up points and treated at different timepoints, as it utilises numerous clinical data inputs, reflecting conventional practice continuing to change over time. For an individual patient it would also be possible to assess PCD-MI over time through calculation at intervals. To our knowledge no other composite longitudinal scores of disease burden exist for clinical or research purposes, and this study presents a demonstration of what is feasible in the era of 'big clinical data'. It is not intended to replace current, validated, disease scores, such as PCDAI, that are used to assess

single timepoint disease activity, and is intended as a first step towards a consensus disease burden tool.

Given the numerous different elements of the score, and the heterogeneity of CD, not many patients were able to achieve very high scores as different phenotypic subgroups are less likely to have overlapping features with other groups, for example penetrating and stricturing disease[13]. This is a comparative strength of the PCD-MI, allowing patients to achieve scores for disease burden from a variety of different disease processes. It also highlights that a number of patients, 25% in this cohort, have long-term quiescent disease which is similar to previously reported by Wintjens *et al,* who described 28.2% of adult patients being classified to a long-term quiescent cluster[19]. Identifying these patients at the point of diagnosis would present huge potential to avoid significant therapy, minimise side effects and costs, whilst also not exposing patients to long-term risk of developing disease complications.

The need for long-term disease assessment have been highlighted by a number of consortia focused on precision medicine in inflammatory bowel disease[20,21]. Where previous prediction studies in CD have used pragmatic proxies of long-term disease activity, such as the need to step up therapy, or specific disease complications, there is a clear requirement to more accurately phenotype longitudinal disease[1,2]. An example of this is reflected by a number of studies utilising surgery as a proxy for severe disease[22,23]. Whilst occurrence of surgery is clearly an important factor in disease burden, for some patients with CD an isolated right hemicolectomy may occur after a long period of quiescent disease and result in further long-term well controlled disease. The PCD-MI allows reflection of the total disease burden, rather than relying on a single measure of disease activity over time.

Similar to our study, previous efforts have been made to classify long-term disease activity. Cosnes and colleagues describe a consensus method for discriminating severe from a mild-to-moderate CD course in adult patients, however this study did not utilise a data-driven approach and relied on expert opinion[24]. More recently Chen *et al* have described a longitudinal cluster analysis of patient's blood results to determine different clusters of disease activity related to infliximab treatment, reflecting a data-driven approach but limited only to blood results[25]. Jiang et al incorporated healthcare costs and disease metrics, including tracking HBI over time to determine subgroups of disease trajectories[26]. The methodology and data used to derive these differing longitudinal scores is largely reflected in our PCD-MI, with the added benefit that we are able to incorporate large amounts of clinical data collected over the follow-up of a patient. The main utility for the PCD-MI may be as a numerical score for incorporation into research, to predict long-term disease activity. Scores are likely to mean little if applied in isolation to a patient but are more useful when observed across a cohort. With the vast amount of data now being generated through initiatives such as the United Kingdom IBD BioResource and the international IBD genetics consortium, there is a need to improve the precision and usefulness of clinical data collected alongside these projects[27,28]. Without accurate clinical data, and disease burden metrics, to incorporate into prediction models, the clinical translation of these resources will remain limited. The largest potential may be for patients with long-term quiescent disease, where therapy can be minimised, prevented associated risks, reducing medication burden and costs, without the risks of disease complications or flares. We also point to the ability to include additional longitudinal blood or faecal results that are used to assess disease activity in a specific cohort or hospital, such as erythrocyte sedimentation rate, into an iteration of PCD-MI, that can then be used to compare disease burden longitudinally within a local cohort.

The study and PCD-MI score have strengths, but we also acknowledge potential weaknesses. A potential criticism of this score is that the values assigned to each clinical variable that comprise the score are not always evidence based. We accept that these values are often pragmatic opinion, however this strategy also reflects the initial PCDAI development[18]. In terms of blood and faecal pathology results, we recognise that retrospectively gathered patient data will frequently have missing data points, but it may be possible to interpolate results in some circumstances. However, through a longitudinal classification of the pattern of patient results there is also the ability to compensate for this. Another potential strength is the flexibility the score offers when including blood results. Different cohorts will have differing blood results available and the score could accommodate different tests or values that reflect disease activity or inflammation. We have also focused on paediatric-onset patients, however with minor adjustments this score could be utilised as an adult CD morbidity index (ACD-MI). The future inclusion of standardised measures of disease extent through endoscopic, histological and radiological tools would also be desirable [29,30]. To date, the retrospective scoring of these tools and the lack of standardisation of repeating radiological or endoscopic investigations was too limited to include in this iteration, however as electronic data capture improves then we would envisage these tools being incorporated into PCD-MI. It is also not possible to have retrospective measures of the impact of disease on a patient, which would be an important prospective modifier for the score. Some of the calculations require manual steps, requiring additional time to calculate the score. We acknowledge the rates of EIMs in this cohort is approximately 8%, lower than the estimated 20% generally reported in the literature[31]. Despite this our data are restricted to arthritis, liver disease, oral and dermatological manifestations, or another autoimmune complication of IBD, and we have not included aphthous stomatitis/ulcers, which

account for a large proportion of the described EIM in the literature[31]. The PCD-MI will require validation on external cohorts to assess its ability to translate across healthcare systems, however with the increasing use of electronic healthcare records the ability to calculate this score will become easier. Further modifications and refinements of the PCD-MI will be useful over time, including refinement of the data included, the weighting of individual clinical variables, and the calculation of follow-up coefficient. The result of the score is a continuous variable related to disease burden, rather than a categorical high/low burden. For the purposes of integration into subsequent data analyses we believe a numerical value is more useful. Validation on external cohorts, and expert opinion, could results in a future consensus on values deemed to reflect high/low disease burden.

Conclusions

This study uses an evidence and data-driven approach to demonstrate the feasibility of a longitudinal paediatric CD morbidity index. This score was calculated on a cohort of patients diagnosed over an 8-year period and assessed for variation between years of diagnosis and number of years of follow-up. We envisage future iterations of the PCD-MI becoming a useful tool for incorporation into disease prediction research, with the next steps to refine the included features, optimise scores and validate on external cohorts.

Figures and Suppl Content Legends

Figure 1- Blood and calprotectin results patients included in paediatric Crohn's Disease morbidity index (PCD-MI) validation plotted over the duration of follow-up, shown as collated results and individual graphs for separate patient groups- initially abnormal (red), persistently normal (green) and relapsing and remitting abnormality (blue). 1A) Haemoglobin collated results, 1B) Haemoglobin by patient group, 1C) C-reactive protein collated results, 1D) C-reactive protein by patient group, 1E) Albumin collated results, 1F) Albumin by patient group, 1G) platelets collated results, 1H) platelets by patient group, 1I) Faecal calprotectin collated results, 1J) Faecal calprotectin by patient group

Figure 2- Histograms of patients included in paediatric Crohn's Disease morbidity index (PCD-MI). validation, bin size 5 points. 2A) Distribution of patients prior to application of follow-up coefficient, 2B) Distribution of patients after the application of follow-up coefficient.

Supplementary table 1- Number of patients included in the PCD-MI validation cohort, included by year of study, with mean PCD-MI (with correction by follow-up coefficient). ANOVA indicates no significant difference between groups, F statistic 1.625, p=0.147

References

1. Lee JC., Lyons PA., McKinney EF., Sowerby JM., Carr EJ., Bredin F., et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *Journal of Clinical Investigation* 2011;**121**(10):4170–9. Doi: 10.1172/JCI59255.
2. Kugathasan S., Denson LA., Walters TD., Kim M-O., Marigorta UM., Schirmer M., et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *The Lancet* 2017;**389**(10080):1710–8. Doi: 10.1016/S0140-6736(17)30317-3.
3. Douglas GM., Hansen R., Jones CMA., Dunn KA., Comeau AM., Bielawski JP., et al. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome* 2018;**6**(1):13. Doi: 10.1186/s40168-018-0398-3.
4. Atia O., Kang B., Orlansky-Meyer E., Ledder O., Tzion RL., Choi S., et al. Existing Prediction Models of Disease Course in Paediatric Crohn's Disease Are Poorly Replicated in a Prospective Inception Cohort. *J Crohns Colitis* 2022;**16**(7):1039–48. Doi: 10.1093/ECCO-JCC/JJAC005.
5. Ricciuto A., Rauter I., McGovern DPB., Mader RM., Reinisch W. Precision Medicine in Inflammatory Bowel Diseases: Challenges and Considerations for the Path Forward. *Gastroenterology* 2022;**162**(7):1815–21. Doi: 10.1053/J.GASTRO.2022.02.049.
6. Brooks-Warburton J., Ashton J., Dhar A., Tham T., Allen PB., Hoque S., et al. Artificial intelligence and inflammatory bowel disease: practicalities and future prospects. *Frontline Gastroenterol* 2021;**0**:flgastro-2021-102003. Doi: 10.1136/FLGASTRO-2021-102003.

7. Ledder O., Turner D. Multi-item Measures for Pediatric Inflammatory Bowel Diseases: The ABCs of All Those Acronyms. *J Crohns Colitis* 2023. Doi: 10.1093/ECCO-JCC/JJAD019.

8. Sun H., Papadopoulos EJ., Hyams JS., Griebel D., Lee JJ., Tomaino J., et al. Well-defined and reliable clinical outcome assessments for pediatric Crohn disease: A critical need for drug development. *J Pediatr Gastroenterol Nutr* 2015;**60**(6):729–36. Doi: 10.1097/MPG.0000000000000793.

9. Peyrin-Biroulet L., Panés 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**(3):348-354.e17. Doi: 10.1016/J.CGH.2015.06.001.

10. van Rheenen PF., Aloi M., Assa A., Bronsky J., Escher JC., Fagerberg UL., et al. The Medical Management of Paediatric Crohn's Disease: an ECCO-ESPGHAN Guideline Update. *J Crohns Colitis* 2020;**2020**:1–24. Doi: 10.1093/ecco-jcc/jjaa161.

11. Ashton JJ., Borca F., Mossotto E., Phan HTT., Ennis S., Beattie RM. Analysis and Hierarchical Clustering of Blood Results Before Diagnosis in Pediatric Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2018. Doi: 10.1093/ibd/izy369.

12. Phan HTT., Borca F., Cable D., Batchelor J., Davies JH., Ennis S. Automated data cleaning of paediatric anthropometric data from longitudinal electronic health records: protocol and application to a large patient cohort. *Scientific Reports 2020 10:1* 2020;**10**(1):1–9. Doi: 10.1038/s41598-020-66925-7.

13. Louis E., Collard A., Oger AF., Degroote E., Aboul Nasr El Yafi FA., Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001;**49**(6):777–82. Doi: 10.1136/GUT.49.6.777.

14. Ashton JJ., Green Z., Kolimarala V., Beattie RM. Inflammatory bowel disease: long-term therapeutic challenges. *Expert Rev Gastroenterol Hepatol* 2019:1049–63. Doi: 10.1080/17474124.2019.1685872.

15. Vernier-Massouille G., Balde M., Salleron J., Turck D., Dupas JL., Mouterde O., et al. Natural history of pediatric Crohn's disease: a population-based cohort study. *Gastroenterology* 2008;**135**(4):1106–13. Doi: 10.1053/j.gastro.2008.06.079.

16. Best WR., Becktel JM., Singleton JW., Kern F. Development of a Crohn's Disease Activity Index: National Cooperative Crohn's Disease Study. *Gastroenterology* 1976;**70**(3):439–44. Doi: 10.1016/S0016-5085(76)80163-1.

17. Harvey RF., Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980;**1**(8167):514. Doi: 10.1016/S0140-6736(80)92767-1.

18. Otley A., Loonen H., Parekh N., Corey M., Sherman PM., Griffiths AM. Assessing activity of pediatric Crohn's disease: which index to use? *Gastroenterology* 1999;**116**(3):527–31. Doi: 10.1016/S0016-5085(99)70173-3.

19. Wintjens D., Bergey F., Saccenti E., Jeuring S., van den Heuvel T., Romberg-Camps M., et al. Disease Activity Patterns of Crohn's Disease in the First Ten Years After Diagnosis in the Population-based IBD South Limburg Cohort. *J Crohns Colitis* 2021;**15**(3):391–400. Doi: 10.1093/ECCO-JCC/JJAA173.

20. Verstockt B., Noor NM., Marigorta UM., Pavlidis P., Deepak P., Ungaro RC., et al. Results of the Seventh Scientific Workshop of ECCO: Precision Medicine in IBD—

Disease Outcome and Response to Therapy. *J Crohns Colitis* 2021;**15**(9):1431–42. Doi: 10.1093/ECCO-JCC/JJAB050.

21. Verstockt B., Parkes M., Lee JC. How Do We Predict a Patient's Disease Course and Whether They Will Respond to Specific Treatments? *Gastroenterology* 2022;**162**(5):1383–95. Doi: 10.1053/J.GASTRO.2021.12.245.

22. Burisch J., Lophaven S., Langholz E., Munkholm P. The clinical course of Crohn's disease in a Danish population-based inception cohort with more than 50 years of follow-up, 1962-2017. *Aliment Pharmacol Ther* 2022;**55**(1):73–82. Doi: 10.1111/APT.16615.

23. Hojsak I., Pavić AM., Mišak Z., Kolaček S. Risk factors for relapse and surgery rate in children with Crohn's disease. *Eur J Pediatr* 2014;**173**(5):617–21. Doi: 10.1007/s00431-013-2230-1.

24. Cosnes J., Bourrier A., Nion-Larmurier I., Sokol H., Beaugerie L., Seksik P. Factors affecting outcomes in Crohn's disease over 15 years. *Gut* 2012;**61**(8):1140–5. Doi: 10.1136/GUTJNL-2011-301971.

25. Chen A., Stein R., Baldassano RN., Huang J. Learning Longitudinal Patterns and Subtypes of Pediatric Crohn Disease Treated with Infliximab via Trajectory Cluster Analysis. *J Pediatr Gastroenterol Nutr* 2022;**74**(3):383–8. Doi: 10.1097/MPG.0000000000003370.

26. Jiang J., Click B., Anderson AM., Koutroubakis IE., Rivers CR., Hashash JG., et al. Group-Based Trajectory Modeling of Healthcare Financial Charges in Inflammatory Bowel Disease: A Comprehensive Phenotype. *Clin Transl Gastroenterol* 2016;**7**(7):e181. Doi: 10.1038/CTG.2016.39.

27. Parkes M. IBD BioResource: an open-access platform of 25 000 patients to accelerate research in Crohn's and Colitis. *Gut* 2019;**68**(9):1537–40. Doi: 10.1136/GUTJNL-2019-318835.

28. Chen G-B., Lee SH., Montgomery GW., Wray NR., Visscher PM., Gearry RB., et al. Performance of risk prediction for inflammatory bowel disease based on genotyping platform and genomic risk score method. *BMC Med Genet* 2017;**18**(1):94. Doi: 10.1186/s12881-017-0451-2.

29. Adler J., Eder SJ., Gebremariam A., French KR., Moncion I., Singer AAM., et al. Development and Testing of a New Simplified Endoscopic Mucosal Assessment for Crohn's Disease: The SEMA-CD. *Inflamm Bowel Dis* 2021;**27**(10):1585–92. Doi: 10.1093/IBD/IZAA307.

30. Focht G., Cytter-Kuint R., Greer MLC., Pratt LT., Castro DA., Church PC., et al. Development, Validation, and Evaluation of the Pediatric Inflammatory Crohn's Magnetic Resonance Enterography Index From the ImageKids Study. *Gastroenterology* 2022;**163**(5):1306–20. Doi: 10.1053/j.gastro.2022.07.048.

31. Jang HJ., Kang B., Choe BH. The difference in extraintestinal manifestations of inflammatory bowel disease for children and adults. *Transl Pediatr* 2019;**8**(1):45–15. Doi: 10.21037/TP.2019.01.06.

	Paediatric Crohn's disease activity index	Crohn's disease activity index (adult tool)	Harvey-Bradshaw index (adult tool)	ECCO-ESPGHAN 2020 guidelines
Medication and treatments	None	Anti-diarrhoeal drugs	None	Treatment algorithms with stepwise escalation through immunomodulation, biologics, and additional therapy such as surgery
Disease complications- fistulating, stricturing or abdominal masses	Includes fistulating and abdominal masses	Includes fistulating and abdominal masses	Includes fistulating and abdominal masses	Risk stratification of patient based on disease behaviour- B2 and B3 disease = 'high risk'
Extraintestinal manifestations	Yes- arthritis, uveitis, erythema nodosum, or pyoderma gangrenosum	Yes- arthritis or arthralgias, iritis or uveitis, erythema nodosum, pyoderma gangrenosum	Yes- arthralgia, uveitis, erythema nodosum, pyoderma gangrenosum	Not mentioned
Investigation results	Yes- Full blood count related measures, inflammatory markers and albumin	Yes- Full blood count related measures	None	Monitoring using faecal calprotectin and prospectively collected endoscopic measures of disease extent/severity- Simple Endoscopic Score for Crohn's Disease (SES-CD), or Crohn's Disease Endoscopic Index of severity (CDEIS)
Growth measures	Height and weight assessment	Assesses weight loss	None	Growth delay is associated with severe disease and 'medium risk' stratification

Table 1- *Summary of features common to the three major disease activity score indices- Paediatric Crohn's disease activity index, Crohn's disease activity index (adult tool), and the Harvey-Bradshaw index (adult tool), alongside the latest ECCO ESPGHAN paediatric Crohn's disease guidelines*

Disease burden domain	Components	Score
Longitudinal blood results	Haemoglobin group 1, 2 + 3* C-reactive protein group 1, 2 + 3* Platelets group 1, 2 + 3* Albumin group 1, 2 + 3*	0, 1, 3 0, 1, 3 0, 1, 3 0, 1, 3
Longitudinal calprotectin	Faecal calprotectin group 1, 2 + 3*	0, 1, 3
Medication	5-ASA/topical treatment Immunomodulator First line monoclonal Second line monoclonal or small molecule	1 2 3 4
Complications	Fistulating disease Stricturing disease	3 3
Surgery	Intestinal resection (ever) Perianal procedure (ever)	4 2
Growth	Fall of >1.0 weight SDS from baseline, dropping to below -2.0 SDS during follow-up or no improvement from SDS <-2.0 at baseline Fall of >1.0 height SDS from baseline, dropping to below -2.0 SDS during follow-up or no improvement from SDS <-2.0 at baseline	2 2
Extraintestinal manifestation	Autoimmune liver disease- Primary sclerosing cholangitis, autoimmune sclerosing cholangitis, autoimmune hepatitis Chronic skin disease- psoriasis, pyoderma gangrenosum etc. Autoimmune Eye disease- uveitis, episcleritis etc. Arthritis- peripheral or axial And/Or other autoimmune comorbidity related to IBD	3 2 2 2 (2)
		Min= 0 Max = 50 (+)

Table 2- *Clinical variables included in the paediatric Crohn's Disease morbidity index (PCD-MI).*

with the accompanying score. Follow-up coefficients related to duration of time since diagnosis

for which an individual patient has follow-up clinical data for, supplementary data 1.

*group 1 = persistently normal result, group 2 = initially abnormal and normalises over disease course, group 3 = relapsing and remitting abnormality over disease course

Follow-up time modification	Multiplicative factor
Score calculated <1 year	*2.0
Score calculated 1-3 years	*1.5
Score calculated 3-5 years	*1.25
Score calculated 5-10 years	*1.1
Score calculated >10 years	*1.0

Figure 1

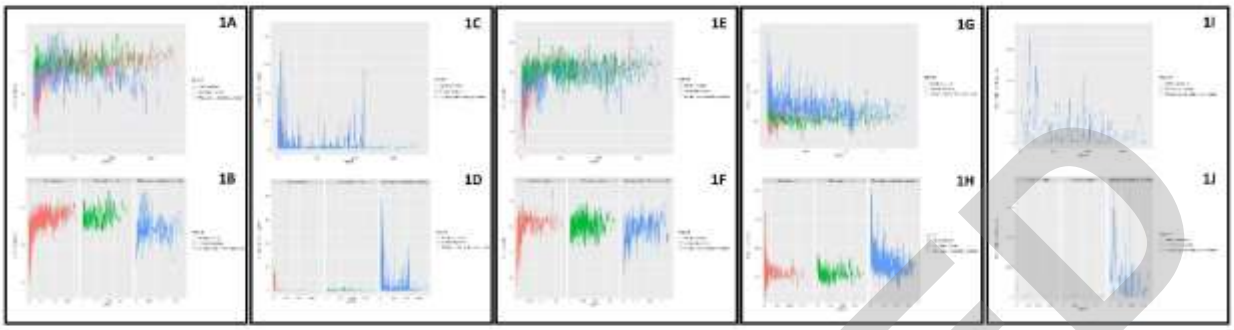

Figure 2