Characterizing the pre-clinical phase of inflammatory bowel disease

Vestergaard, Marie Vibeke; Allin, Kristine H.; Poulsen, Gry J.; Lee, James C.; Jess, Tine

Published in:
Cell reports. Medicine

DOI (link to publication from Publisher):
10.1016/j.xcrm.2023.101263

Creative Commons License
CC BY-NC-ND 4.0

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

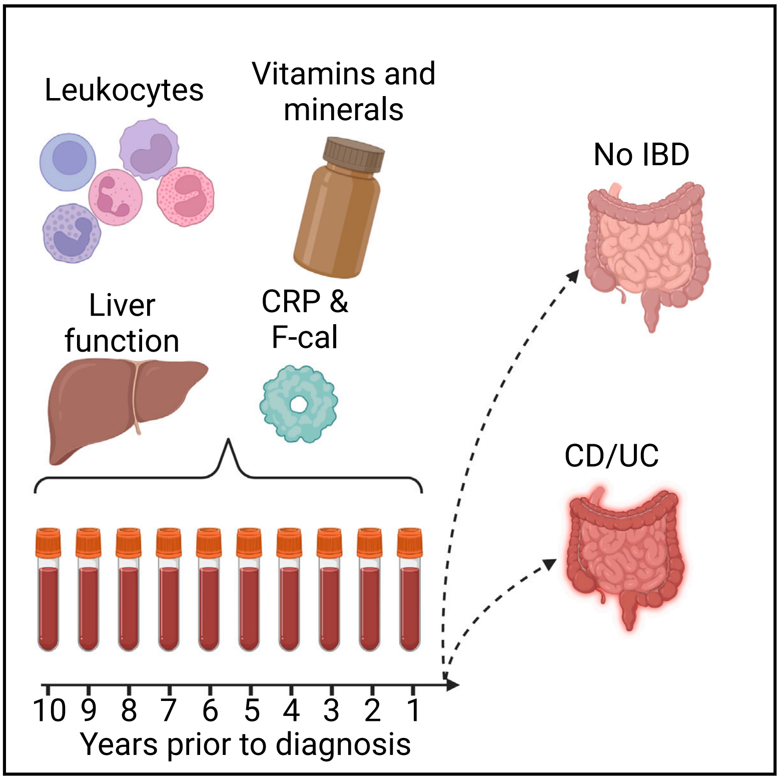
Citation for published version (APA):
Vestergaard, M. V., Allin, K. H., Poulsen, G. J., Lee, J. C., & Jess, T. (2023). Characterizing the pre-clinical phase of inflammatory bowel disease. *Cell reports. Medicine*, *4*(11), 101263. Article 101263. https://doi.org/10.1016/j.xcrm.2023.101263

Cell Reports
Medicine

Article

Characterizing the pre-clinical phase of inflammatory bowel disease

Graphical abstract



Highlights

- Early diagnosis of IBD enables timely interventions and improved clinical outcomes
- A pre-clinical phase of IBD is well recognized but poorly understood
- Biological changes occur up to 8 years before CD diagnosis and 3 years before UC diagnosis
- The ability of these changes to predict future IBD is modest

Authors

Marie Vibeke Vestergaard,
Kristine H. Allin, Gry J. Poulsen,
James C. Lee, Tine Jess

Correspondence

jess@dcm.aau.dk

In brief

Vestergaard et al. systematically characterize the pre-clinical phase of inflammatory bowel disease, delineating changes that occur in patients' blood many years before symptom onset and an eventual diagnosis. These hematological and biochemical changes reveal a pre-clinical disease phase whose duration far exceeds previous estimates.

Vestergaard et al., 2023, Cell Reports Medicine *4*, 101263
November 21, 2023 © 2023 The Authors.
https://doi.org/10.1016/j.xcrm.2023.101263

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Characterizing the pre-clinical phase of inflammatory bowel disease

Marie Vibeke Vestergaard,[1] Kristine H. Allin,[1,2] Gry J. Poulsen,[1] James C. Lee,[3,4,5] and Tine Jess[1,2,5,6,*]

[1]Center for Molecular Prediction of Inflammatory Bowel Disease, PREDICT, Department of Clinical Medicine, Aalborg University, Copenhagen, Denmark
[2]Department of Gastroenterology & Hepatology, Aalborg University Hospital, Aalborg, Denmark
[3]Genetic Mechanisms of Disease Laboratory, The Francis Crick Institute, London, UK
[4]Institute of Liver and Digestive Health, Division of Medicine, Royal Free Hospital, University College London, London, UK
[5]Senior author
[6]Lead contact
*Correspondence: jess@dcm.aau.dk

https://doi.org/10.1016/j.xcrm.2023.101263

SUMMARY

Understanding the biological changes that precede a diagnosis of inflammatory bowel disease (IBD) could facilitate pre-emptive interventions, including risk factor modification, but this pre-clinical phase of disease remains poorly characterized. Using measurements from 17 hematological and biochemical parameters taken up to 10 years before diagnosis in over 20,000 IBD patients and population controls, we address this at massive scale. We observe widespread significant changes in multiple biochemical and hematological parameters that occur up to 8 years before diagnosis of Crohn's disease (CD) and up to 3 years before diagnosis of ulcerative colitis. These changes far exceed previous expectations regarding the length of this pre-diagnostic phase, revealing an opportunity for earlier intervention, especially in CD. In summary, using a nationwide, case-control dataset—obtained from the Danish registers—we provide a comprehensive characterization of the hematological and biochemical changes that occur in the pre-clinical phase of IBD.

INTRODUCTION

The inflammatory bowel diseases (IBDs), ulcerative colitis (UC) and Crohn's disease (CD), are incurable inflammatory disorders of the gastrointestinal tract, which mainly occur in younger individuals.[1] The incidence of IBD is on the rise globally with 6.8 million IBD patients in 2017, an 85% increase from 1990.[2] Currently there is no cure for IBD, and while there are increasing numbers of treatments, these often fail to halt or reverse the progression of disease, resulting in the need for life-changing surgery.[3]

As with most autoimmune or inflammatory diseases, early diagnosis is critical to allow timely intervention and thereby improve treatment outcomes,[4] reduce surgery rates, and enhance overall quality of life.[5–7] Unfortunately, despite this ambition, a proportion of IBD patients will already have established bowel damage at the time they are diagnosed, even though only a minority will have experienced gastrointestinal symptoms for more than 6 months[8]—highlighting the existence of a pre-clinical phase of disease. To improve the overall treatment of IBD patients and minimize avoidable bowel damage, a better understanding of this pre-clinical phase of disease is required. However, studying the pre-clinical phase of IBD is challenging—since patients do not have an established diagnosis at this point—and studies to date have either been limited by small numbers of clinical samples and pre-diagnostic data or have relied on patients having to best recall symptoms from many years earlier. In the American PREDICTS cohort, which collected serum samples from military personnel, 400 individuals were identified who subsequently developed IBD. By analyzing pre-diagnostic samples, several anti-microbial antibodies[9] and serum proteins[10] were found to correlate with subsequent IBD development. Similar results were found in the GEM study, which recruited unaffected relatives of CD patients and followed them prospectively for development of CD (n = 77), and in a Swedish population-study, which included 72 patients who later developed UC.[11,12] However, the small sizes of such pre-diagnosis cohorts limits a more comprehensive study of this phase of disease.

IBD is associated with several biochemical and hematological changes, which are often used to support the diagnostic process, including vitamin and mineral deficiencies,[13,14] anemia,[15] increases in inflammatory markers, including C-reactive protein (CRP) and fecal calprotectin (f-cal),[16] and/or clinically measurable manifestations of associated liver disease.[17]

In this study, we aimed to determine whether hematological or biochemical changes would be detectable within the pre-clinical phase of IBD using Danish laboratory data from ∼20,000 IBD patients (diagnosed between 2008 and 2018) and 4.6 million potential population-based controls. By considering results up to 10 years before diagnosis, we provide a systematic characterization of the pre-clinical phase of IBD at nationwide scale. Finally, we assess whether the pre-diagnostic changes we detect might be useful predictively, both as a proof of concept and to help inform future prediction models of IBD.

Table 1. Sample summary

	CD controls	CD patients	UC controls	UC patients
Tests	N measurements/N individuals	N measurements/N individuals	N measurements/N individuals	N measurements/N individuals
CRP	55145/53466	11030/4585	84736/80926	16950/7543
F-cal	4402/4009	885/693	6148/5567	1237/1056
Leukocytes	75520/72931	15104/5374	118060/112135	23615/9036
Neutrophils	35131/33890	7027/3018	53065/50379	10616/4871
Lymphocytes	49721/47993	9945/4246	73675/69905	14738/6783
Monocytes	47506/45929	9502/4167	69595/66478	13922/6600
Eosinophils	34762/33706	6953/3143	49330/47312	9869/4836
Basophils	46871/45341	9375/4120	68930/65708	13789/6543
Platelets	63935/61821	12787/4764	99333/94387	19869/7959
Hemoglobin	81575/78865	16315/5442	130008/123318	26004/9236
Iron	20509/19934	4103/2300	29293/28322	5861/3484
Folate	11221/10974	2245/1462	15150/14765	3030/2030
Vitamin B12	32075/31453	6415/3325	46743/45459	9350/5024
Vitamin D2+D3	27033/26555	5407/2719	40077/39094	8016/4117
ALAT	81940/79426	16388/5743	137667/130802	27536/9751
Albumin	38235/36935	7647/3339	62979/59534	12597/5537
Bilirubin	36435/35458	7287/3565	59350/56973	11872/5782

The table shows the number of measurements and number of individuals for CD and UC patients and matched controls. Each measurement from a CD or UC patient is matched 1:5 to measurement from a control using matching with replacement. Additional information within each clinical feature is available in Table S5. CD, Crohn's disease; UC, ulcerative colitis; CRP, C-reactive protein; f-cal, fecal calprotectin; ALAT, alanine-aminotransferase.

RESULTS

We identified 12,466 patients with CD and 23,533 patients with UC who were diagnosed between 2008 and 2018. Median age at diagnosis of CD was 37.0 years (interquartile range, IQR, 22.7–57.2) and of UC was 45.1 years (IQR, 29.1–62.8). 54.6% of CD patients and 52.6% of UC patients were women. Within these datasets, 7,739 CD patients and 12,934 UC patients had at least one of the selected tests performed before their first diagnosis of CD or UC. During the same period, 7,288,145 individuals lived in Denmark, and of these, 4,550,623 had at least one available result for a study test in the Register of Laboratory Results for Research (RLRR). Table 1 shows number of test results and number of individuals for each test after matching. Table S3 provides corresponding pre-clinical numbers, defined as measurements more than 1 year prior to diagnosis. Table S4 provides additional phenotype data including sex and age of the patients and controls.

Association between test results and development of inflammatory bowel disease

We first analyzed all test results for both static and time-dependent differences between IBD cases and controls (Figure S3A). Static differences account for a different baseline in test results between cases and controls, while time-dependent differences account for a different slope of test results over time (until diagnosis) between cases and controls. All tests were significantly associated with CD or UC in either the static and/or time-dependent analyses. Levels of 16 of 17 tests showed statistically significant static differences between CD cases and controls. When analyzing results over time, 12 of 17 tests also showed time-dependent differences—implying that there were temporal trends in test results within the pre-diagnostic period that differed between individuals who developed CD and controls. For UC, 16 of 17 tests showed statistically significant static differences, and 15 of 17 tests showed time-dependent differences. All tests were corrected for multiple testing.

We next investigated all 17 tests to further explore associations with IBD at specific time intervals before diagnosis. When examining 1-year time intervals before diagnosis, we observed the largest differences in the year immediately preceding diagnosis (Figure 1). For many tests, this association became gradually weaker at earlier time points. However, 8 years before a diagnosis of CD, levels of leukocytes, neutrophils, and platelets remained significantly higher in CD cases compared to controls. 7 years before CD diagnosis, levels of CRP were also higher, while levels of hemoglobin were lower. 5 years before CD diagnosis, levels of monocytes were higher in cases than controls, while levels of iron, albumin, and bilirubin were lower. In UC, we also observed significant differences in several blood tests in the pre-diagnostic phase of disease, although these differences were detectable over a much shorter period compared to CD. For example, 3 years before UC diagnosis, cases had higher levels of CRP, leukocytes, neutrophils, eosinophils, and platelets compared to controls, but these differences were not apparent at earlier time points (Figure 1). In total, 14/17 and 9/17 tests showed pre-clinical changes (within time interval −2 to −9) that were associated with CD and UC, respectively (Figure 1). We exclude time interval −1 (from 1 day until 1 year before diagnosis) from the pre-clinical window since we have previously

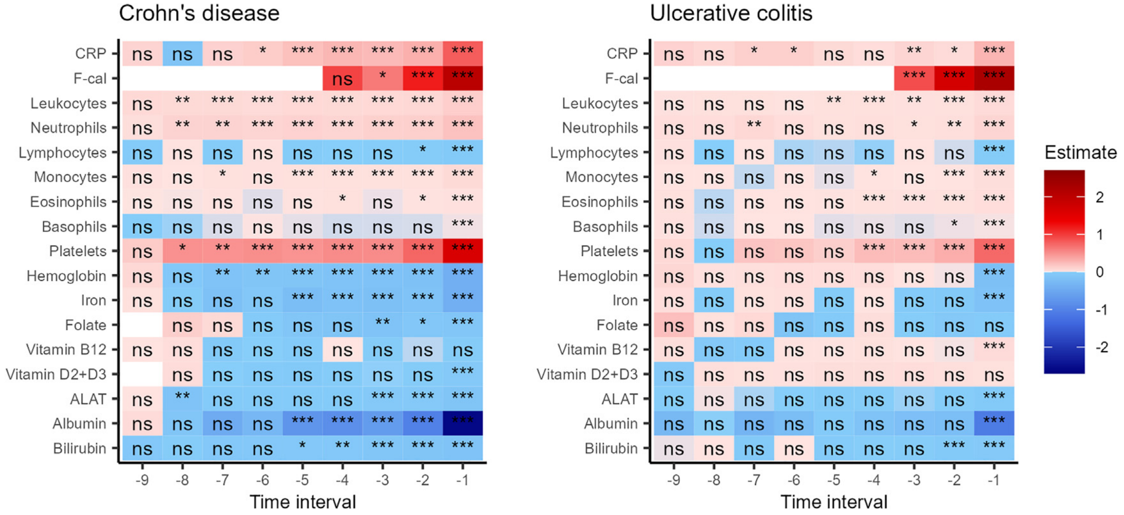


Figure 1. Biochemical and hematological changes in the pre-clinical phase of IBD

The dataset was divided into 1-year pre-diagnosis time intervals so that time interval −1 contains measurements from the year preceding diagnosis, etc. Left panels show results for CD, and the right panels show results for UC. Tiles are colored by the estimated beta-values of the transformed test results. A positive (red) estimate indicates that the test measurements were higher in future CD or UC patients compared to controls. Empty tiles indicate insufficient available measurements within the time interval to fit a model. Number of participants for each measure is shown in Table 1. ns, adjusted p value ≥ 0.05; *, adjusted p value < 0.05; **, adjusted p value < 0.01; ***, adjusted p value < 0.001; CRP, C-reactive protein; f-cal, fecal calprotectin; ALAT, alanine-aminotransferase.

shown that 46.9% of IBD patients had an IBD-relevant procedure before the date of first IBD hospital admission in an overlapping Danish health dataset.[18] This could indicate a delay in diagnosis registration, although reassuringly this number approached zero within the 180 days before admission.

Strikingly, the median values for all blood test results (indeed, all test results apart from fecal calprotectin) were within their respective normal ranges (Figures 2 and S3B). This suggests that while pre-diagnostic results in future IBD patients may have significantly differed from controls, these results would generally not have been considered abnormal. The estimated strength of associations is shown in Table S5.

Sensitivity analysis

We conducted three sensitivity analyses. First, to minimize misclassification of IBD, we used a more stringent definition of IBD based on registration of two diagnoses of IBD within 2 years. When using this definition, results were similar to the main analysis results (Figure S4A). Second, we used an alternative model, including only one sample per IBD case per time interval. This decreased the statistical power marginally, but it did not change the results substantially—suggesting that the adjustment for multiple samples per individual was appropriate (Figure S4B). Third, to take comorbidities into account, we additionally adjusted for the Charlson comorbidity index. Notably, comorbidities were not abundant in either the case or control populations (Table S2), and the main results were robust to the correction for Charlson comorbidity index (Figure S4C).

Prediction of future diagnosis of Crohn's disease and ulcerative colitis

Based on our findings, we further investigated whether pre-diagnostic levels of CRP, neutrophils, monocytes, platelets, hemoglobin, and eosinophils together with the covariates age, sex, and laboratory ID could predict the later development of CD or UC using a logistic regression model (Table 2). The mean predicted probability for developing IBD increased as we approached date of IBD diagnosis in cases. No change was noted in the controls over the equivalent time periods (Figure 3A). The model was better at predicting future CD cases than future UC cases (compared to controls). In the validation dataset, the area under the receiver operating characteristic (ROC) curve (AUC) for the logistic regression model was 72.7% for separating CD and controls and 64.6% for separating UC and controls in the year preceding diagnosis.

Finally, we examined whether the separation of IBD cases and controls was improved by fitting non-linear support vector machine (SVM) and random forest (RF) models. The model fitted using RF had the marginally better performance compared to SVM (Figure S5) and is shown in Figure 3B. We generated ROC curves for the prediction of CD/UC cases from controls for all time intervals (Figures 3C and 3D). AUC values were again generally better for predicting CD than UC and improved closer to the date of diagnosis. The best AUC using RF models was 73.6% for predicting CD in the year before diagnosis.

DISCUSSION

The existence of a pre-clinical phase of IBD has long been suspected. This phenomenon, which cannot simply be ascribed to delayed diagnosis, is supported by several small studies that have identified biological changes in healthy individuals who would later develop IBD.[10,11,19,20]

Unfortunately, relatively little is known about this pre-clinical phase—since patients do not have an established diagnosis and often no clinical symptoms at this point—and most studies

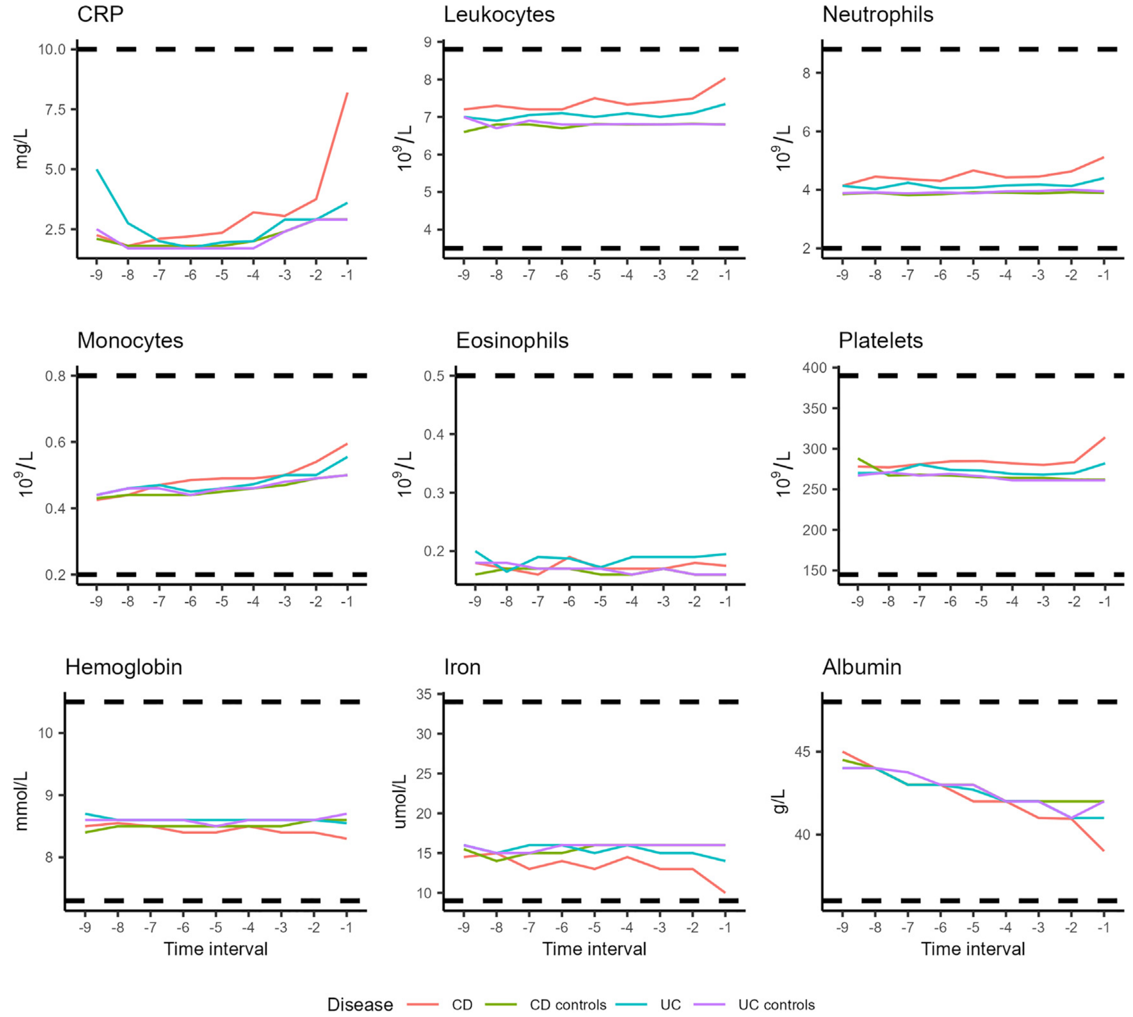


Figure 2. Biochemical and hematological values in relation to the medical references. For each test parameter in CD, UC, and respective controls, the median value was calculated and plotted for each time interval before diagnosis.

If the same individual had multiple results within a time interval, the median value was used. Medical reference limits for each test are shown as horizontal dashed lines. Plots of the remaining features can be found in Figure S3B. Number of participants for each measure is shown in Table 1.

to date have either had limited clinical samples and pre-diagnostic data or had to rely on patients recalling events from many years earlier.[8] One point of general agreement is that, where present, this phase probably precedes a diagnosis of IBD by between 1 and 5 years, but exactly what happens in this period remains unknown.

Here, we present the largest ever objective study into the pre-clinical phase of IBD. Using pre-diagnosis blood results from a nationwide cohort of ∼20,000 IBD patients and 4.6 million potential controls, we identify widespread hematological and biochemical changes that precede a diagnosis of IBD by up to 8 years. Specifically, we detect differences in parameters including CRP, leukocytes, neutrophils, monocytes, platelets, hemoglobin, iron, and albumin, which occur up to 8 years before a diagnosis of CD and up to 3 years before a diagnosis of UC. These results indicate that disease initiation is likely to begin far earlier than previously thought, especially in CD. This has important implications for the future development of strategies aimed at preventing the onset of disease and conversely highlights a considerable window of opportunity that could be targeted pharmacologically for early therapy or by addressing modifiable risk factors (e.g., smoking and diet).

Notably, most of the changes detected (except for fecal calprotectin) did not exceed the normal ranges of the affected tests, which is presumably why they have not been noted previously. Indeed, this discovery was only made possible because of the availability of nationwide electronic health data in a country with a free healthcare system and a low percentage of undiagnosed IBD cases.[21] While the absolute test results were generally not abnormal, it was striking that the pre-clinical changes phenocopied alterations that are typically associated with CD and UC. For example, in CD, progressive reductions in iron, hemoglobin, and albumin and concomitant increases in CRP, monocytes, and platelets were observed years before a diagnosis was made.

Table 2. Dataset description for six selected tests

Time interval	N CD	N CD controls	N UC	N UC controls
−1	1186	1186	1818	1818
−2	470	470	703	703
−3	352	352	541	541
−4	251	251	373	373
−5	186	186	274	274
−6	126	126	196	196
−7	92	92	128	128
−8	35	35	82	82
−9	18	18	33	33

The six blood tests CRP, neutrophils, monocytes, platelets, hemoglobin, and eosinophils were evaluated further in prediction models. Cases who had all six measurements taken within the same time interval were selected and matched with controls. The resulting sample sizes in each time interval are shown in the table. Further, time interval −1 was split 80:20 in a training and validation dataset. All data in the earlier time intervals were used to evaluate the model trained on the −1 interval.

Similar results were observed in UC but for a much shorter time period prior to diagnosis. This discrepancy may partly reflect differences in the symptoms of the two diseases, with the eventual onset of UC symptoms (bloody diarrhea) being more likely to trigger prompt investigations—and less likely to be misdiagnosed as other conditions—than those of CD (abdominal pain, non-bloody diarrhea). However, it is important to note that in large epidemiological studies, only a minority of IBD patients report symptoms more than 6 months before diagnosis,[8] and so differences in the time from symptom onset until diagnosis are unlikely to fully account for this observation.

Since most of the pre-clinical changes were not due to abnormally high or low test results, we concluded that these would not be individually useful as diagnostic tools. However, we also reasoned that it might be possible to build predictive classifiers by considering these in combination. Using results of CRP, neutrophils, monocytes, platelets, hemoglobin, and eosinophils, we showed that predictive classifiers could be built using various approaches (including logistic regression, RF, and SVM) but that their overall predictive performance in an independent validation set was modest and worsened the further they were from diagnosis. Notably, this analysis did not consider the possibility that associations between predictors and future IBD status may vary between time intervals, since the models were fitted on time interval −1 and applied to the other time intervals. This may have contributed to the poor performance of the model at the earliest time points. Unsurprisingly, performance in CD was better than in UC (since significant changes are detectable earlier in pre-clinical CD), but the best AUC obtained 1 year before diagnosis for CD was only 73.6%. It is possible that the inclusion of other blood tests could improve this performance, and serum protein biomarkers have been reported to be able to predict a diagnosis of CD within 1 year with an AUC of 87%, although this result was not externally validated.[10] In any case, it seems likely that better biomarkers will be needed to fulfill the potential of targeting the pre-clinical phase of IBD. Incorporating additional predictors and investigating more advanced learning algorithms, including neural networks, could improve model performance and should be the focus of future work.

Our study design used matching and statistical testing at the level of individual tests—rather than at an individual patient level—to facilitate the inclusion of historical data from 2008 to 2015 (before the RLRR was initiated and thus when the database may have been incomplete). This enabled analysis of earlier time periods than would otherwise have been possible but means we cannot examine individual patient trajectories. This should be possible in future studies when the RLRR has been complete for a longer time period.

Another important limitation of this study is that the indications for the tests were unknown. This creates a potential bias, although this effect should be limited as it would affect both the case and control datasets. We attempted to reduce the influence of this bias by restricting measurements to those sampled by general practitioners (thus excluding hospitalized patients) and by adjusting for the Charlson comorbidity index (although the value of this index was generally low, reflecting the relatively young age of the patients in the datasets). However, it was not possible to include more granular data relating to medication use or other morbidities than those included in the Charlson comorbidity index. Another limitation in this study is that despite its size, the study power varied at different time points. This reflected the availability of samples, with fewer being available for time intervals further away from the date of diagnosis and with differential availability of different test results at different time points. This will have negatively affected the power to detect differences at the earliest time points, meaning that the results described here may underestimate the true duration of the pre-clinical phase of CD in particular. By only requiring the diagnosis of IBD to have been registered once, there is also a potential for some patients to have been misclassified. However, we have previously shown that this approach is associated with a low misclassification rate in these databases,[22] and a sensitivity analysis using two registered diagnoses replicated findings from the main analysis. Finally, by including all available tests from IBD patients, there is a possibility that this study could be influenced by the regression toward the mean phenomenon. However, a sensitivity analysis including only one sample per IBD patient did not change the results substantially.

The study is potentially influenced by Berkson-like bias, since the case and control samples were both from the subpopulation of citizens who attended the Danish healthcare system. We sought to limit the effect of this bias by excluding hospitalized patients and by using standard tests that are commonly taken in the general population for a variety of indications.

It is not known whether the changes we detect in the pre-clinical phase of IBD are specific to IBD or would also be seen in other inflammatory disorders, e.g., rheumatoid arthritis (RA). Sokolove et al.[23] detected changes in levels of autoantibodies and cytokines in the pre-clinical phase of RA, some of which (e.g., IL-10) have also been reported in the pre-clinical phase of IBD.[10] A pre-clinical phase of disease may therefore be a common feature among inflammatory disorders, which warrants further study.

In conclusion, we have systematically delineated the extent of the hematological and biochemical changes that occur within

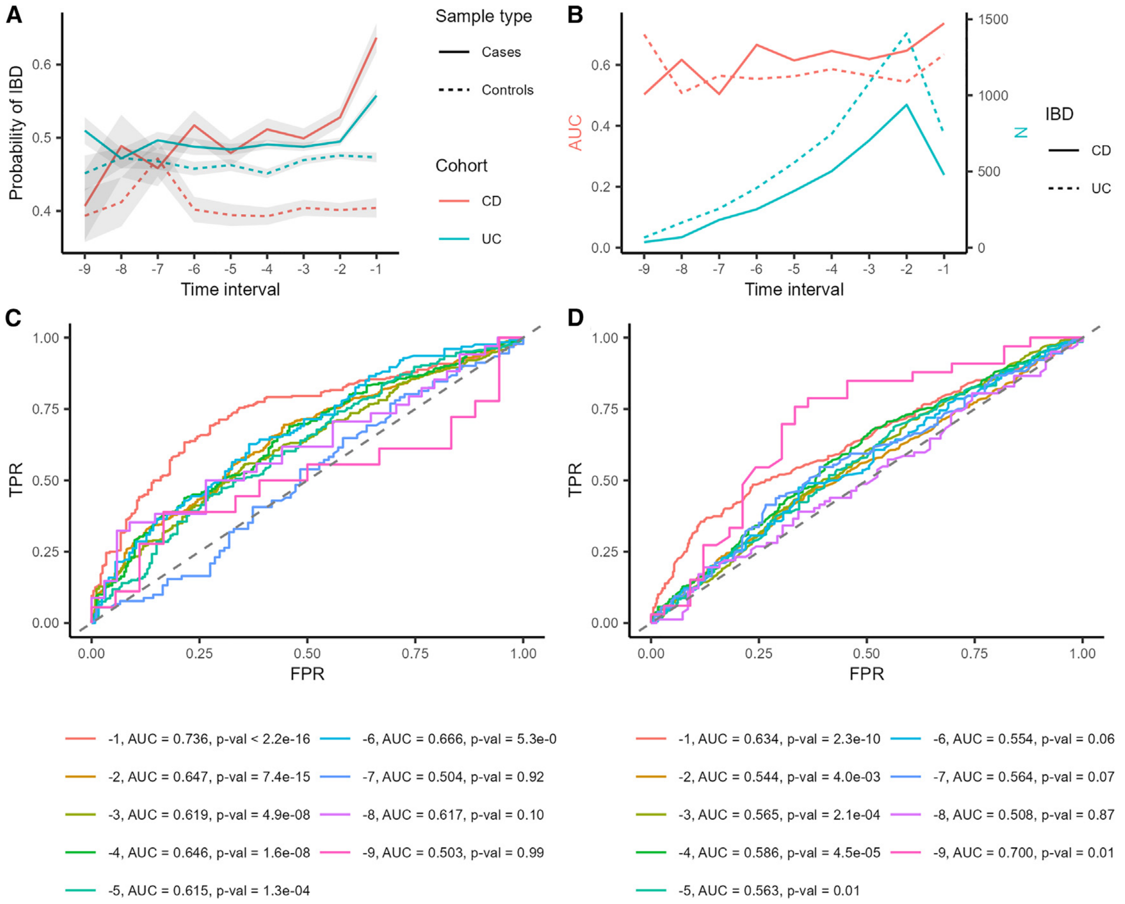


Figure 3. Prediction models of IBD based on clinical features

(A) A logistic regression model to predict future CD/UC cases from controls was fitted on data from time interval −1. The resulting model was used to calculate the predicted probability of developing IBD on all data. The plot shows the mean predicted probability of developing IBD in each time interval. Shaded areas represent standard error.

(B) Random forest (RF) models were fitted to separate CD/UC cases and controls using training data from time interval −1 (80% of data). The final model was applied on the validation data from time interval −1 (20% of data) and the entire datasets from other time intervals. The figure shows the resulting AUC values for the model's ability to predict future CD/UC cases at each time interval, together with the corresponding sample size of the dataset that the model was applied to. In time interval −1, the RF model is applied to the validation dataset only.

(C) ROC curves and their AUCs for each time interval for predicting CD from controls.

(D) ROC curves and their AUCs for each time interval for predicting UC from controls. AUCs were evaluated for whether they differed significantly from 0.5 using a Mann-Whitney U test. Number of participants for each measure is shown in Table 2. IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; FPR, false positive rate; TPR, true positive rate; AUC, area under the receiver operating characteristic curve.

the pre-clinical phase of CD and UC. These changes far exceed previous expectations regarding the length of this pre-clinical phase of disease and thereby provide important insights that will need to be considered if future treatment strategies aspire to disease prevention.

Limitations of the study

Our study is based on register data and is thus influenced by the selection of the subpopulation with available data. Data availability, and thus study power, reduced as time to diagnosis increased. Future prospective studies with predefined sampling could overcome this limitation, but assembling cohorts of equivalent size is likely to be prohibitive.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - ○ Blood and stool tests

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.101263.

ACKNOWLEDGMENTS

Funding sources: Danish National Research Foundation Center of Excellence Grant (DNRF148) to T.J. J.C.L. is a Lister Institute Prize Fellow and is supported by the Francis Crick Institute, which receives its core funding from the UK Medical Research Council (CC2219), Cancer Research UK (CC2219), and the Wellcome Trust (CC2219).

AUTHOR CONTRIBUTIONS

Conception and design: all authors. Acquisition and analysis: M.V.V., G.J.P., and T.J. Interpretation: M.V.V., K.H.A., J.C.L., and T.J. Drafting the work: M.V.V., K.H.A., J.C.L., and T.J. Revising the work for critically important intellectual content: all authors. Final approval: all authors. Agreement to be accountable for the work: all authors.

DECLARATION OF INTERESTS

J.C.L. reports financial support for research from GSK and consultancy fees from Abbvie, AgPlus Diagnostics, PredictImmune, and C4X Discovery.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 8, 2023
Revised: July 21, 2023
Accepted: October 10, 2023
Published: November 7, 2023

REFERENCES

1. Agrawal, M., Christensen, H.S., Bøgsted, M., Colombel, J.F., Jess, T., and Allin, K.H. (2022). The rising burden of IBD in Denmark over two decades: a nationwide cohort study. Gastroenterology *163*, 1547–1554. https://doi.org/10.1053/j.gastro.2022.07.062.
2. GBD 2017 Inflammatory Bowel Disease Collaborators; Sepanlou, S.G., Ikuta, K., Vahedi, H., Bisignano, C., Safiri, S., Sadeghi, A., Nixon, M.R., Abdoli, A., Abolhassani, H., et al. (2020). 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. *5*, 17–30. https://doi.org/10.1016/S2468-1253(19)30333-4.
3. Cai, Z., Wang, S., and Li, J. (2021). Treatment of Inflammatory Bowel Disease: A Comprehensive Review. Front. Med. *8*, 2681. https://doi.org/10.3389/FMED.2021.765474/BIBTEX.
4. Park, S.-K., Ye, B.D., Yang, S.-K., Kim, S.-O., Kim, J., Kim, J.W., Park, S.H., Yang, D.-H., Jung, K.W., Kim, K.-J., et al. (2014). Clinical features and course of ulcerative colitis diagnosed in asymptomatic subjects. J. Crohn's Colitis *8*, 1254–1260. https://doi.org/10.1016/J.CROHNS.2014.03.002.
5. Schoepfer, A.M., Dehlavi, M.A., Fournier, N., Safroneeva, E., Straumann, A., Pittet, V., Peyrin-Biroulet, L., Michetti, P., Rogler, G., and Vavricka, S.R.; IBD Cohort Study Group (2013). Diagnostic delay in Crohn's disease is associated with a complicated disease course and increased operation rate. Am. J. Gastroenterol. *108*, 1744–1753, quiz 1754. https://doi.org/10.1038/AJG.2013.248.
6. Li, Y., Ren, J., Wang, G., Gu, G., Wu, X., Ren, H., Hong, Z., Hu, D., Wu, Q., Li, G., et al. (2015). Diagnostic delay in Crohn's disease is associated with increased rate of abdominal surgery: A retrospective study in Chinese patients. Dig. Liver Dis. *47*, 544–548. https://doi.org/10.1016/J.DLD.2015.03.004.
7. Pellino, G., Sciaudone, G., Selvaggi, F., and Riegler, G. (2015). Delayed diagnosis is influenced by the clinical pattern of Crohn's disease and affects treatment outcomes and quality of life in the long term: a cross-sectional study of 361 patients in Southern Italy. Eur. J. Gastroenterol. Hepatol. *27*, 175–181. https://doi.org/10.1097/MEG.0000000000000244.
8. Blackwell, J., Saxena, S., Jayasooriya, N., Bottle, A., Petersen, I., Hotopf, M., Alexakis, C., and Pollok, R.C.; POP-IBD study group (2020). Prevalence and Duration of Gastrointestinal Symptoms Before Diagnosis of Inflammatory Bowel Disease and Predictors of Timely Specialist Review: A Population-Based Study. J. Crohn's Colitis *15*, jjaa146-211. https://doi.org/10.1093/ECCO-JCC/JJAA146.
9. Choung, R.S., Princen, F., Stockfisch, T.P., Torres, J., Maue, A.C., Porter, C.K., Leon, F., De Vroey, B., Singh, S., Riddle, M.S., et al. (2016). Serologic microbial associated markers can predict Crohn's disease behaviour years before disease diagnosis. Aliment. Pharmacol. Ther. *43*, 1300–1310. https://doi.org/10.1111/APT.13641.
10. Torres, J., Petralia, F., Sato, T., Wang, P., Telesco, S.E., Choung, R.S., Strauss, R., Li, X.J., Laird, R.M., Gutierrez, R.L., et al. (2020). Serum Biomarkers Identify Patients Who Will Develop Inflammatory Bowel Diseases Up to 5 Years Before Diagnosis. Gastroenterology *159*, 96–104. https://doi.org/10.1053/J.GASTRO.2020.03.007.
11. Bergemalm, D., Andersson, E., Hultdin, J., Eriksson, C., Rush, S.T., Kalla, R., Adams, A.T., Keita, Å.V., D'Amato, M., Gomollon, F., et al. (2021). Systemic Inflammation in Preclinical Ulcerative Colitis. Gastroenterology *161*, 1526–1539.e9. https://doi.org/10.1053/j.gastro.2021.07.026.
12. Lee, S.H., Turpin, W., Espin-Garcia, O., Raygoza Garay, J.A., Smith, M.I., Leibovitzh, H., Goethel, A., Turner, D., Mack, D., Deslandres, C., et al. (2021). Anti-Microbial Antibody Response is Associated With Future Onset of Crohn's Disease Independent of Biomarkers of Altered Gut Barrier Function, Subclinical Inflammation, and Genetic Risk. Gastroenterology *161*, 1540–1551. https://doi.org/10.1053/j.gastro.2021.07.009.
13. Del Pinto, R., Pietropaoli, D., Chandar, A.K., Ferri, C., and Cominelli, F. (2015). Association Between Inflammatory Bowel Disease and Vitamin D Deficiency: A Systematic Review and Meta-analysis. Inflamm. Bowel Dis. *21*, 2708–2717. https://doi.org/10.1097/MIB.0000000000000546.
14. Vagianos, K., Bector, S., McConnell, J., and Bernstein, C.N. (2007). Nutrition Assessment of Patients With Inflammatory Bowel Disease. J. Parenter. Enter. Nutr. *31*, 311–319. https://doi.org/10.1177/0148607107031004311.
15. Stein, J., Hartmann, F., and Dignass, A.U. (2010). Diagnosis and management of iron deficiency anemia in patients with IBD. Nat. Rev. Gastroenterol. Hepatol. *7*, 599–610. https://doi.org/10.1038/NRGASTRO.2010.151.
16. Chang, S., Malter, L., and Hudesman, D. (2015). Disease monitoring in inflammatory bowel disease. World J. Gastroenterol. *21*, 11246–11259. https://doi.org/10.3748/WJG.V21.I40.11246.
17. Karaivazoglou, K., Konstantakis, C., Tourkochristou, E., Assimakopoulos, S.F., and Triantos, C. (2020). Non-alcoholic fatty liver disease in inflammatory bowel disease patients. Eur. J. Gastroenterol. Hepatol. *32*, 903–906. https://doi.org/10.1097/MEG.0000000000001679.
18. Rasmussen, N.F., Green, A., Allin, K.H., Iversen, A.T., Madsen, G.I., Pedersen, A.K., Wolff, D.L., Jess, T., and Andersen, V. (2022). Clinical procedures used to diagnose inflammatory bowel disease: real-world evidence from a Danish nationwide population-based study. BMJ Open Gastroenterol. *9*, e000958. https://doi.org/10.1136/BMJGAST-2022-000958.

19. Turpin, W., Lee, S.H., Raygoza Garay, J.A., Madsen, K.L., Meddings, J.B., Bedrani, L., Power, N., Espin-Garcia, O., Xu, W., Smith, M.I., et al. (2020). Increased Intestinal Permeability Is Associated With Later Development of Crohn's Disease. Gastroenterology *159*, 2092–2100.e5. https://doi.org/10.1053/J.GASTRO.2020.08.005.

20. Israeli, E., Grotto, I., Gilburd, B., Balicer, R.D., Goldin, E., Wiik, A., and Shoenfeld, Y. (2005). Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut *54*, 1232–1236. https://doi.org/10.1136/GUT.2004.060228.

21. Jess, T., Vestergaard, M.V., Iversen, A.T., and Allin, K.H. (2022). Undiagnosed Inflammatory Bowel Disease Among Individuals Undergoing Colorectal Cancer Screening: A Nationwide Danish Cohort Study 2014–2018. Gut *72*, 214–216. https://doi.org/10.1136/GUTJNL-2022-327296.

22. Albaek Jacobsen, H., Jess, T., and Larsen, L. (2022). Validity of Inflammatory Bowel Disease Diagnoses in the Danish National Patient Registry: A Population-Based Study from the North Denmark Region. Clin. Epidemiol. *14*, 1099–1109. https://doi.org/10.2147/CLEP.S378003.

23. Sokolove, J., Bromberg, R., Deane, K.D., Lahey, L.J., Derber, L.A., Chandra, P.E., Edison, J.D., Gilliland, W.R., Tibshirani, R.J., Norris, J.M., et al. (2012). Autoantibody Epitope Spreading in the Pre-Clinical Phase Predicts Progression to Rheumatoid Arthritis. PLoS One *7*, e35296. https://doi.org/10.1371/JOURNAL.PONE.0035296.

24. Laboratoriedatabasen - Sundhedsdatastyrelsen https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/doedsaarsager-og-biologisk-materiale/laboratoriedatabasen.

25. Arendt, J.F.H., Hansen, A.T., Ladefoged, S.A., Sørensen, H.T., Pedersen, L., and Adelborg, K. (2020). Existing Data Sources in Clinical Epidemiology: Laboratory Information System Databases in Denmark. Clin. Epidemiol. *12*, 469–475. https://doi.org/10.2147/CLEP.S245060.

26. Schmidt, M., Schmidt, S.A.J., Sandegaard, J.L., Ehrenstein, V., Pedersen, L., and Sørensen, H.T. (2015). The Danish National Patient Registry: a review of content, data quality, and research potential. Clin. Epidemiol. *7*, 449–490. https://doi.org/10.2147/CLEP.S91125.

27. Chen, P., Zhou, G., Lin, J., Li, L., Zeng, Z., Chen, M., and Zhang, S. (2020). Serum Biomarkers for Inflammatory Bowel Disease. Front. Med. *7*, 123. https://doi.org/10.3389/FMED.2020.00123.

28. Liu, D., Saikam, V., Skrada, K.A., Merlin, D., and Iyer, S.S. (2022). Inflammatory bowel disease biomarkers. Med. Res. Rev. *42*, 1856–1887. https://doi.org/10.1002/MED.21893.

29. Okba, A.M., Amin, M.M., Abdelmoaty, A.S., Ebada, H.E., Kamel, A.H., Allam, A.S., and Sobhy, O.M. (2019). Neutrophil/lymphocyte ratio and lymphocyte/monocyte ratio in ulcerative colitis as non.invasive biomarkers of disease activity and severity. Autoimmun. Highlights *10*, 4–9. https://doi.org/10.1186/S13317-019-0114-8/TABLES/6.

30. Harries, A.D., Beeching, N.J., Rogerson, S.J., and Nye, F.J. (1991). The platelet count as a simple measure to distinguish inflammatory bowel disease from infective diarrhoea. J. Infect. *22*, 247–250. https://doi.org/10.1016/S0163-4453(05)80006-4.

31. Kaitha, S., Bashir, M., and Ali, T. (2015). Iron deficiency anemia in inflammatory bowel disease. World J. Gastrointest. Pathophysiol. *6*, 62–72. https://doi.org/10.4291/WJGP.V6.I3.62.

32. Franklin, J.L., and Rosenberg, H.H. (1973). Impaired Folic Acid Absorption in Inflammatory Bowel Disease: Effects of Salicylazosulfapyridine (Azulfidine). Gastroenterology *64*, 517–525. https://doi.org/10.1016/S0016-5085(73)80120-9.

33. Cappello, M., Randazzo, C., Bravatà, I., Licata, A., Peralta, S., Craxì, A., and Almasio, P.L. (2014). Liver Function Test Abnormalities in Patients with Inflammatory Bowel Diseases: A Hospital-based Survey. Clin. Med. Insights Gastroenterol. *7*, 25–31. https://doi.org/10.4137/CGAST.S13125.

34. Khan, N., Patel, D., Shah, Y., Trivedi, C., and Yang, Y.X. (2017). Albumin as a prognostic marker for ulcerative colitis. World J. Gastroenterol. *23*, 8008–8016. https://doi.org/10.3748/WJG.V23.I45.8008.

35. Roggenbuck, D., Reinhold, D., Schierack, P., Bogdanos, D.P., Conrad, K., and Laass, M.W. (2014). Crohn's disease specific pancreatic antibodies: clinical and pathophysiological challenges. Clin. Chem. Lab. Med. *52*, 483–494. https://doi.org/10.1515/CCLM-2013-0801.

36. Ibbs, S., and Muhammed, R. (2017). PTH-055 Vitamin b12 deficiency is common in children with ulcerative colitis as well as Crohn's disease. Gut *66*, A233. https://doi.org/10.1136/GUTJNL-2017-314472.454.

37. Agrawal, M., Christensen, H.S., Bøgsted, M., Colombel, J.-F., Jess, T., and Allin, K.H. (2022). The Rising Burden of Inflammatory Bowel Disease in Denmark Over Two Decades: A Nationwide Cohort Study. Gastroenterology *163*, 1547–1554.e5. https://doi.org/10.1053/j.gastro.2022.07.062.

38. Christensen, S., Johansen, M.B., Christiansen, C.F., Jensen, R., and Lemeshow, S. (2011). Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care. Clin. Epidemiol. *3*, 203–211. https://doi.org/10.2147/CLEP.S20247.

39. Thygesen, S.K., Christiansen, C.F., Christensen, S., Lash, T.L., and Sørensen, H.T. (2011). The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med. Res. Methodol. *11*, 83. https://doi.org/10.1186/1471-2288-11-83.

40. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B *57*, 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.

41. Benjamini, Y., and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. Ann. Stat. *29*, 1165–1188. https://doi.org/10.1214/AOS/1013699998.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Danish nationwide Register of Laboratory Results for Research	Danish National Health registers	https://sundhedsdatastyrelsen.dk
Danish National Patient Registry	Danish National Health registers	https://sundhedsdatastyrelsen.dk
Software and algorithms		
Original code	GitHub, Zenodo	https://zenodo.org/record/8318775
R (v 4.1.3)	The R Project	http://www.r-project.org
BioRender	BioRender	https://www.biorender.com/

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to the lead contact, Tine Jess (jess@dcm.aau.dk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This study is based on data from the Danish National Health registers (https://sundhedsdatastyrelsen.dk). The register data are protected by the Danish Act on Processing of Personal Data and can be accessed through application to and approval from the Danish Data Protection Agency and the Danish Health Data Authority. Application requires a project description, a list of requested registers and variables, and documentation from a Danish institution who will be responsible for handling the data. The lead author can be contacted for further information on accessing the data.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOI is listed in the key resources table.
- Any additional information is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study is based on results from the Danish nationwide Register of Laboratory Results for Research (RLRR).[24,25] In brief, RLRR covers all regions of Denmark and includes the results of biochemistry and hematology tests taken at both hospitals and general practitioners. RLRR was initiated in 2015 but includes historical data from many hospitals in Denmark going back to 2008. We linked the RLRR to the Danish National Patient Registry[26] (NPR). IBD cases were defined as all individuals with a hospital contact with a diagnosis of either CD (ICD-10: K50) or UC (ICD-10: K51) who had lived in Denmark for at least one year before diagnosis.

METHOD DETAILS

Blood and stool tests

In total, 17 different clinical tests were selected for analysis based on prior knowledge and including parameters that are known to differ between IBD patients and healthy controls.[27–36] If a test was covered by more than one Nomenclature for Properties and Units (NPU) code, the NPU code with the largest amount of data was used. In addition to the selected tests, we had planned to include perinuclear anti-neutrophil cytoplasmic antibodies, anti-Saccharomyces cerevisiae antibodies, anti-pancreatic antibody, erythrocyte sedimentation rate, and aspartate transaminase, however there was insufficient data for these measurements to be included.

For the IBD cases, we included available test results from the RLRR that had been registered up until the day before the first hospital contact (date of admission) for IBD. Only results from tests taken at general practitioners or non-hospital clinics were included to minimize biases that may be associated with the indications for tests performed in hospitals. This was achieved by only including results with a requester ID type ‘‘external number’’ (‘‘ydernummer’’).

Matching of tests from IBD patients and controls

Each IBD test result was matched with up to five equivalent test results from controls based on sex, age (±5 years), laboratory and time of sampling (±0.5 years). Samples from individuals who died, emigrated, or subsequently received an IBD diagnosis within half a year of the matched IBD case were ineligible for matching. Matching was performed with replacement and if no controls were available, the IBD test was excluded from further analyses.

We calculated the number of days between test sampling and diagnosis for all IBD patients. For the matched controls, we similarly calculated the number of days between test sampling and diagnosis of the matched IBD patient (Figure S1).

Evaluation of data

We replaced categorical values of tests (for example "<10") by imputed values using the median value of all datapoints that matched the categorical description (i.e., the median of all values < 10).

Distributions of the different test results were examined graphically. In cases of non-normal distributions, data were transformed using the natural logarithm or square root to approach normality (Figure S2; Table S1). No outliers were identified using the criteria of more than three standard deviations (SD) from the mean of the transformed values.

Sensitivity analyses

In a sensitivity analysis we restricted the IBD cases to patients having at least two IBD hospital contacts within a two-year period.[22,37] The matching and assigned IBD diagnosis (CD or UC) were performed based on the second recorded diagnosis date, whereas the date of diagnosis in the analyses was based on the first diagnosis obtained.

Some IBD cases had several measurements of the same test within a one-year time interval. This was adjusted for in the model by including a random intercept for the person ID. In addition, we performed a sensitivity analysis including only one sample per IBD case per time interval to confirm that this adjustment was sufficient.

To ensure that the analyses were not biased by different representation of comorbidities in cases and controls, we included the Charlson comorbidity index[38,39] (based on the nine years preceding diagnosis date) (Table S2) in sensitivity analyses to adjust for any differences in comorbidity profiles.

Predictive performance of selected blood tests

In a secondary analysis, we aimed to predict IBD status based on six selected blood tests that were associated with CD and/or UC several years before the date of diagnosis and had a sufficiently large number of available samples at each time point. For each one-year time interval before diagnosis, cases and controls with all six blood tests within the interval were matched 1:1 on age, sex and laboratory. Matching was performed in three monthly intervals, where all CD/UC patients diagnosed within that period were identified and all controls that died, emigrated, or received an IBD diagnosis before the end of the period were excluded. One control was randomly selected for each case to obtain a balanced dataset for prediction models. If the same individual had more than one measurement of the same blood tests within the same one-year time interval, the median value was used. An overview of the resulting dataset is available in Table 2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of single clinical test results

All analyses were conducted in R (version 4.1.3). The test results were modeled as dependent on IBD diagnosis and time until diagnosis using a linear mixed effects model (LMM). The model included an interaction term between IBD diagnosis and time to diagnosis to capture both static (different intercept for cases versus control) and time-dependent changes (different slope for cases versus controls) in the test values between future IBD patients and controls. The model included sex and age as fixed effects and person ID and laboratory as random intercepts to adjust for multiple samples per individual and for possible differences in analysis methods across laboratories. A threshold for statistical significance of 0.05 was used after correcting p values for multiple testing using the Benjamini-Hochberg method.[40]

We further evaluated significant test results using a linear mixed model (LMM) fitted on subsets of the dataset – divided into one-year time intervals before diagnosis – to evaluate the effect size and significance between IBD cases and controls at different times pre-diagnosis. The model similarly included age and sex as fixed effects and patient ID and laboratory as random intercepts. The analysis of different time intervals for the same test were considered dependent and thus the Benjamini-Yekutieli method[41] was used for multiple testing correction.

Predictive performance of selected blood tests

Data in the one-year interval before diagnosis (one day until one year before diagnosis) was randomly split into training (80%) and validation (20%) datasets and a logistic regression model - to predict CD/UC cases from controls - was trained on the training dataset. The predictive performance was subsequently tested for all CD/UC cases and controls in the validation dataset at each time intervals. The regression model was based on the transformed blood test values and adjusted for sex, age, and laboratory.

We further investigated two different machine learning approaches: support vector machine (SVM) and Random Forests (RF). The SVM and RF models were trained on the training dataset including sex, age, and laboratory as covariates. For SVM, the optimal choices of cost and sigma were selected using a grid search, which evaluated by rooted mean squared error using three times 10-fold cross validation. The same approach was performed for selecting the optimal parameters for the number of trees and the number of splitting predictors for RF. The final models were fit using the parameter settings detected through the grid searches (incorporating three times 10-fold cross validation). The optimal models were then applied to the validation data and other time intervals to estimate performance.

The performance of the models in the validation data were visualised using the ROC curve and summarised by calculating the AUC. The ROC curve visualises the true positive rate as a function of the false positive rate. We evaluated whether the AUCs differed significantly from 0.5 using a Mann-Whitney U test.