Genetic risk of inflammatory bowel disease is associated with disease course severity

Authors: Marie Vibeke Vestergaard[1]*, Kristine Højgaard Allin[1,2]*, Anne Krogh Nøhr[1]*, Jesus Vicente Torresano Lominchar[1], Heidi Søgaard Christensen[1], Henrik Krarup[3], Kaspar René Nielsen[4], Henrik Albæk Jacobsen[1,2], Jonas Bybjerg-Grauholm[5], Marie Bækvad-Hansen[5], Rasmus Froberg Brøndum[1], Martin Bøgsted[1], Lone Larsen[1,2#], Aleksejs Sazonovs[1#], Tine Jess[1,2#]

[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]Department of Molecular Diagnostics, Aalborg University Hospital, and Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[4]Department of Clinical Immunology, Zealand University Hospital, Næstved, Denmark

[5]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institute, Copenhagen, Denmark

*, [#] Contributed equally

Corresponding author:

Marie Vibeke Vestergaard

Center for Molecular Prediction of Inflammatory Bowel Disease (PREDICT), Department of Clinical Medicine, Aalborg University

A. C. Meyers Vænge 15, 2450 København, Denmark

Email: marievv@dcm.aau.dk, Phone: +45 42963689

Word count: 3358

Number of main tables and figures: 5

Abstract

Background

Genetic susceptibility to inflammatory bowel disease (IBD) has been widely studied, whereas the genetic contribution to disease progression over time remains largely unknown. In a unique population-based cohort, we explored if genetic susceptibility to IBD associated with disease course severity.

Methods

In a Danish nationwide cohort of 3688 patients with Crohn's disease (CD), 4491 patients with ulcerative colitis (UC), and 9469 controls, we estimated polygenic scores (PGS) for IBD susceptibility. We then investigated the association between susceptibility PGS and severe versus less severe disease courses. Patients were categorized as having a severe disease course if needing I) at least two IBD-related hospitalizations exceeding two days, II) at least two IBD-related major surgeries, III) one hospitalization and one major surgery not overlapping in time, or IV) total use of minimum 5000 mg systemic corticosteroids within the first three years of diagnosis. Secondary analyses explored the association with other severity measures including inflammatory markers, exposure to biologics and immunomodulators, and time to hospitalization and major IBD-related surgery. Statistical analyses included logistic, linear, and Cox proportional hazards regression, and predictive modeling using random forest.

Findings

Patients with severe disease courses had higher susceptibility PGS than patients with less severe disease courses (CD: OR=1·27, P=7·73×10^{-10}, UC: OR=1·35, P=7·29×10^{-17} per SD increase in susceptibility PGS). When comparing the highest versus lowest PGS quintile, we observed a hazard ratio (HR) for major surgery of 2·74 (P=7·19×10^{-18}) in patients with CD and 2·04 (P=4·36×10^{-7}) in patients with UC. A higher susceptibility PGS was also associated with longer and more frequent hospitalizations, higher levels of C-reactive protein, decreased hemoglobin, and a higher need for corticosteroids, immunomodulators, and biologic therapies.

Interpretation

Patients with a higher genetic burden for developing IBD also experience a more severe disease course, suggesting a shared genetic architecture between disease susceptibility and severity.

Funding

Danish National Research Foundation, DNRF148, Lundbeck Foundation, R313-2019-857, Novo Nordisk Foundation, NNF23OC0087616, Danish Colitis Crohn Association.

Introduction

Inflammatory bowel disease (IBD) is a chronic, progressive, immune-mediated intestinal disease with an increasing incidence and prevalence globally. The two most common subtypes of the disease are Crohn's disease (CD) and ulcerative colitis (UC).[1,2] Patients are often diagnosed in early adulthood and experience a serious, life-altering condition that does not currently have a curative treatment. The disease course is highly heterogeneous, ranging from relatively mild symptoms to severe disease characterized by uncontrolled gastrointestinal inflammation and repeated hospitalizations and surgeries.[3,4] A better understanding of which patients will experience a more severe disease course would be the first step to improve treatment of this subgroup. However, our current limited understanding of IBD and the factors driving its heterogenicity makes it difficult to anticipate a patient's disease course.

Genome-wide association studies (GWAS) of IBD have primarily focused on disease onset by comparing patients with IBD to population controls. To date, more than 600 susceptibility loci have been identified.[5,6] However, the genetic contribution to disease progression over time remains largely unknown. In an analysis of approximately 2700 patients with CD performed by Lee *et al.*,[7] four individual loci were associated with the need for immunomodulators or gastrointestinal surgeries. These loci were not associated with disease susceptibility.

Polygenic scores (PGS) use the results from GWAS to calculate the aggregated effect of trait-associated variants carried by an individual. To date, PGS has primarily been used to model the individual risk of disease onset for complex diseases where a well-powered, large-scale GWAS exists. For IBD, PGS can predict disease onset with limited performance (area under the receiver operating characteristic curve (AUC) = 0·63) and only identifies ~3.2% of the population with a threefold increased risk of IBD.[8] As for genetics overall, very little is known about the impact of genetics on the course of IBD. Although a strong association between PGS and disease location has been reported,[9] previous studies have failed to identify an association between susceptibility PGS and anti-TNF response,[10] early onset,[11] and CD severity stratified by disease location.[7]

In a large, unique population-based IBD cohort with both genetics and long-term clinical data, we investigated the hypothesis that an individual's genetic burden, as measured by their IBD susceptibility PGS, impacts the severity of the disease course.

Methods and materials

Danish registries

The study was based on data from Danish registries, including the Danish Civil Registration System,[12] The Danish National Patient Registry (DNPR),[13] The Danish National Prescription Registry,[14] and The Danish nationwide Registry of Laboratory Results for Research (RLRR)[15]. They hold information on inpatient hospital contacts, International Classification of Diseases 8th and 10th revision (ICD-8 and -10) codes, surgical and other procedures since 1977, outpatient contacts since 1995, redeemed prescriptions for Danish residents at all pharmacies since 1994, and results of biochemistry and hematological tests from hospitals and general practitioners since 2008 with complete coverage from 2015. We had data available until September 2022. All registry codes used are listed in Table S1.

Source populations

The PREDICT neonatal blood spot cohort (PREDICT-NBS cohort). The Danish National Biobank stores neonatal blood spots from almost all Danes born since 1982.[16] The PREDICT-NBS cohort includes neonatal blood spots from individuals born and diagnosed with IBD between April 1981 and December 2018. Identification of patients with IBD was based on previously described criteria.[2] IBD patients should have at least two IBD-related hospital registrations within a two-year period recorded in the DNPR. Assignment of CD or UC was based on the two most recent hospital contacts if these were concordant, and otherwise according to the highest number of diagnoses. Date of diagnosis was the date of the first IBD-related hospital registration. Patients not living in Denmark during the two years before this first IBD-related hospital registration were excluded to ensure a valid date of diagnosis. Each IBD patient was matched to controls based on sex, date of birth, and no diagnosis of IBD prior to the patient's diagnosis date. Matched controls that developed IBD from end of cohort selection (December 2018) to end of available registry data (September 2022) were considered IBD patients.

The North Denmark IBD cohort (NorDIBD). The NorDIBD cohort is a population-based cohort of all patients from the North Denmark Region with a confirmed IBD diagnosis from 1978 to 2020.[17] From this cohort, 940 patients with IBD and a further 973 blood donors with no IBD diagnosis were genotyped from whole blood. The date of IBD diagnosis was based on patient health records.

Overlapping individuals between the two cohorts (n=233) were excluded randomly from one of the cohorts. Details on genotyping, quality control, and generation of genetic principal components (PCs) are previously described[18] and available in Supplement 1 and summarized in Figure S1.

114 **Study population**

115 We excluded patients diagnosed with IBD before January 1, 1996, and after September 2019, to ensure at
116 least three years of post-diagnosis follow-up in the registries. This resulted in a final study population of
117 3688 patients with CD (3340 from the PREDICT-NBS cohort and 348 from the NorDIBD cohort), and 4491
118 patients with UC (4153 from the PREDICT-NBS cohort and 338 from the NorDIBD cohort) (Figure S2).

119 **Polygenic score**

120 We downloaded and applied to our dataset variant loadings for calculating PGS for CD and UC susceptibility
121 generated by Middha *et al.*[19] They generated variant loadings from summary statistics from IIBDGC[20]
122 calibrated on UK Biobank data. We calculated PGS using Plink; 738,417 out of 744,682 (>99%) and 738,314
123 out of 744,575 (>99%) available variants were included for calculating CD and UC susceptibility PGS
124 respectively. We also calculated CD and UC susceptibility PGS for the controls. The mean (SD) of CD PGS
125 across controls and CD patients was 1.00 (1.27) and the corresponding mean (SD) of UC PGS across controls
126 and UC patients was -0.95 (1.01). The PGS scores were normalized to standard deviations (SDs) from the
127 mean value to improve interpretability.

128 **Outcomes**

129 Our primary outcome was severe versus less severe disease course. We defined a severe IBD course based
130 on having one of the following outcomes within three years after date of diagnosis: I) two or more IBD-
131 related hospitalizations exceeding two days, II) two or more IBD-related major surgeries, III) one IBD-related
132 hospitalization and one IBD-related major surgery not overlapping in time, or IV) a total use of at least 5000
133 mg systemic corticosteroids (Figure S3). The threshold for systemic corticosteroids was set based on
134 visualization of the data. Patients not fulfilling these criteria were considered less severe.

135 Secondary outcomes were age at diagnosis, and the following outcomes within three years after date of
136 diagnosis: two or more IBD-related hospitalizations exceeding two days, two or more IBD-related major
137 surgeries, one or more IBD-related major surgeries, a total use of at least 5000 mg systemic corticosteroids,
138 treatment with biologics, treatment with immunomodulators, other inflammatory immune mediated
139 diseases (IMDs), and days spent in hospital due to IBD. Additionally, we analyzed time to first IBD-related
140 hospitalization and time to first IBD-related major surgery. Finally, we examined markers of inflammation
141 i.e., C-reactive protein (CRP), fecal calprotectin (f-calpro) and hemoglobin at time of diagnosis and during
142 follow-up. At diagnosis was defined as ± six months from the date of diagnosis, and the post-diagnosis
143 phase was defined as three years starting from six months after date of diagnosis. Definitions and codes for
144 primary and secondary outcomes are listed in Table S1.

145 **Statistical analyses**

146 We analyzed the association between susceptibility PGS and the primary outcome (severe versus less
147 severe IBD) using logistic regression, adjusting for sex, cohort, calendar year, age at diagnosis, and the first
148 ten PCs. A similar model was used for the analysis of the binary secondary outcomes (two or more IBD-
149 related hospitalizations exceeding two days, two or more IBD-related major surgeries, one or more IBD-
150 related major surgeries, a total use of at least 5000 mg systemic corticosteroids, treatment with biologics,
151 treatment with immunomodulators, and other IMIDs). For the continuous secondary outcomes (age at IBD
152 diagnosis and days spent in hospital due to IBD), we used linear regression adjusting for the above-
153 mentioned covariates. All results are per SD increase of susceptibility PGS.

154 For analysis of time to first hospitalization and first major IBD-related surgery, we divided patients into PGS
155 quintiles and used Cox proportional hazards regression models to test for differences across PGS quintiles.
156 These survival analyses included patients previously excluded based on not having three years of follow-up
157 (CD=44, UC=44), and censoring was based on death, emigration, or end of study. The model included
158 adjustment for sex, cohort, calendar year, age at diagnosis, and the first ten PCs.

159 At- and post-diagnosis biochemistry test values of CRP, f-calpro, and hemoglobin were evaluated for
160 association with susceptibility PGS (per SD increase) using a linear regression model including the above-
161 mentioned covariates. Where there were multiple samples per individual, the sample closest to date of
162 diagnosis was chosen for the at-diagnosis interval, whereas the median value was chosen for the post-
163 diagnosis interval. Test values were transformed using log transformation for CRP and f-calpro to approach a
164 Gaussian distribution.

165 Reported p-values from tests of secondary outcomes and from biochemistry test values were adjusted for
166 multiple testing using the Bonferroni correction.

167 **Prediction models**

168 Random forest models were generated for predicting severe CD and UC based on patient information
169 available at date of diagnosis. We fitted a baseline model only including the covariates sex, cohort, calendar
170 year and age at diagnosis, parental IBD status, and having other IMIDs at time of diagnosis. The baseline
171 model was expanded to evaluate the increase in predictive power of adding a) susceptibility PGS, b) CRP,
172 and f-calpro test results within ± six months from the date of diagnosis, and c) both susceptibility PGS, CRP,
173 and f-calpro.

174 The random forest models were fitted on a random subset of 80% of the data, of which the less severe
175 group was randomly down-sampled to have the same number of patients as the severe group. Parameter
176 settings were tuned using three times ten-fold cross validation. The final model evaluation was based on
177 the remaining 20% validation data. The performance of the models was reported as the area under the
178 receiver operating curves (ROC-AUC).

179 **Ethical approval**

180 The PREDICT-NBS cohort was approved by the Research Ethical Committee of the Capital Region of
181 Denmark (H-20048987) and no direct consent was required. The NorDIBD cohort was approved by the
182 Research Ethical Committee of the North Denmark Region (N-20170005-N-20200071). Oral and written
183 consents were obtained, and we followed up on whether any participant withdrew their consent.

184 ## Results

185 We included 8179 patients with IBD (3688 (45%) patients with CD and 4491 (55%) with UC) and 9469
186 control individuals (Table 1). Among patients with CD, the mean age at diagnosis was 21·7 years (SD=7·4),
187 56% were female, and 1165 patients (31·6%) had a severe disease course based on three years of follow-up.
188 Among patients with UC, the mean age at diagnosis was 23·3 years (SD=8·4), 53% were female, and 1203
189 patients (26·8%) had a severe disease course. Control individuals had a mean normalized PGS of -0.221 for
190 CD PGS and -0.235 for UC PGS. Patients with CD had a mean normalized CD PGS of 0.528, whereas patients
191 with UC had a mean normalized UC PGS of 0.469.

192 **Higher genetic burden of IBD susceptibility is associated with a more severe disease course**

193 The PGS of IBD susceptibility was significantly associated with severe versus less severe disease for both CD
194 and UC (Figure 1A-B). Patients with severe CD had higher PGS than the remaining patients with CD (odds
195 ratio [OR]=1·26, 95% confidence interval [CI] =1·17-1·35 per SD increase). Accordingly, 39·3% of CD patients
196 in the highest PGS quintile experienced a severe disease course compared to 23·6% of CD patients in the
197 lowest PGS quintile. This was also the case for patients with UC (OR=1·35, 95% CI=1·26-1·45 per SD
198 increase), where 34·7% of patients in the highest PGS quintile experienced a severe disease course
199 compared to 20·1% in the lowest PGS quintile.

200 We investigated the association between susceptibility PGS and the secondary severity outcomes within the
201 follow-up period: two or more IBD-related hospitalizations exceeding two days, total hospitalization days,
202 two or more IBD-related major surgeries, one or more IBD-related major surgeries, a total use of at least
203 5000 mg systemic corticosteroids, treatment with biologics, treatment with immunomodulators,
204 occurrences of other IMIDs, and age at IBD onset (Figure 1C-D). In general, we observe that all individual
205 components of the main outcome were associated with susceptibility PGS with a consistent effect size
206 ranging from 1·24-1·50 and with overlapping CIs.

207 The most significant association between susceptibility PGS and CD severity was detected for treatment
208 with immunomodulators (OR=1·47, 95% CI=1·36-1·58 per SD increase), treatment with biologic therapies
209 (OR=1·35, 95% CI=1·25-1·45 per SD increase) and need for major surgery (OR=1·49, 95% CI=1·35-1·65 per
210 SD increase). CD PGS was further associated with having at least two hospitalization events, a higher
211 number of hospitalization days, use of at least 5000 mg of systemic corticosteroids, and with younger age at
212 diagnosis.

213 The most significant association between susceptibility PGS and UC severity was detected for having at least
214 two separate hospitalization events (OR=1·43, 95% CI=1·31-1·56 per SD increase) and for more days in

215 hospital (β=1·68, 95% CI=1·26-2·09 days of hospitalization per SD increase). UC PGS further associated with
216 need for treatment with biologic therapies and immunomodulators, use of at least 5000 mg of systemic
217 corticosteroids, and with the need of major surgery.

218 **Extending beyond three years of follow-up**

219 Separating the CD and UC cohorts into quintiles based on susceptibility PGS revealed cumulative incidence
220 curves ranking based on their PGS quintile, where the quintile with lowest PGS had the longest event-free
221 time (Figure 2). This pattern was consistent for both time to hospitalization and time to major surgery for
222 both CD and UC. Comparing the highest PGS quintile to the lowest revealed a quicker time to hospitalization
223 (CD: hazard ratio [HR]=1·81, 95% CI=1·57-2·09, UC: HR=1·69, 95% CI=1·47-1·95) and to major surgery (CD:
224 HR=2·74, 95% CI=2·18-3·45, UC: HR=2·04, 95% CI=1·55-2·69).

225 **Biochemistry profile is affected by susceptibility genetics**

226 Increased susceptibility PGS was associated with disease activity markers of gastrointestinal and systemic
227 inflammation (Figure 3). At diagnosis, f-calpro was increased (CD: β=0·32, 95% CI=0·23-0·42 mg/kg per SD
228 increase, UC: β=0·23, 95% CI=0·14-0·33 mg/kg per SD increase) and hemoglobin decreased (CD: β=-0·11,
229 95% CI=-0·16--0·07 mmol/L per SD increase, UC: β=-0·08, 95% CI=-0·12--0·03 mmol/L per SD increase) with
230 higher PGS, whereas CRP was only increased with higher PGS for patients with CD (β=0·17, 95% CI=0·10-
231 0·24 mg/L per SD increase). For patients with CD, increased f-calpro levels remained significantly associated
232 with higher PGS in the post-diagnosis phase (β=0·25, 95% CI=0·18-0·33 mg/kg per SD increase), whereas for
233 patients with UC, decreased hemoglobin levels remained significant (β=-0·04, 95% CI=-0·07--0·01 mmol/L
234 per SD increase).

235 **Predicting IBD severity at the time of diagnosis**

236 We constructed random forest models to separate severe and less severe disease courses of patients
237 diagnosed with IBD using information available at time of diagnosis (Figure 4). The down-sampled, balanced
238 dataset consisted of 367 severe and non-severe patients with CD and 440 severe **and** non-severe patients
239 with UC with information on all predictors. The AUC of the full models on the test data were 0·651 and
240 0·752 for CD and UC respectively. Susceptibility PGS ranked as the most important predictor of CD severity
241 and second most important predictor in UC severity. Adding CRP and f-calpro values at time of diagnosis
242 diminished the effect of PGS and ranked to be the most important predictors of both CD and UC severity.
243 Explicitly modelling HLA-DRB1*01:03[18] that is associated with both susceptibility to IBD and severity of UC
244 did not increase the performance of the model in UC (AUC=0·726).

Discussion

In this unique, population-wide study of >8000 patients with available GWAS data and long-term follow-up, we investigated the influence of IBD susceptibility genetics on the course of IBD using known susceptibility PGS for IBD and a range of longitudinal severity outcomes within the first three years, including hospitalizations, major surgeries, medication use, and test results from biochemistry laboratory tests.

We found an association between the aggregated genetic susceptibility to CD and UC and disease severity over time. In patients with CD, the susceptibility PGS associated most strongly with undergoing major surgery and needing treatment with biologics or immunomodulators. In patients with UC, the susceptibility PGS associated strongly with more frequent and longer hospitalizations, as well as with a need for systemic corticosteroids, immunomodulators, and biologics.

When splitting time to hospitalization and major surgery by PGS quintiles and introducing follow-up beyond three years, we observed an additive pattern for both CD and UC, where a higher genetic susceptibility to CD and UC associated with a proportionally higher hazard of hospitalization and major surgery.

Using longitudinal nationwide data from biochemistry laboratory tests, we further showed that the IBD susceptibility PGS is associated with more gastrointestinal inflammation in the form of higher values of f-calpro at time of diagnosis of both CD and UC and during follow-up in CD. Furthermore, susceptibility PGS associated with lower hemoglobin at diagnosis of CD and UC and during follow-up in UC, which is part of the diagnostic criteria for acute severe UC[21].

Our prediction models of IBD severity showed that CRP and f-calpro at time of diagnosis were the most predictive markers for future severe disease and that adding susceptibility PGS did not improve predictive power. The performance of future prediction models could potentially be improved by the creation of a future IBD "severity PGS" based on genetics studies of IBD disease course.

Complementing the findings from Lee *et al.*[7] and Cleynen *et al.*[9], our results suggest that IBD susceptibility genetics only explain part of the severity of IBD, hence leaving room for other factors at play. These should be investigated together with genetics and included in future prognostic prediction models. This observation is promising, as it points towards the possibility for modulating the effects of genetic susceptibility through lifestyle interventions directed towards potential environmental factors that are yet to be uncovered. Future prediction models of IBD severity may also include the gut microbiota profile[22] and additional biochemistry measurements to CRP and f-calpro.[23] The present study is hopefully a step towards the development of algorithms based on genetics, environmental exposures, and molecular pathways

275 which in the future can identify patients in need of increased monitoring and intensified treatment to
276 prevent their disease from progressing and improving their quality of life.

277 The primary strength of this study is its nationwide character and unselected nature, which assures
278 generalizability of findings, as well as the complete longitudinal follow-up of patients. Identifying IBD
279 patients using the Danish registries has previously been validated with a positive predictive value of 95%[24]
280 and completeness of 75-94%.[24,25] A further strength is the use of unique population-based samples for
281 genetic analyses, which was combined with lifetime health data.

282 Our study also has potential limitations, which need to be considered. The cohort is rather young, and our
283 results may primarily reflect the course of IBD in this part of the patient population. Also, we did not have
284 nationwide data on disease extent and behavior to study if the genetic burden of susceptibility modulates
285 disease severity through these factors.[9] However, the linearity of the association (especially for UC)
286 between the different susceptibility PGS quintiles and hospitalization and surgery risk suggests that the
287 effect is not limited to disease types that require more surgical interventions (ileal CD, extensive UC)
288 suggesting that the effect is driven by other factors also. We used comparable definitions of severe disease
289 for CD and UC which may be debated which is why we performed additional sensitivity analysis that
290 demonstrated the consistency of the association. Laboratory analyses were not available for all patients and
291 their availability may potentially reflect severity. However, no study has to our knowledge been able to test
292 all these parameters over time in a cohort where genetics were also available.

293 **Conclusion**

294 This study presents novel evidence of an association between the genetic susceptibility burden and
295 subsequent severity of IBD over time. Patients with a higher susceptibility PGS experience more
296 hospitalizations, surgeries and medical treatments, suggesting that genetics of disease susceptibility and
297 disease severity are not orthogonal. Our findings of a dose-response relationship between the genetic
298 burden of IBD susceptibility and disease severity provide a new biological understanding of disease severity.

Acknowledgements

We thank the nurses at the outpatient clinic at the Department of Gastroenterology and Hepatology at Aalborg University Hospital, Denmark, for their important contribution to the sample collection.

Declaration of interests

TJ reports consultancy for Ferring and Pfizer. LL reports speaker fee from Takeda and advisory board for Tillotts. HAJ reports speaker fee for Tillotts. The remaining authors have no disclosures.

Data sharing statement

This work uses data from the Danish National Health registries (https://sundhedsdatastyrelsen.dk), which is protected by the Danish Act on Processing of Personal Data. To access the data an application is required, which must be approved from the Danish Data Protection Agency and the Danish Health Data Authority. Handling of the data must be carried out by a Danish institution.

References

1 Ng SC, Shi HY, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017; **390**: 2769–78.

2 Agrawal M, Christensen HS, Bøgsted M, Colombel J-F, Jess T, Allin KH. The Rising Burden of Inflammatory Bowel Disease in Denmark Over Two Decades: A Nationwide Cohort Study. *Gastroenterology* 2022; **163**: 1547-1554.e5.

3 Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *The Lancet* 2017; **389**: 1741–55.

4 Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *The Lancet* 2017; **389**: 1756–70.

5 Liu Z, Liu R, Gao H, *et al.* Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet* 2023; **55**: 796–806.

6 Fachal L, Consortium on behalf of the IIG. OP08 Multi-ancestry genome-wide association study of inflammatory bowel disease identifies 125 novel loci and directly implicates new genes in disease susceptibility. *J Crohns Colitis* 2024; **18**: i14–5.

7 Lee JC, Biasci D, Roberts R, *et al.* Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nature Genetics 2017 49:2* 2017; **49**: 262–8.

8 Khera A V., Chaffin M, Aragam KG, *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; **50**: 1219.

9 Cleynen I, Boucher G, Jostins L, *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: A genetic association study. *The Lancet* 2016; **387**: 156–67.

10 Karmi N, Bangma A, Spekhorst LM, *et al.* Polygenetic risk scores do not add predictive power to clinical models for response to anti-TNFα therapy in inflammatory bowel disease. *PLoS One* 2021; **16**. DOI:10.1371/JOURNAL.PONE.0256860.

11 Serra EG, Schwerd T, Moutsianas L, *et al.* Somatic mosaicism and common genetic variation contribute to the risk of very-early-onset inflammatory bowel disease. *Nature Communications 2020 11:1* 2020; **11**: 1–15.

338 12 Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology.
339 *Eur J Epidemiol* 2014; **29**: 541–9.

340 13 Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National
341 Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**:
342 449.

343 14 Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource
344 Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017; **46**: 798–798f.

345 15 Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing Data Sources in
346 Clinical Epidemiology: Laboratory Information System Databases in Denmark. *Clin Epidemiol* 2020;
347 **12**: 469–75.

348 16 Nørgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening
349 Biobank. *J Inherit Metab Dis* 2007; **30**: 530–6.

350 17 Larsen L, Karachalia Sandri A, Fallingborg J, *et al.* Has the Incidence of Inflammatory Bowel Disease
351 Peaked? Evidence From the Population-Based NorDIBD Cohort 1978-2020. *Am J Gastroenterol* 2023;
352 **118**: 501–10.

353 18 Vestergaard MV, Nøhr AK, Allin KH, *et al.* HLA-DRB1*01:03 and Severe Ulcerative Colitis. *JAMA* 2024;
354 published online Oct 15. DOI:10.1001/JAMA.2024.20429.

355 19 Middha P, Thummalapalli R, Betti MJ, *et al.* Polygenic risk score for ulcerative colitis predicts immune
356 checkpoint inhibitor-mediated colitis. *Nature Communications 2024 15:1* 2024; **15**: 1–10.

357 20 De Lange KM, Moutsianas L, Lee JC, *et al.* Genome-wide association study implicates immune
358 activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017; **49**: 256.

359 21 Kedia S, Ahuja V, Tandon R. Management of acute severe ulcerative colitis. *World J Gastrointest*
360 *Pathophysiol* 2014; **5**: 579.

361 22 Buffet-Bataillon S, Bouguen G, Fleury F, Cattoir V, Le Cunff Y. Gut microbiota analysis for prediction of
362 clinical relapse in Crohn's disease. *Scientific Reports 2022 12:1* 2022; **12**: 1–8.

363 23 Vestergaard MV, Allin KH, Poulsen GJ, Lee JC, Jess T. Characterizing the pre-clinical phase of
364 inflammatory bowel disease. *Cell Rep Med* 2023; **4**: 101263.

365 24 Jacobsen HA, Jess T, Larsen L. Validity of Inflammatory Bowel Disease Diagnoses in the Danish
366 National Patient Registry: A Population-Based Study from the North Denmark Region. *Clin Epidemiol*
367 2022; **14**: 1099–109.

368 25 Fonager K, Toft Sørensen H, Nørby Rasmussen S, Møller-Petersen J, Vyberg M. Assessment of the
369 diagnoses of Crohn's disease and ulcerative colitis in a Danish hospital information system. *Scand J*
370 *Gastroenterol* 1996; **31**: 154–9.

371

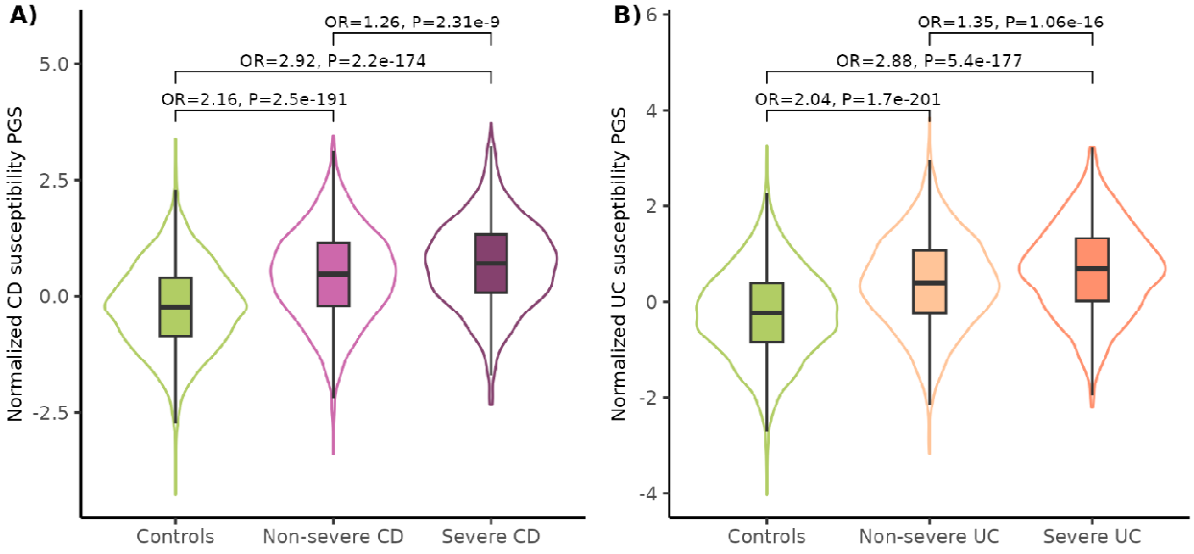
372

Tables and figures

Table 1: Baseline characteristics of study population. Numbers (percentage) are reported in the table.

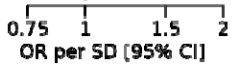
	All N = 8179 (100)		The PREDICT-NBS cohort N = 7493 (91.6)		The NorDIBD cohort N = 686 (8.4)	
	CD	**UC**	**CD**	**UC**	**CD**	**UC**
N	3688 (45.1)	4491 (54.9)	3340 (44.6)	4153 (55.4)	348 (50.7)	338 (49.3)
Sex						
Female	2056 (55.7)	2400 (53.4)	1875 (56.1)	2211 (53.2)	181 (52.0)	189 (55.9)
Male	1632 (44.3)	2091 (46.6)	1465 (43.9)	1942 (46.8)	167 (48.0)	149 (44.1)
Age at diagnosis						
<17 years	822 (22.3)	772 (17.2)	783 (23.4)	754 (18.2)	39 (11.2)	18 (5.3)
17-40 years	2790 (75.7)	3580 (79.7)	2557 (76.6)	3399 (81.8)	233 (67.0)	181 (53.6)
≥40 years	76 (2.1)	139 (3.1)	-	-	76 (21.8)	139 (41.1)
Calendar year at diagnosis						
1996-2005	534 (14.5)	576 (12.8)	433 (13.0)	488 (11.8)	101 (29.0)	88 (26.0)
2005-2015	1936 (52.5)	2470 (55.0)	1783 (53.4)	2308 (55.6)	153 (44.0)	162 (47.9)
2015-2020	1218 (33.0)	1445 (32.2)	1124 (33.7)	1357 (32.7)	94 (27.0)	88 (26.0)

CD: Crohn's disease, UC: ulcerative colitis, PREDICT-NBS: PREDICT neonatal blood spots, NorDIBD: North Danmark inflammatory bowel disease



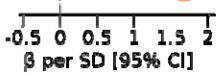
C)

	Disease	N	OR per SD [95% CI]	P adjusted
Main outcome	CD	1165 vs. 2523	1.26 [1.17—1.35]	2.31×10^{-8}
	UC	1203 vs. 3288	1.35 [1.26—1.45]	1.07×10^{-15}
≥2 hospitalizations	CD	718 vs. 2523	1.27 [1.16—1.38]	1.76×10^{-6}
	UC	692 vs. 3288	1.43 [1.31—1.56]	1.92×10^{-14}
≥2 major surgeries	CD	718 vs. 2523	1.27 [1.16—1.38]	1.76×10^{-6}
	UC	242 vs. 3288	1.50 [1.30—1.72]	1.4×10^{-7}
≥1 major surgery	CD	567 vs. 2373	1.49 [1.35—1.65]	7.31×10^{-14}
	UC	358 vs. 3266	1.43 [1.28—1.61]	1.08×10^{-8}
≥5000 mg systemic corticosteroids	CD	654 vs. 2523	1.24 [1.13—1.36]	6.52×10^{-5}
	UC	821 vs. 3288	1.32 [1.22—1.44]	1.95×10^{-10}
Required biologic therapies	CD	1438 vs. 1914	1.35 [1.25—1.45]	6.01×10^{-14}
	UC	908 vs. 3241	1.31 [1.21—1.42]	4.56×10^{-10}
Required immunomodulators	CD	2081 vs. 1271	1.47 [1.36—1.58]	3.51×10^{-22}
	UC	1292 vs. 2857	1.29 [1.20—1.38]	1.08×10^{-11}
Other IMIDs	CD	561 vs. 3127	1.03 [0.94—1.14]	1
	UC	566 vs. 3925	0.95 [0.86—1.04]	1



D)

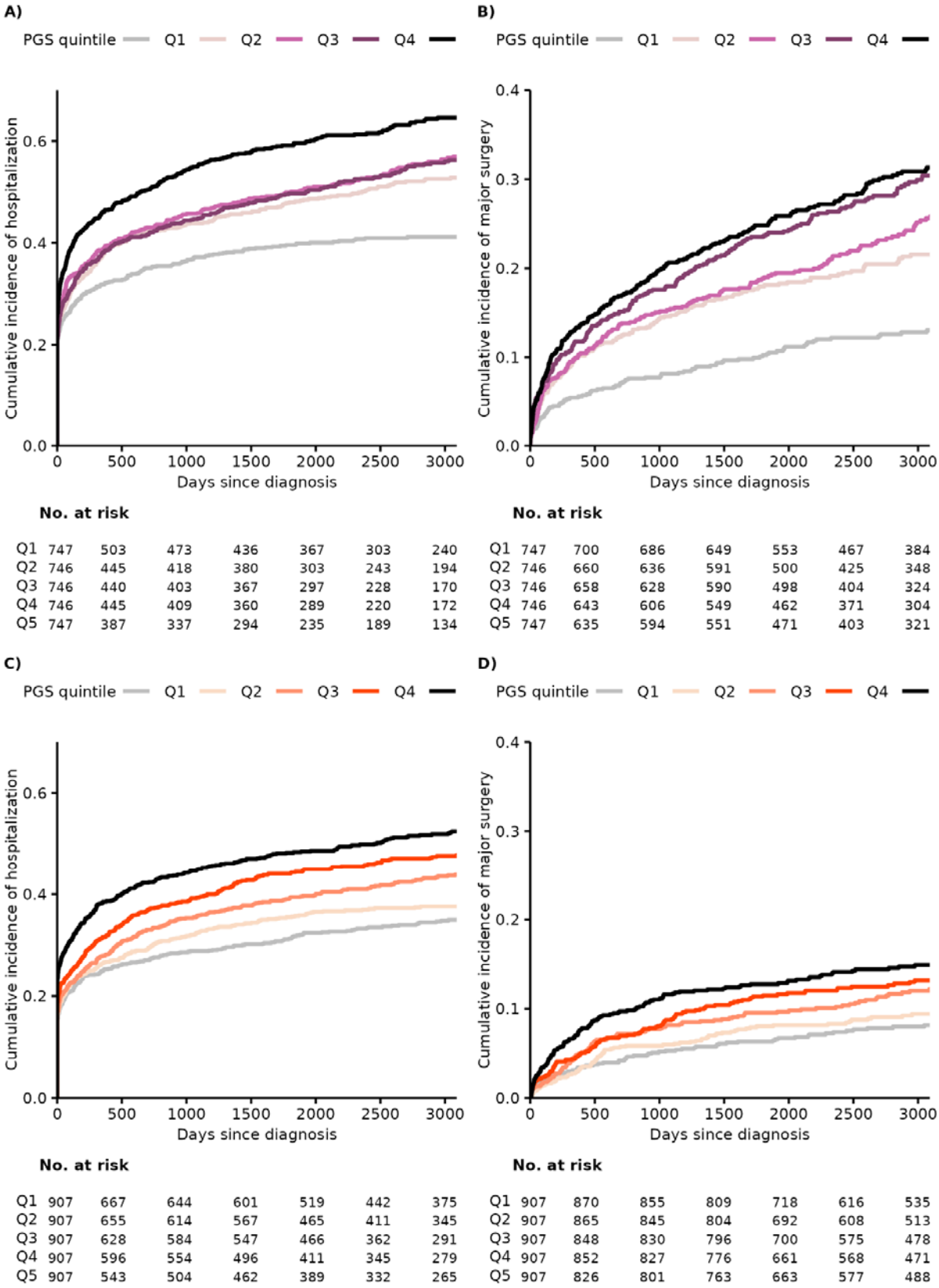
	Disease	β per SD [95% CI]	P adjusted
Age at diagnosis (years)	CD	-0.34 [-0.55—-0.12]	0.021
	UC	-0.04 [-0.24—0.16]	1
Hospitalization days (N)	CD	1.01 [0.50—1.51]	9.76×10^{-4}
	UC	1.68 [1.26—2.09]	2.25×10^{-14}



379 **Figure 1: Susceptibility PGS related to IBD severity. A)** Violin plot of PGS values of CD susceptibility
380 (normalized to SDs from the mean) for controls and CD severity groups. Summary statistics are outputted
381 from logistic regressions. **B)** Violin plot of PGS values of UC susceptibility (normalized to SDs from the mean)
382 for controls and UC severity groups. Summary statistics are outputted from logistic regressions. We
383 analyzed the association between the PGS of susceptibility for developing CD and UC and ten different
384 severity outcomes. Binary outcomes **C)** were analyzed using logistic regression and the OR (±95% CI) per SD
385 increase in PGS is shown for CD and UC patients separately. Numerical outcomes **D)** were analyzed using
386 linear regression and the β estimates (±95% CI) per SD increase in PGS are shown for CD and UC patients
387 separately. *P*-values were adjusted for multiple testing using the Bonferroni correction for ten tests.

388 CD: Crohn's disease, UC: ulcerative colitis, IBD: inflammatory bowel disease, OR: odds ratio, PGS: polygenic
389 score, SD: standard deviation, IMID: immune-mediated inflammatory disease, CI: confidence interval

390


A)
PGS quintile Q1 Q2 Q3 Q4
Cumulative incidence of hospitalization
0.0 0.2 0.4 0.6
0 500 1000 1500 2000 2500 3000
Days since diagnosis
No. at risk
Q1 747 503 473 436 367 303 240
Q2 746 445 418 380 303 243 194
Q3 746 440 403 367 297 228 170
Q4 746 445 409 360 289 220 172
Q5 747 387 337 294 235 189 134
B)
PGS quintile Q1 Q2 Q3 Q4
Cumulative incidence of major surgery
0.0 0.1 0.2 0.3 0.4
0 500 1000 1500 2000 2500 3000
Days since diagnosis
No. at risk
Q1 747 700 686 649 553 467 384
Q2 746 660 636 591 500 425 348
Q3 746 658 628 590 498 404 324
Q4 746 643 606 549 462 371 304
Q5 747 635 594 551 471 403 321
C)
PGS quintile Q1 Q2 Q3 Q4
Cumulative incidence of hospitalization
0.0 0.2 0.4 0.6
0 500 1000 1500 2000 2500 3000
Days since diagnosis
No. at risk
Q1 907 667 644 601 519 442 375
Q2 907 655 614 567 465 411 345
Q3 907 628 584 547 466 362 291
Q4 907 596 554 496 411 345 279
Q5 907 543 504 462 389 332 265
D)
PGS quintile Q1 Q2 Q3 Q4
Cumulative incidence of major surgery
0.0 0.1 0.2 0.3 0.4
0 500 1000 1500 2000 2500 3000
Days since diagnosis
No. at risk
Q1 907 870 855 809 718 616 535
Q2 907 865 845 804 692 608 513
Q3 907 848 830 796 700 575 478
Q4 907 852 827 776 661 568 471
Q5 907 826 801 763 663 577 488

391

392 **Figure 2: Time to severity event split by PGS quintiles. A)** Cumulative incidence of IBD-related
393 hospitalization split by CD susceptibility PGS quintiles within the CD cohort. **B)** Cumulative incidence of IBD-
394 related major surgery split by CD susceptibility PGS quintiles within the CD cohort. **C)** Cumulative incidence
395 of IBD-related hospitalization split by UC susceptibility PGS quintiles within the UC cohort. **D)** Cumulative
396 incidence of IBD-related major surgery split by UC susceptibility PGS quintiles within the UC cohort.

397 IBD: Inflammatory bowel disease, UC: ulcerative colitis, CD: Crohn's disease, PGS: polygenic score, Q1-Q5:
398 Quintiles 1 to 5, where Q1 has the lowest quintile of PGS scores, and Q5 the highest.

399

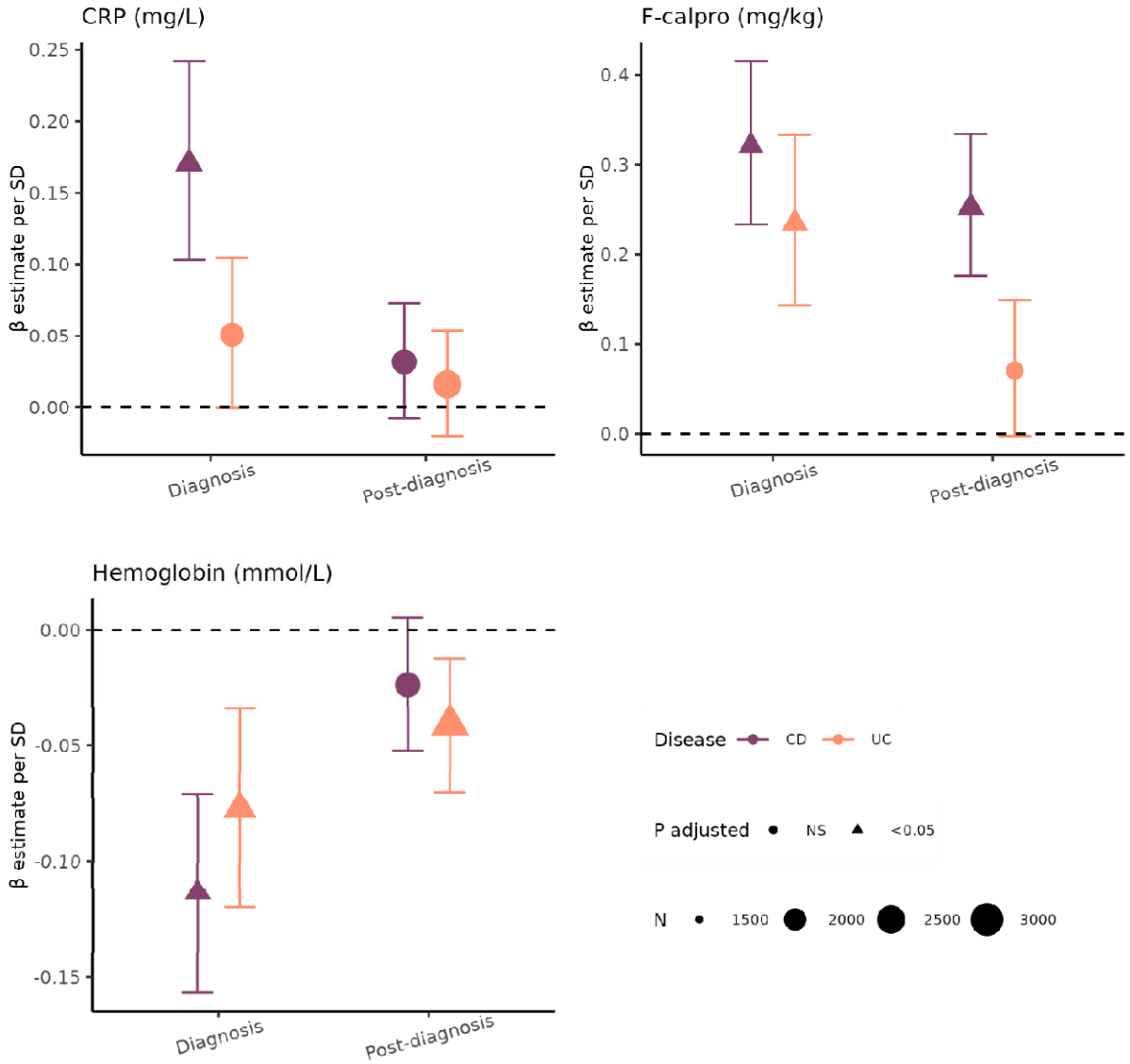


Figure 3: Association between susceptibility PGS and laboratory biochemical test results at different time points. Plot showing β estimates and their 95% CI for association between laboratory test results (CRP, f-calpro, and hemoglobin) and SD standardized susceptibility PGS for CD and UC separately. For each individual, the test results were calculated as following: Diagnosis = test result closest to date of diagnosis (±six months), and post-diagnosis = the median value of all tests from six months to three years after date of diagnosis. Back-transformed estimates and 95% CIs are showed. *P*-values were adjusted for multiple testing using the Bonferroni correction for three tests. The dots are shaped based on significance after adjusting for multiple testing. Dot sizes are based on number of individuals with available test results in that given period.

410 CD: Crohn's disease, UC: ulcerative colitis, PGS: polygenic score, SD: standard deviation, CRP: C-reactive
411 protein, f-calpro: fecal calprotectin, CI: confidence interval, NS: not significant
412

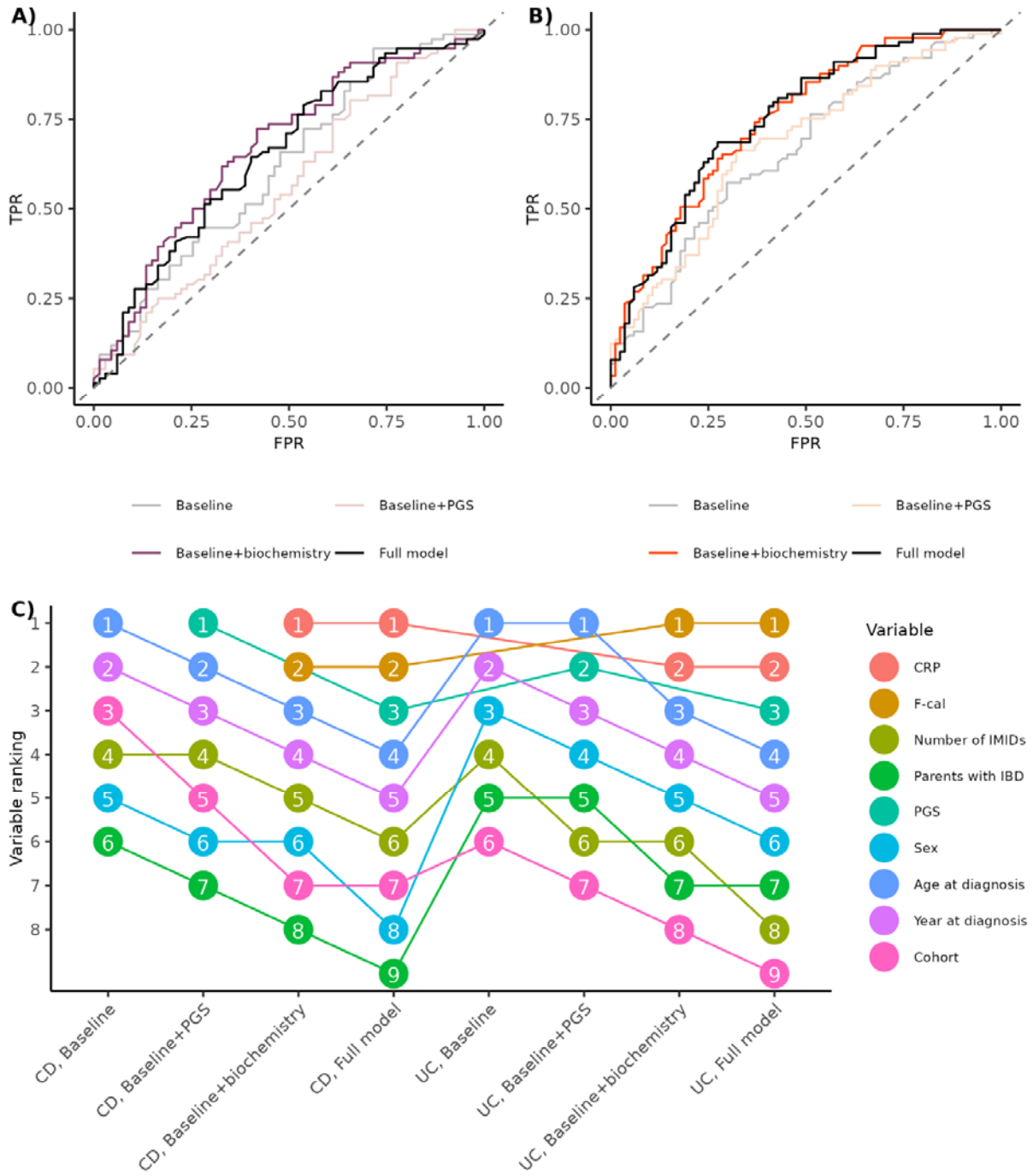


Figure 4: Results of IBD severity prediction models. Random forest models were fitted using patient information available at date of diagnosis as predicters. Receiver operating characteristic (ROC) curves for the performance on a test dataset is shown for **A)** CD and **B)** UC. **C)** Variable importance for predicting IBD severity is shown based on ranking by decreasing mean decrease in Gini index.

418 CD: Crohn's disease, UC: ulcerative colitis, IBD: inflammatory bowel disease, PGS: polygenic score, IMID:
419 immune-mediated inflammatory disease, CRP: C-reactive protein, f-calpro: fecal calprotectin, TPR: true
420 positive rate, FPR: false positive rate