Title

Large-scale clustering of longitudinal faecal calprotectin and C-reactive protein profiles in inflammatory bowel disease

Short title

Longitudinal biomarker clustering in inflammatory bowel disease

Authors

Nathan Constantine-Cooke[1,2], Nikolas Plevris[1,3], Karla Monterrubio-Gómez[1,2], Clara Ramos Belinchón[1,3], Solomon Ong[3], Alexander T. Elford[3,4], Beatriz Gros[3,5], Gareth-Rhys Jones[3,6], Charlie W. Lees[† 1,3] and Catalina A. Vallejos[† 2].

[†] Joint senior authors

Affiliations

1. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
2. MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
3. Edinburgh IBD Unit, Western General Hospital, Edinburgh, UK
4. Faculty of Medicine, The University of Melbourne, Melbourne, Australia
5. Reina Sofía University Hospital, Gastroenterology and Hepatology. IMIBIC. University of Cordoba. Cordoba, Spain
6. Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK

Abstract

Background

Crohn's disease (CD) and ulcerative colitis (UC) are highly heterogeneous, dynamic and unpredictable, with a marked disconnect between symptoms and intestinal

inflammation. Attempts to classify inflammatory bowel disease (IBD) subphenotypes to inform clinical decision making have been limited. We aimed to describe the latent disease heterogeneity by modelling routinely collected faecal calprotectin (FC) and CRP data, describing dynamic longitudinal inflammatory patterns in IBD.

Methods

In this retrospective study, we analysed patient-level longitudinal measurements of FC and CRP recorded within seven years since diagnosis. Latent class mixed models (LCMMs) were used to cluster individuals with similar longitudinal FC or CRP profiles. Associations between cluster assignment and information available at diagnosis (e.g. age, sex, and Montreal classification) were quantified using multinomial logistic regression. Differences in advanced therapy use across clusters were also explored using cumulative distributions over time. Finally, we considered uncertainty in cluster assignments with respect to follow-up length and explored the overlap between clusters identified based on FC and CRP.

Findings

We included 1036 patients (544 CD, 380 UC, 112 IBD-unclassified (IBDU)) in the FC analysis with a total of 10545 FC observations (median 9 per subject, IQR 6–13). The CRP analysis consisted of 1838 patients (805 CD, 847 UC, 186 IBDU) with 49364 CRP measurements (median 20 per subject; IQR 10–36). Eight distinct clusters of inflammatory behaviour over time were identified by LCMM in each analysis. The clusters, FC1-8 and CRP1-8, were ordered from the lowest cumulative inflammatory burden to the highest. The clusters included groups with high diagnostic levels of inflammation which rapidly normalised, groups where high inflammation levels persisted throughout the full seven years of observation, and a series of intermediates including delayed remitters and relapsing remitters.

CD and UC patients were unevenly distributed across the clusters. In CD, whilst patients with upper GI involvement (L4) were less likely to be in FC1 and FC2, there was no impact on ileal versus colonic disease on cluster assignment. In UC, male sex was associated with the poorest prognostic cluster (FC8). The use and timing of advanced therapy was associated with cluster assignment, with the highest use of early advanced therapy in FC1. Of note, FC8 and CRP8 captured consistently high

patterns of inflammation despite a high proportion of patients receiving advanced therapy, particularly for CD individuals (56.8% and 33.3%, respectively). We observed that uncertainty in cluster assignments was higher for individuals with short longitudinal follow-up, particularly between clusters capturing similar earlier inflammation patterns. There was broadly poor agreement between FC and CRP clusters in keeping with the need to monitor both in clinical practice.

Interpretation

Distinct patterns of inflammatory behaviour over time are evident in patients with IBD. Cluster assignment is associated with disease type and both the use and timing of advanced therapy. These data pave the way for a deeper understanding of disease heterogeneity in IBD and enhanced patient stratification in the clinic.

Introduction

Inflammatory bowel disease (IBD), an umbrella term for Crohn's disease (CD), ulcerative colitis (UC) and inflammatory bowel disease unclassified (IBDU), has a prevalence of almost 1% in the UK population.[1,2] The condition is characterised by chronic relapsing and remitting inflammation of the gastrointestinal (GI) tract that confers a host of debilitating symptoms, negatively impacting quality of life.[3,4] Studies have clearly demonstrated that uncontrolled inflammation of the GI tract increases the risk of disease progression and complications including the development of colorectal cancer,[5] stricturing/penetrating complications and surgery.[6,7] However, IBD is highly heterogeneous with respect to symptoms, inflammatory burden, treatment response, and long-term outcomes.

Current IBD classification methods are mostly based on historic nomenclature which utilise baseline phenotypic characteristics and do not take into account the dynamic and unpredictable nature of the disease. Attempts at developing prediction tools to identify high risk patients have been made but again these use static clinical parameters, dismissing the changing nature of the disease.[8] Furthermore, they do not take into account the influence of other factors, such as advanced therapy timing, on the disease course. In the wake of the increasing prevalence of IBD and associated healthcare burden, new methods to characterise the dynamic disease course and help

identify at-risk individuals who require aggressive therapy with close monitoring versus those that need less intense input are essential.

The original IBSEN studies provided the first data on the clinical course of patients with UC and CD during the first 10-years of diagnosis.[9,10] However, these data were based on predefined disease patterns, rather than data driven, utilising symptoms alone. It is now widely accepted that there is a clear disconnect between inflammation and symptoms in IBD,[11] therefore it is imperative characterisations of disease course include objective parameters of inflammation.

C-reactive protein (CRP) and faecal calprotectin (FC) are well established tools for monitoring patients with IBD, but are typically interpreted in terms of the most recent measurement or short-term trends. Interrogation of long-term trends of inflammation could greatly assist clinical decision making and improve prediction of future events. Moreover, modelling inflammatory behaviour over time might be a key tool for characterising the largely unexplained heterogeneity seen in IBD, and provide new tools for disease sub-phenotyping beyond the current Montreal classification.[12]

In this study, we aimed to 1) identify groups of IBD patients with similar longitudinal patterns of inflammation 2) determine if these groupings were associated with age, sex, IBD type, Montreal classification, or early advanced therapy, and 3) explore whether subjects with similar longitudinal FC profiles also share similar CRP profiles.

Methods

Ethics

This project was approved by the local Caldicott Guardian (Project ID: CRD18002, registered NHS Lothian information asset #IAR-954). Patients or the public were not involved in the design, conduct, reporting or dissemination plans of our research.

Study design

This was a retrospective cohort study. Patients with a confirmed diagnosis of IBD (as per Lennard-Jones criteria)[13] were followed up for a period of seven years from the date of diagnosis. Baseline phenotype data (sex, age at diagnosis, IBD type, date of diagnosis) were obtained from the Lothian IBD registry (LIBDR), a retrospective cohort

of patients receiving IBD care in Lothian, Scotland. The LIBDR is estimated to have identified 94·3% of all true IBD patients in the area using a capture-recapture approach.[1] Using a population level cohort reduces potential biases associated with cohort recruitment.[14] When available, additional phenotyping information was extracted by the clinical team from electronic health records (TrakCare; InterSystems, Cambridge, MA). This includes smoking and Montreal classification for disease location, behaviour and extent, all recorded as per patient status at the time of diagnosis. Data on prescribing of all advanced therapies, including start/stop dates, were extracted from both TrakCare and NHS Lothian pharmacy databases. Primary care prescribing data were not available. See Supplementary Note 1 for more detailed definitions.

Inclusion/exclusion criteria

Subjects were required to have a confirmed diagnosis of IBD at any age and receive secondary care for their condition from the NHS Lothian health board. Only subjects with a recorded date of IBD diagnosis between 2005 and 2019 were included. The lower bound for this criteria was established as FC testing was not routinely performed prior to this date. The upper bound ensured subjects had the possibility of having at least five years of follow-up at the time of data extraction.

For subjects which met the above requirements, the following criteria was applied to their FC and/or CRP longitudinal measurements. Subjects were required to have a diagnostic measurement (± three months of diagnosis, Figures S1 and S2) and have a further two observations available within seven years of diagnosis. If any biomarker measurements were observed within three months prior to the recorded diagnosis date, measurement time scales were realigned with respect to diagnosis (Figure S3). Only non-censored observations were considered in this calculation for FC. For CRP, this filtering was applied after preprocessing (see “Longitudinal measurements and preprocessing” section), and subjects with constant biomarker measurements over time were excluded. As the clustering analyses were performed separately for FC and CRP, subjects did not need to meet the criteria for both biomarkers.

Statistical analysis

Cohort description

Continuous variables were summarised as their median and interquartile range (IQR). Categorical variables were summarised as counts and percentages.

Longitudinal measurements and preprocessing

FC and CRP measurements were obtained from an extract by the local biochemistry team describing tests recorded up to August 13, 2024. For each individual, all measurements made within seven years from diagnosis were considered. Failed tests, for example due to contamination, were discarded. All FC tests were performed from stool samples using the same ELISA technology.[15] Due to limits of detection, observations < 20 µg/g were recorded to 20 µg/g whilst observations > 1250 µg/g were mapped to 1250 µg/g. Such observations were treated as censored when applying the inclusion exclusion criteria described above. CRP was measured from blood samples; observations for which only an upper bound was available, e.g. < 1 mg/L, were mapped to the corresponding upper bound.

Further processing was applied to CRP data to smooth out short-term fluctuations. Measurements were grouped into intervals of t: [0, 0·5), [0·5, 1·5), [1·5, 2·5), [2·5, 3·5), [3·5, 4·5), [4·5, 5·5), [5·5, 7], where t = 0 (years) is the time of diagnosis. The median CRP for each interval was calculated for each subject and used as input for subsequent analyses. The centre of each interval was used as the corresponding observation time.

Longitudinal biomarker clustering

Prior to model fitting, FC and CRP observations were log-transformed. FC and CRP trajectories were modelled separately using latent class mixed models (LCMMs),[16] an extension of linear mixed effects models that enables clustering of individuals that share similar longitudinal biomarker trajectories. LCMM consists of two submodels: one which captures the longitudinal behaviour of the biomarker, and one which captures cluster assignment. Fixed effects for the longitudinal submodel were specified using natural cubic splines with three interior knots placed at quantiles with respect to observation times. An alternative specification was considered leading to

similar results. Only the intercept was treated as a random effect. The cluster assignment submodel used IBD type (CD, UC, and IBDU) as a covariate. Formal definitions of the LCMM models considered here and the associated hyper-parameter choices are provided in Supplementary Note 2.

As the number of clusters is not known a priori, the optimal model was found using a grid search approach. We considered models with 2–10 clusters for both FC and CRP. The likelihood based statistics,[17] Akaike information criterion (AIC) and Bayesian information criterion (BIC), and visual inspection of cluster trajectories were used to compare models with different specifications and determine the most appropriate number of assumed clusters.

LCMM calculates, for each individual, the probability of being assigned to each cluster. In subsequent analysis, each individual was assigned to the cluster with the highest probability. The distribution of cluster assignment probabilities was used to assess uncertainty in these allocations with respect to follow-up length. This was defined as the time difference between diagnosis and the last available biomarker measurement (FC or CRP, depending on the analysis). For each cluster, the average probability of individual-specific probabilities of cluster assignment were reported.

To avoid displaying potentially identifiable individual-level data, exemplar trajectories within each cluster are visualised as aggregated trends, where measurements were summarised as the median across six randomly selected individuals. Cluster labels were ordered based on the area under the overall biomarker trend inferred for each cluster (a proxy for cumulative inflammatory burden).

Associations with respect to cluster assignments

To facilitate the interpretation of each cluster, we considered potential associations between cluster assignments and patient-level information. Violin plots and percentage bar plots were used as a visual summary when considering continuous (age) and discrete (sex, IBD type, additional phenotyping) patient-level covariates, respectively. Associations with respect to additional phenotyping were only explored after stratifying by IBD type (CD and UC only). Multinomial logistic regression was used to quantify associations between cluster assignments and patient-level covariates. The cluster used as the reference class was chosen to closely resemble the overall distribution of IBD types within the corresponding cohort. Univariate and

multivariate models were considered, and the associated 95% confidence intervals are reported. Individuals with missing covariate values were excluded when fitting each model. Due to small cluster sizes and low frequency of some covariate levels, the analysis was repeated after merging clusters with similar cumulative inflammatory burden. Effect sizes were largely consistent (data not shown).

Advanced therapy use

To compare patterns of advanced therapy (AT) across clusters, the cumulative distribution of first-line advanced therapy use was calculated. Results are reported stratified by IBD type (CD and UC only).

Comparison between FC and CRP cluster assignments

For subjects meeting the criteria for both FC and CRP analyses, the relationship between FC and CRP cluster assignment was visualised using alluvial plots and side-by-side comparisons of mean cluster trajectories for the optimal models. Stratified results for CD and UC subjects are also reported.

Software

R (v. 4·4·0), extended using the lcmm (v. 2·1·0),[18] ggalluvial (v.0·12·5),[19] and datefixR (v.1·6·1)[20] packages, was used for all analyses. Analytical reports have been generated using the Quarto scientific publishing system and are hosted online (https://vallejosgroup.github.io/Lothian-IBDR/). An R package, libdr (v.1·0·0), has also been produced, supporting the reuse of our R code with other datasets.

Role of the funding source

Funders were not involved in the study design, collection, analysis, or interpretation of the data, writing, or decision to submit the paper for publication.

Results

Cohort derivation and description

Of the 10153 subjects with a confirmed IBD diagnosis, 5508 were reported as being diagnosed between 2005 and 2019 (Figure 1). Of these subjects, 1036 and 1838 subjects were included in the FC and CRP analysis, respectively. We identified 808 subjects which met the inclusion criteria for both biomarkers. Table 1 describes key demographic factors for subjects included in the FC and CRP analyses.

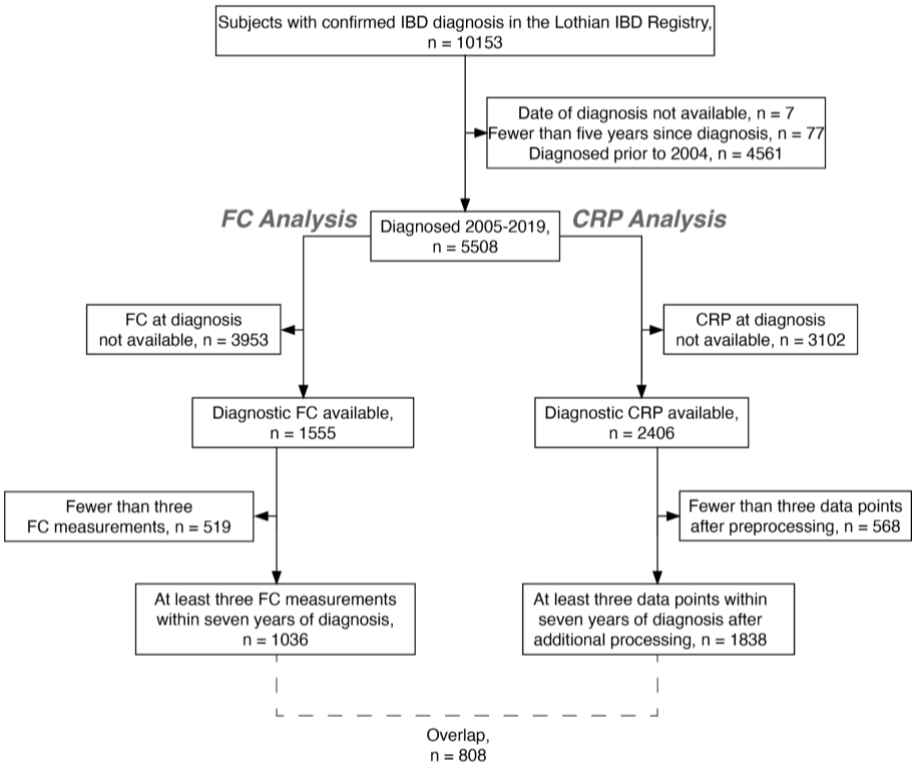

Figure 1. Derivation of the study cohorts based on separate faecal calprotectin (FC) and C-reactive protein (CRP) inclusion/exclusion criteria.

	FC analysis (n = 1036)	CRP analysis (n = 1838)	Overlap (n = 808)
Age at diagnosis	33 (25–75)	37 (24–55)	32 (20–50)
Male sex	503 (48·6%)	956 (52·0%)	397 (49·1%)
IBD type			
Crohn's disease (CD)	544 (52·5%)	805 (43·8%)	451 (55·8%)
Ulcerative colitis (UC)	380 (36·7%)	847 (46·1%)	276 (34·2%)

IBDU	112 (10·8%)	186 (10·1%)	81 (10·0%)
Baseline FC (µg/g)	740 (320–1070)	..	760 (330–1081)
FC observations	9 (6–13)	..	9 (6–14)
Baseline CRP (µg/mL)	..	8 (3–24)	8 (3–20)
Total unprocessed CRP observations	..	20 (10–36)	26 (14–36)
Total CRP observations after pre-processing	..	6 (4–7)	6 (5–7)
Advanced therapies (AT)			
AT within seven years for CD	270 (49·6%)	374 (46·5%)	
AT within one year for CD	129 (23·7%)	167 (20·7%)	
AT within seven years for UC	108 (28·4%)	85 (10·0%)	
AT within one year for UC	42 (11·1%)	35 (4·1%)	
Crohn's disease			
Montreal location			
L1	168 (30·9%)	249 (30·9%)	136 (30·2%)
L2	200 (36·8%)	287 (35·7%)	157 (34·8%)
L3	169 (31·1%)	246 (30·6%)	153 (33·9%)
Missing	7 (1·3%)	23 (2·9%)	5 (1·1%)
Upper GI inflammation			
Present	77 (14·2%)	87 (10·8%)	68 (15·1%)
Not Present	467 (85.8·4%)	718 (89·2%)	383 (84·9%)
Montreal behaviour			
B1	443 (81·4%)	607 (75·4%)	364 (80·7%)
B2	63 (11·5%)	110 (13·7%)	54 (12·0%)
B3	28 (5·1%)	61 (7·6%)	25 (5·5%)

Missing	10 (1·8%)	27 (3·4%)	8 (1·8%)
Perianal disease			
Present	79 (14·5%)	113 (14·0%)	68 (15·1%)
Not Present	459 (84·4%)	673 (83·6%)	379 (84·0%)
Missing	6 (1·1%)	19 (2·4%)	4 (0·9%)
Smoking status			
Current or previously	166 (30·5%)	274 (34·0%)	138 (30·6%)
Never	352 (64·7%)	484 (60·1%)	293 (65·0%)
Missing	26 (4·8%)	47 (5·8%)	20 (4·4%)

Table 1. Demographic and clinical data at diagnosis for subjects meeting the faecal calprotectin (FC) or C-reactive protein (CRP) study inclusion criteria. Continuous data are presented as median and interquartile range. Categorical data are presented as counts and percentages. Missingness is only directly reported if values were missing. The column labelled as "Overlap" denotes subjects which met the inclusion criteria for both FC and CRP modelling. Missing observations for upper gastrointestinal inflammation were assumed to be "not present" (Supplementary Note 1). Missingness was not inferred to be a value for any other variable.

Longitudinal measurements of FC and CRP

For subjects in the FC analysis, 10545 FC observations were available (median 9 per subject, IQR 6–13). Prior to processing 49364 CRP observations were available (median 20 per subject, IQR 10–36). Following the pre-processing of CRP observations, there were 9898 data points (median 6 per subject, IQR 4–7). The distribution of log-transformed FC and CRP values is shown in Figure S4.

Modelling of FC trajectories

AIC suggested the 10-cluster model (Figure S5), whilst BIC suggested the 9-cluster model (Figure S6) was more appropriate (Table S1, Figure S7A). However, the 8-cluster model was chosen as a parsimonious choice, as it captures the main observed longitudinal patterns without generating very small clusters (<50 individuals) which could be difficult to interpret.

Figure 2 shows representative cluster profiles for the 8-cluster model, ordered from lowest (FC1) to highest (FC8) cumulative inflammatory burden. FC2 (n=67; ~6%)

represents low FC values throughout the whole observation period. Instead, FC1 (n=140; ~14%) and FC3 (n=157; ~15%) and FC7 (n=244; ~24%) were characterised by initially high FC values (>250 µg/g) which decreased over time at different rates. Whilst FC1 exhibited a sharp decrease within the first year, the decrease was more gradual for FC3 and FC7, where FC was normalised (<250 µg/g) approximately around two and five years post diagnosis, respectively. Furthermore, FC4 (n=103; ~10%), FC5 (n=67; ~6%) and FC6 (n=64; ~6%) capture relapsing and remitting patterns of gastrointestinal inflammation. Finally, FC8 (n=194; ~19%) represented individuals with consistently high FC values.

Associations with respect to FC cluster assignments

Associations with respect to age, sex and IBD type were first considered. Whilst the effect of age and sex was not substantial (Figure S8), this was not the case for IBD type (Figure S9). For example, IBDU and UC patients were less likely to be assigned to FC6 (1·56% IBDU and 28·1% UC vs 11·4% and 37·3% respectively elsewhere).
A similar analysis was performed after stratifying by IBD type (UC and CD only) and considering additional phenotyping (Figures S10 - S16). In most cases, effect sizes were not statistically significant (in some cases this was due to low counts and small cluster sizes). However, amongst CD patients, those without upper GI inflammation were more likely to be assigned to FC1 or FC2 (2·4% and 3·03% L4 vs 18·1% elsewhere). In UC, and to some extent CD, males were more likely to be assigned to FC8 (72% male versus 49·7% elsewhere). Finally, UC patients with ulcerative proctitis were found to be less likely to be assigned to FC3 (2·44% E1 versus 14·2% elsewhere).

FC cluster assignment and advanced therapy usage

In total, 270 (49·6%) CD and 108 (28·4%) UC subjects received an advanced therapy within the seven year observation period (Table 1). Overall AT rates and the distribution of time to first AT were not homogeneous across clusters (Figure 3). For example, whilst AT prescription rates in FC1 largely matched the overall FC cohort, prescriptions were generally earlier in this cluster, especially in CD patients. It was noteworthy that patients in FC8, with consistently high FC levels and later onset of AT,

had cumulative AT rates of 56·8% in CD and 37·9% in UC by the end of the seven year observation period.

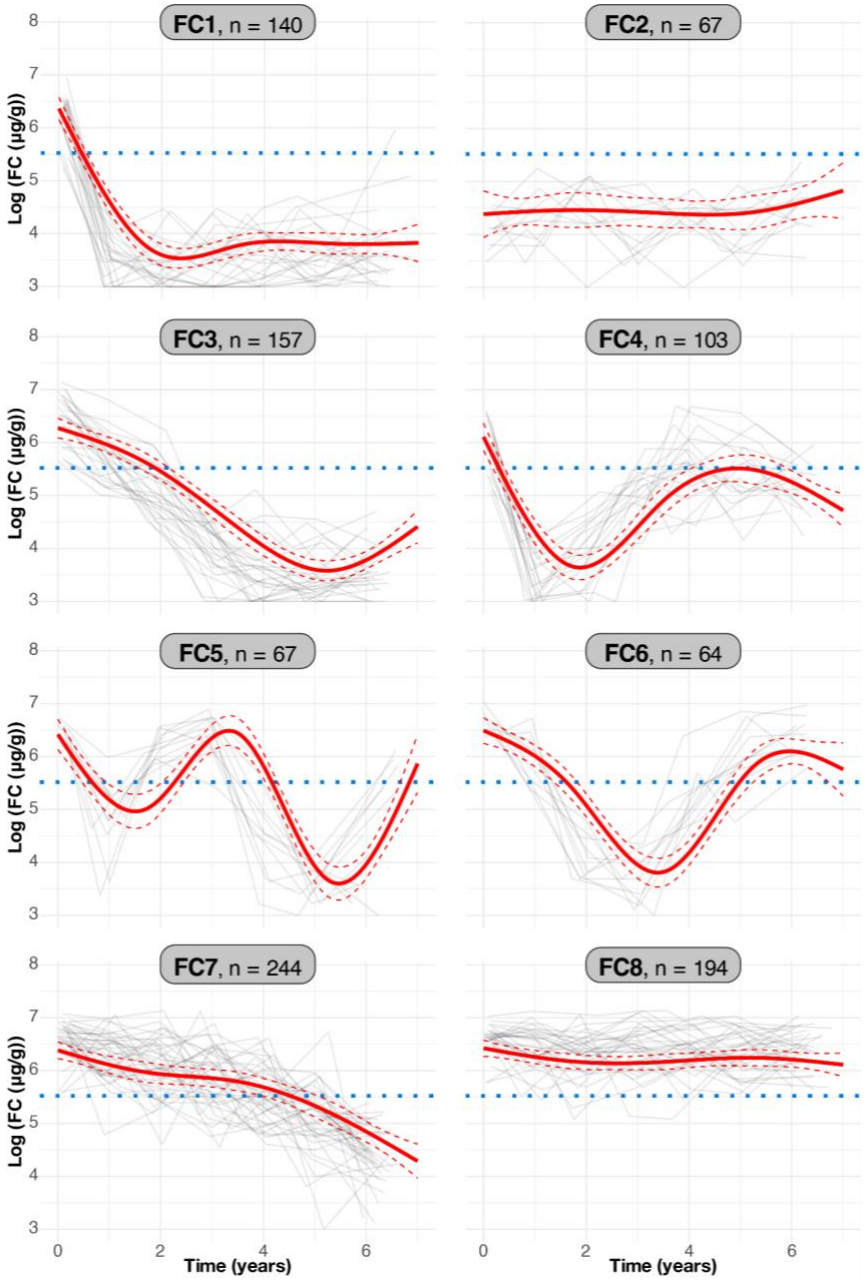

Figure 2. Cluster trajectories obtained from LCMM assuming eight clusters fitted to FC data (log-transformed). Red lines indicate predicted mean cluster profiles with 95% confidence intervals. The blue dotted lines indicate log(250µg/g). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from randomly selected groups of six subjects. Clusters are ordered from lowest (FC1) to highest (FC8) cumulative inflammatory burden. Cluster sizes are shown as panel titles.

Modelling of CRP trajectories

AIC and BIC both suggested the 8-cluster model was the most appropriate (Table S4). Visual inspection also supported this finding as the 9-cluster model did not identify new trajectories when compared to the 8-cluster model, producing two trajectories with consistently low CRP (Figure S8). In contrast, the 7-cluster model (Figure S9) lacks one of the clinically interesting trajectories, characterised by an initially elevated CRP which then decreases to slightly above biochemical remission after one year, when compared to the 8-cluster model.

Figure 4 presents exemplar cluster profiles for the 8-cluster model. Over a third of subjects (n=702; ~38%) were assigned to CRP1 which was defined by consistently low CRP. CRP2 (n=225; ~12%) was characterised by high CRP at diagnosis which rapidly decreased shortly thereafter remaining low. CRP3 (n=51; ~3%) and CRP4 (n=60; ~3%) are both small clusters with the former described by low CRP until the last year of follow-up and the latter presenting as low inflammation within the first year of diagnosis before increasing until the third year where the inflammation then decreases again. CRP5 (n=110; 6%) is characterised by elevated CRP at diagnosis which then decreases gradually over time. CRP6 (n=434; 24%) consists of trajectories which are elevated at diagnosis which then falls slightly for the first two years after diagnosis, remaining elevated across the remaining duration of follow-up. CRP8 (n=172; 9%) is consistently elevated and does not change over time.

Associations with CRP cluster assignments

Figure S19 and S20 visualise the distribution of age, sex and IBD type within each CRP cluster. Older patients were more likely to be assigned to a CRP cluster with higher cumulative inflammatory burden. IBD type was not evenly distributed among CRP clusters, with CRP1, CRP3, CRP4 and CRP7 enriched for UC patients (59·3%, 56·8%, 61·7% and 72·6% UC vs 32·3% elsewhere). Amongst CD patients, higher smoking rates were generally observed for CRP clusters with higher inflammatory burden (e.g. 26·9% in CRP1, 54.0% in CRP8; Figure S21), but there were not substantial differences when considering Montreal location and behaviour, or upper gastrointestinal inflammation (Figures S22 - S24). AT prescribing patterns in CRP clusters are shown in Figure S25.

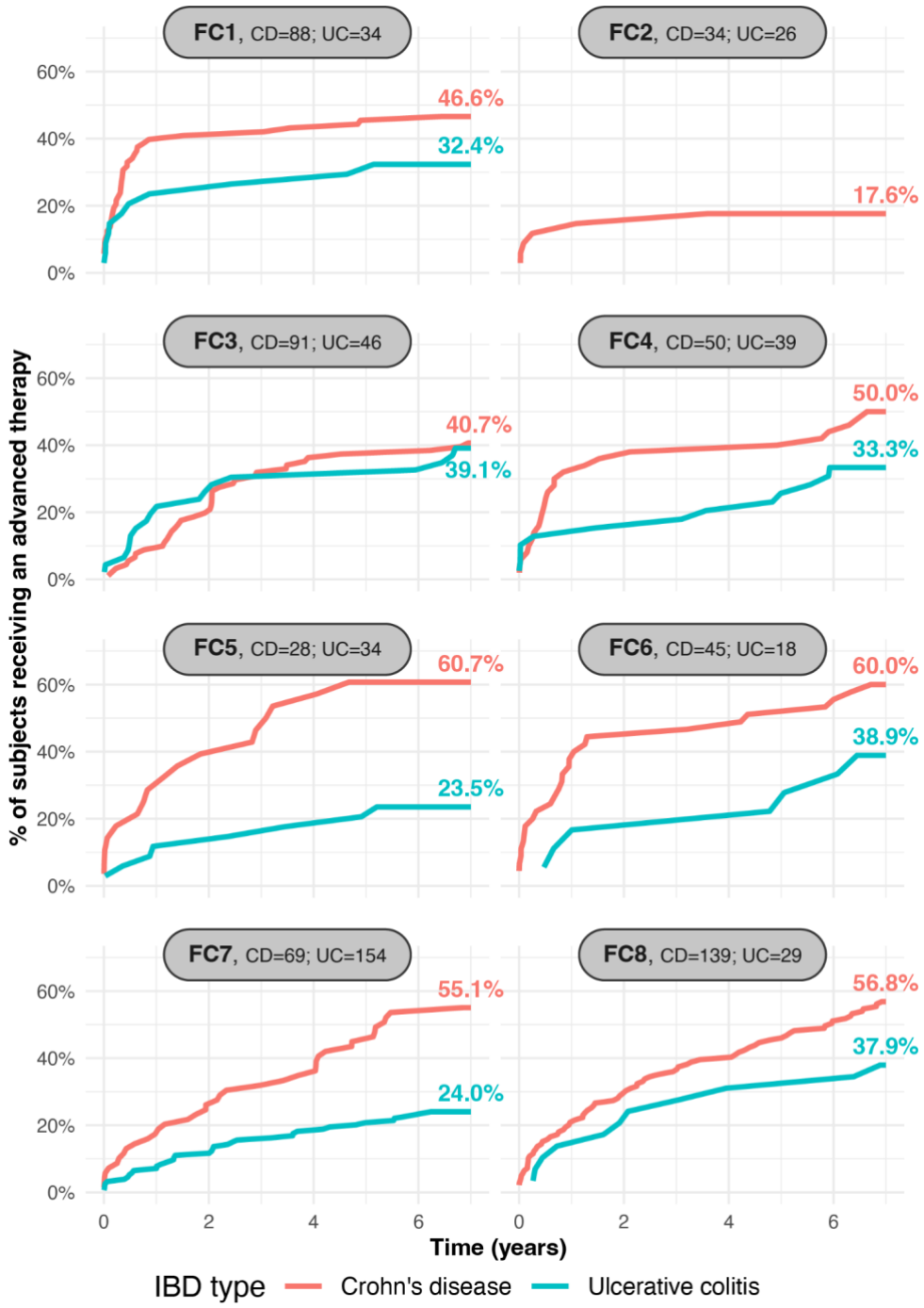

Figure 3. FC cluster-specific cumulative distribution for first-line advanced therapy prescribing for Crohn's disease (red) and ulcerative colitis (teal) subjects. Clusters are ordered from lowest (FC1) to highest (FC8) cumulative inflammatory burden. The number of CD and UC subjects present in each cluster is displayed as panel titles. Total advance therapy prescribing (as a percentage of the corresponding group) within seven years from diagnosis is shown next to each distribution curve. Curves which would describe fewer than five subjects are not shown.

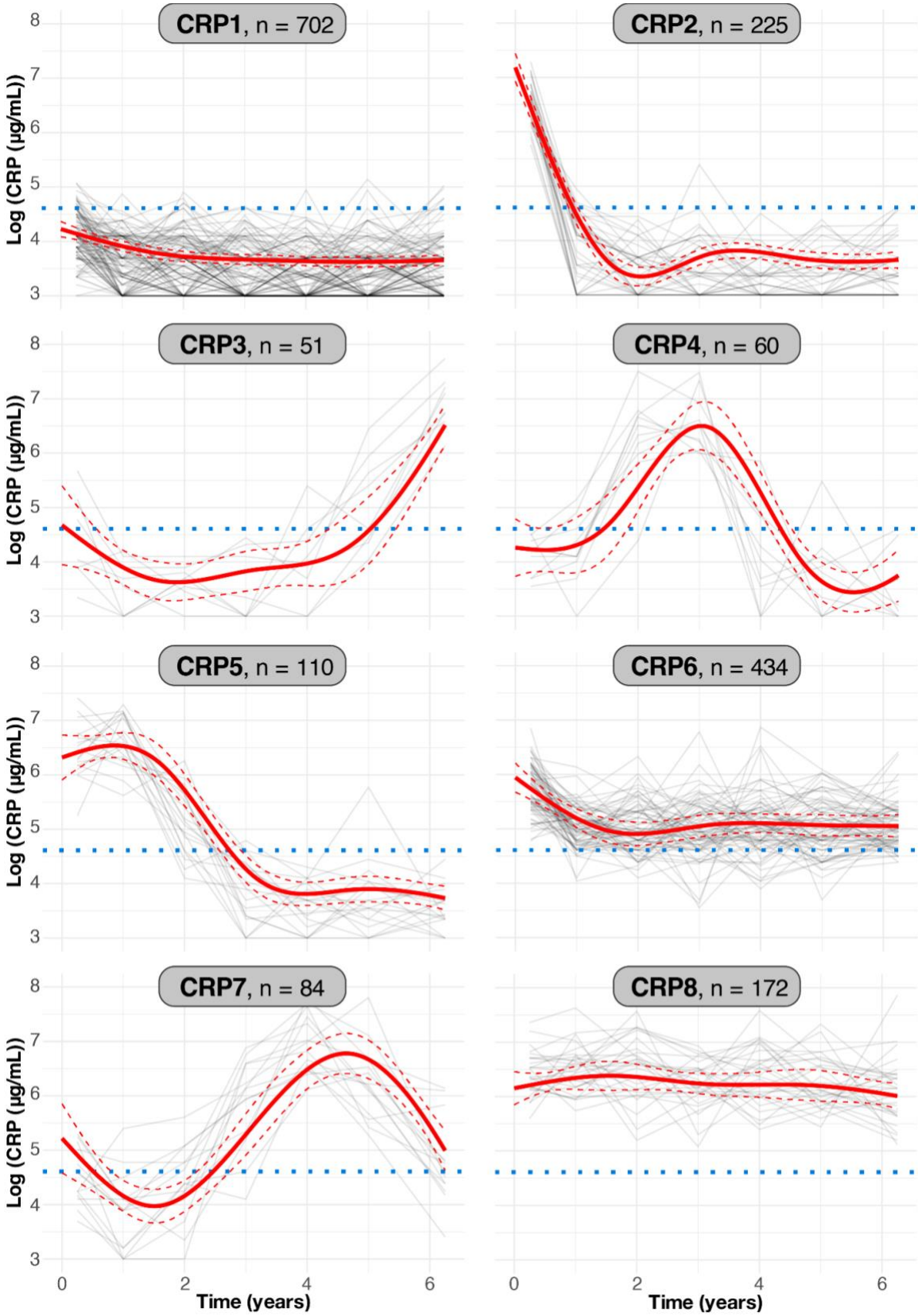

Figure 4. Cluster trajectories obtained from LCMM assuming eight clusters fitted to processed CRP data (log-transformed). Red lines indicate predicted mean cluster profiles with 95% confidence intervals. The blue dotted lines indicate log(5µg/mL). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from randomly selected

groups of six subjects. Clusters are ordered from lowest (CRP1) to highest (CRP8) cumulative inflammatory burden. Cluster sizes are shown as panel titles.

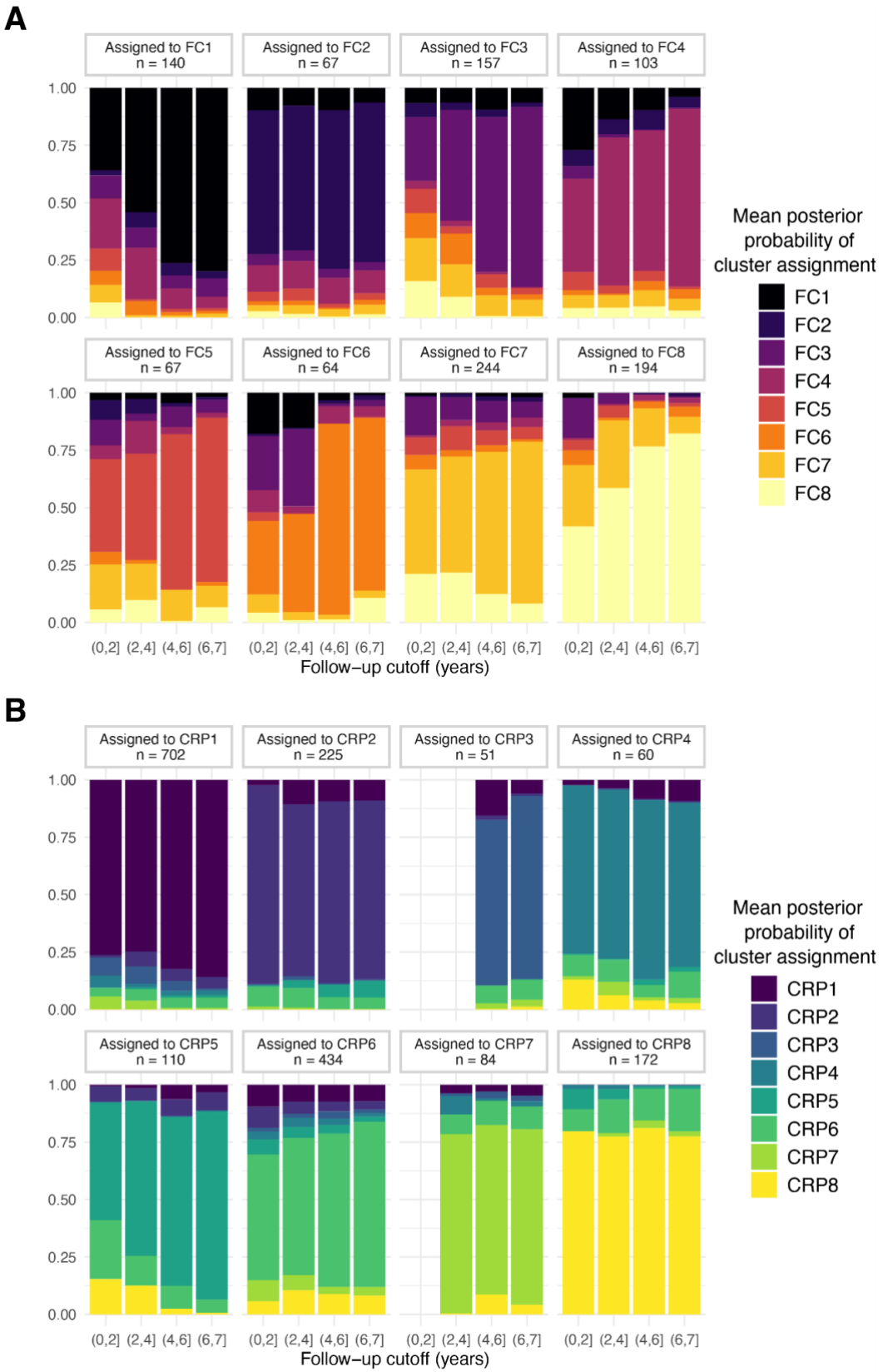

Figure 5. Exploration of cluster assignment uncertainty for A) faecal calprotectin (FC) and B) CRP clusters. LCMM assigns individuals to the cluster with the highest estimated probability. For

individuals assigned to a given cluster, bars show the average probability of cluster assignment to each possible cluster. Results are stratified according to follow-up length, defined as the time difference between diagnosis and the last available biomarker measurement (FC or CRP for A) and B), respectively). Clusters are ordered from lowest (FC1 and CRP1) to highest (FC8 and CRP8) cumulative inflammatory burden, with adjacent clusters coloured sequentially in the plots for FC (black to yellow) and CRP (blue to yellow).

Uncertainty in cluster assignments

In the FC analysis, with the exception of FC2, cluster assignments were on average more uncertain for subjects with a short follow-up (Figure 5 (A)). This is particularly the case for FC clusters that share similar earlier trends. For example, individuals assigned to FC1 (rapid FC normalisation) had a low average probability of being assigned to FC8 (consistently high FC) and vice-versa, even for those with a short follow-up. This is not the case when comparing FC3 and FC6, both of which capture similar FC trajectories within the first two years. Indeed, those assigned to FC6 with less than two years of follow-up also have, on average, a high probability of being assigned to FC3. On average, cluster assignments were less uncertain in the CRP analysis, even for individuals with a short follow-up (Figure 5 (B)). This is expected as CRP clusters are associated with more distinct early trajectories.

Comparison of FC and CRP clustering

Overall, all FC clusters were well represented amongst the 808 subjects included in both analyses (*overlap cohort*), but CRP8 was underrepresented (~27% of CRP8 was in the overlap cohort vs ~46% elsewhere; Figure S26). When comparing FC and CRP clustering in the overlap cohort, there was little agreement (Figure 6 (A)). Whilst most subjects in FC1 (71.6%) were also in CRP1 or CRP2, this relationship was not mirrored as 81.2% of subjects in the latter two clusters were assigned to substantially different FC clusters. However, CRP8 did overlap with elevated FC as the vast majority of subjects in this cluster (80.8%) were assigned to either FC7 or FC8. For the 451 CD subjects in the overlap cohort, largely similar patterns were observed (Figure 6 (B)). For the 276 UC subjects, a substantial proportion of FC7 (45.1%) were assigned to CRP1 (Figure 6 (C)), perhaps reflecting the characteristic where UC patients are less likely to mount an abnormal CRP response than CD patients[21].

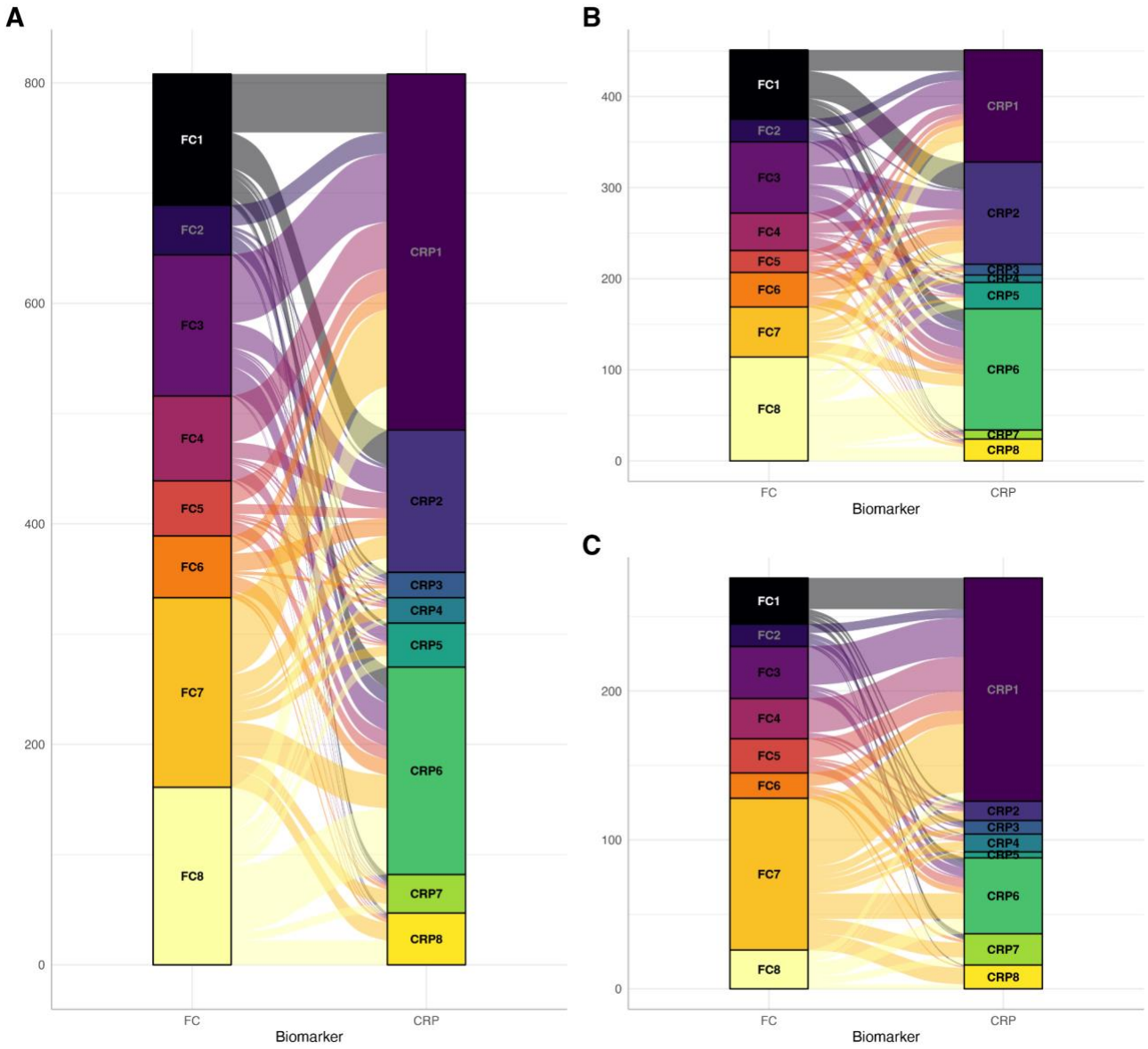

Figure 6. Comparison between faecal calprotectin (FC) and processed CRP for models with chosen specification (three NCS) assuming eight clusters. Results are reported based on the overlap cohort, consisting of 808 subjects included in both the FC and CRP analysis. (A) all subjects; (B) Crohn's disease; and (C) ulcerative colitis. Each segment denotes the size of the cluster whilst the alluvial segments connecting the nodes visualises the number of subjects shared between clusters.

Discussion

We have characterised IBD behaviour using long-term longitudinal trends of objective inflammatory markers routinely collected for clinical care. Our model has uncovered, in a large IBD cohort, eight clusters with distinct inflammatory profiles based on FC and CRP, respectively. Our data highlights the heterogeneity of the disease course,

and present a novel approach for understanding real-world patterns of inflammatory activity in IBD patients. For the first time, we are capturing the dynamic nature of the disease in a more biologically nuanced way than traditional behaviour endpoints such as treatment escalation, surgery and "Montreal progression". Moreover, this represents a marked shift from the traditional symptom-based behaviour profiles exemplified by the IBSEN cohorts over a decade ago.[9,10]

This study builds on our earlier proof-of-concept work, where we first demonstrated the feasibility of using long-term individualised profiles of FC to cluster IBD patients with CD.[15] In our previous analysis we identified four distinct FC clusters: one cluster with persistently high FC (non-remitters) and three clusters with different downward longitudinal trends. Here, we expand upon this by considering a substantially larger IBD cohort (FC cohort, n=1036; CRP cohort, n=1838), with longer follow-up, not excluding patients based on indicators of disease severity, and also including patients with UC or IBDU. In addition, we modelled CRP. This has generated results with greater representativeness across IBD phenotypes, uncovering more granular structure with eight distinct FC clusters rather than four. Whilst the modelling suggested further partitioning of the FC data was possible (Figure S5, Figure S7 (A)), we selected the eight-cluster model as a parsimonious choice that captured the key inflammatory patterns (Figure 2). Notably, the CRP data also partitioned into eight distinct clusters. Together, the clusters broadly fall into four patterns of inflammatory behaviour mirroring those recognised by gastroenterologists managing IBD patients (i) rapid remitters (FC1 and CRP2), (ii) delayed remitters (FC3, FC7 and CRP5), (iii) relapsing-remitters (FC4-6 and CRP4 and CRP7), and (iv) non-remitters (FC8 and CRP6 and CRP8). These classifications provide new insights into the inflammatory course of IBD patients.

We observed broadly poor agreement between FC and CRP clusters, although there was overlap amongst those at the varying ends of the inflammatory spectrum with FC1 and FC2 correlating with CRP1 and CRP2 (rapid remitters with low cumulative inflammatory burden), and FC7 and FC8 correlating with CRP6-8 (non-remitters with high cumulative inflammatory burdens). Although, the lack of overlap between the two is not surprising. CRP is a marker of systemic inflammation, whilst FC is more specific for detecting inflammation at a mucosal level. Hence, why both biomarkers are

complimentary when monitoring patients with IBD. Studies have also shown that a proportion of IBD patients will have lower CRP values at diagnosis,[22] as is seen in CRP1 which is significantly enriched for UC cases.

We observed several important observations regarding cluster assignment. All IBD subtypes featured in each cluster, however cluster assignment was unevenly distributed across CD, UC and IBDU (Figures S9 and S20). In CD, although patients with L4 disease were less represented in FC1 and FC2 there was no association with ileal versus colonic disease location. Males were slightly overrepresented in the FC8 cluster, which is characterised by persistently elevated values. Interestingly, in the recent SEXEII study, they observed male patients with UC were more likely to have extensive colonic involvement and abdominal surgery, which may account for this finding.[23] CRP cluster membership was also associated with smoking and older age, with both more likely to have higher inflammatory burdens.

Multiple lines of evidence, including the recent PROFILE study,[24] have clearly demonstrated improved disease control with and outcomes in Crohn's patients receiving early advanced therapy. As such, we anticipated that the biggest driver of inflammatory patterns over time would be advanced therapies and that this effect would be more pronounced in CD versus UC especially given the era of our cohort. Indeed, rates of advanced therapy use were not homogenous across clusters or IBD subtypes. Overall, in the FC cohort, advanced therapies were used in 49.6% of CD patients of which 47.7% started AT in the first year. In FC1, similar rates of advanced therapies were used for Crohn's patients, however they were used earlier, suggesting their positive benefit in rapidly inducing and maintaining remission. Rates of advanced therapy prescriptions for patients with CD in FC8 (persistently high levels of inflammatory behaviour) were similar, but started later in the disease course, which may have negatively affected the ability to bring about remission of disease. This cluster may also represent a more refractory group of patients, with a higher risk phenotype. Whilst this may provide additional support for the use of early advanced therapy, a causal interpretation of these effects is not possible in this study, which is based on observational data. Additional work with alternative cohorts where treatment assignment is randomised is planned.

Our study also highlights the importance of using a probabilistic approach to account for uncertainty in cluster assignments. Indeed, we observed higher uncertainty for subjects with a shorter longitudinal follow-up, particularly for FC clusters (Figure 5). In such cases, we cannot confidently assign individuals to a specific cluster. Instead, cluster assignment probabilities are sometimes evenly split across multiple clusters, mostly between those with similar earlier behaviour and inflammatory burdens. This effect is less prominent in CRP cluster assignments, partly due to the more distinct early trajectories observed across CRP clusters. As such, we anticipate that a multivariate approach which simultaneously considers other biomarkers of disease activity, such as haemoglobin, albumin, and platelet count, may further increase the robustness of cluster assignments. Such analysis may also consider pre-diagnostic biomarker measurements, following the recent observations in Danish registry data,[25] as well as metabolomics, genetics or microbiome data to inform cluster assignments.

Critically, this study is limited to data generated in a single health board in Scotland. As such, independent validation of the identified longitudinal clusters will be necessary before clinical implementation and to better understand the associations between cluster assignment and disease phenotyping. Furthermore, there are limitations associated with the use of observational data. Beyond advanced therapy use, the frequency and amount of biomarker measurements is likely to be higher for those with a more severe disease, and very mild cases may be excluded from our cohort. Moreover, FC and CRP clusters cannot be directly used in a prognostic way. Indeed, future work is needed to assess associations between clusters and IBD related complications (such as steroid use, hospitalisations and surgery) but also non-conventional complications related to a high cumulative inflammatory burden, such as major cardiovascular adverse events, neuropsychiatric illness, and malignancy. Such analyses will ultimately support the development of a low cost clinical support tool to help deliver precision medicine to patients with IBD.

Classifying patients by their inflammatory behaviour may better inform therapy decisions, including timing of advanced therapy, as well as monitoring and follow-up requirements. This is a paradigm shift in thinking about disease behaviour compared with the previous symptom based profiles first reported by the IBSEN cohorts. Moreover, this approach - rooted in data that is widely available in clinical settings[26] and probabilistic modelling- paves the way for predictive analytics integrated into a

clinical support tool for population wide risk stratification and individual patient level prognostication and treatment.

Data sharing statement

As the data collected for this study has been derived from unconsented patient data, it is not possible to share subject-level data with external entities. Detailed summary level data is available online at https://vallejosgroup.github.io/Lothian-IBDR. The code used to conduct the analysis is also publicly available (https://github.com/VallejosGroup/Lothian-IBDR).

Author contributions

CAV, CWL, and NCC were involved in conceptualising the study. CRB, SO, ATE, GRJ, and NP curated the data. NCC and CAV implemented computer code, tested existing code components, and conducted the formal analysis and visualisation. The original draft was written by NCC, NP, ATE, CMR, CWL and CAV with all authors involved in reviewing and editing the manuscript. CWL and CAV provided supervisory support.

Competing interests

NP has served as a speaker for Janssen, Takeda and Pfizer. BG has acted as consultant to Galapagos and Abbvie and as speaker for Abbvie, Jansen, Takeda, Pfizer and Galapagos. G-RJ has served as a speaker for Takeda, Janssen, Abbvie, Fresnius and Ferring. CWL has acted as a speaker and/or consultant to AbbVie, Janssen, Takeda, Pfizer, Galapagos, GSK, Gilead, Vifor Pharma, Ferring, Dr Falk, BMS, Boehringer Ingelheim, Eli Lilly, Merck, Novartis, Sandoz, Celltrion, Cellgene, Amgen, Samsung Bioepis, Fresenius Kabi, Tillotts, Kuma Health, Trellus Health and Iterative Health. None of the other authors report any conflicts of interest.

Funding

CWL is funded by a UKRI (UK Research and Innovation) Future Leaders Fellowship ‘Predicting outcomes in IBD’ (MR/S034919/1). G-RJ is funded by a Wellcome Trust Clinical Research Career Development Fellowship. NC-C was partially supported by the Medical Research Council and The University of Edinburgh via a Precision Medicine PhD studentship (MR/N013166/1).

References

1. Jones GR, Lyons M, Plevris N, et al. IBD prevalence in Lothian, Scotland, derived by capture–recapture methodology. *Gut*. 2019;68(11):1953-1960. doi:10.1136/gutjnl-2019-318936

2. Hamilton B, Green H, Heerasing N, et al. Incidence and prevalence of inflammatory bowel disease in Devon, UK. *Frontline Gastroenterol*. 2021;12(6):461-470. doi:10.1136/flgastro-2019-101369

3. Cushing K, Higgins PDR. Management of Crohn disease: A review. *JAMA*. 2021;325(1):69. doi:10.1001/jama.2020.18936

4. Gros B, Kaplan GG. Ulcerative colitis in adults: A review. *JAMA*. 2023;330(10):951. doi:10.1001/jama.2023.15389

5. Zhou RW, Harpaz N, Itzkowitz SH, Parsons RE. Molecular mechanisms in colitis-associated colorectal cancer. *Oncogenesis*. 2023;12(1):48. doi:10.1038/s41389-023-00492-0

6. Kennedy NA, Jones GR, Plevris N, Patenden R, Arnott ID, Lees CW. Association between level of fecal calprotectin and progression of Crohn's disease. *Clin Gastroenterol Hepatol*. 2019;17(11):2269-2276.e4. doi:10.1016/j.cgh.2019.02.017

7. Plevris N, Fulforth J, Lyons M, et al. Normalization of fecal calprotectin within 12 months of diagnosis is associated with reduced risk of disease progression in patients with Crohn's disease. *Clin Gastroenterol Hepatol*. 2021;19(9):1835-1844.e6. doi:10.1016/j.cgh.2020.08.022

8. Verstockt B, Noor NM, Marigorta UM, et al. Results of the Seventh Scientific Workshop of ECCO: Precision medicine in IBD—Disease outcome and response to therapy. *J Crohns Colitis*. 2021;15(9):1431-1442. doi:10.1093/ecco-jcc/jjab050

9. Solberg IC, Vatn MH, Høie O, et al. Clinical course in Crohn's disease: Results of a Norwegian population-based ten-year follow-up study. *Clin Gastroenterol Hepatol*. 2007;5(12):1430-1438. doi:10.1016/j.cgh.2007.09.002

10. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: Results from a population-based inception cohort (IBSEN Study). *Scand J Gastroenterol*. 2009;44(4):431-440. doi:10.1080/00365520802600961

11. Peyrin-Biroulet L, Reinisch W, Colombel JF, et al. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. *Gut*. 2014;63(1):88-95. doi:10.1136/gutjnl-2013-304984

12. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of

a working party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol Hepatol*. 2005;19:5A-36A. doi:10.1155/2005/269076

13. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol*. 1989;24(sup170):2-6. doi:10.3109/00365528909091339

14. Sorlie P, Wei GS. Population-based cohort studies: Still relevant? *J Am Coll Cardiol*. 2011;58(19):2010-2013. doi:10.1016/j.jacc.2011.08.020

15. Constantine-Cooke N, Monterrubio-Gómez K, Plevris N, et al. Longitudinal fecal calprotectin profiles characterize disease course heterogeneity in Crohn's disease. *Clin Gastroenterol Hepatol*. 2023;21(11):2918-2927.e6. doi:10.1016/j.cgh.2023.03.026

16. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *J Stat Softw*. 2017;78(2):1-56. doi:10.18637/jss.v078.i02

17. Stoica P, Selen Y. Model-order selection: A review of information criterion rules. *IEEE Signal Process Mag*. 2004;21(4):36-47. doi:10.1109/MSP.2004.1311138

18. Proust-Lima C, Philipps V, Diakite A, Liquet B. lcmm: extended mixed models using latent classes and latent processes. Published online 2021. https://cran.r-project.org/package=lcmm

19. Brunson JC. ggalluvial: layered grammar for alluvial plots. *J Open Source Softw*. 2020;5(49):2017. doi:10.21105/joss.02017

20. Constantine-Cooke N. datefixR: fix really messy dates in R. Published online 2024. doi:10.5281/zenodo.5655311

21. Clough J, Colwill M, Poullis A, Pollok R, Patel K, Honap S. Biomarkers in inflammatory bowel disease: A practical guide. *Ther Adv Gastroenterol*. 2024;17:17562848241251600. doi:10.1177/17562848241251600

22. Henriksen M, Jahnsen J, Lygren I, et al. C-reactive protein: A predictive factor and marker of inflammation in inflammatory bowel disease. Results from a prospective population-based study. *Gut*. 2008;57(11):1518-1523. doi:10.1136/gut.2007.146357

23. Gargallo-Puyuelo CJ, Ricart E, Iglesias E, et al. Sex-related differences in the phenotype and course of inflammatory bowel disease: SEXEII study of ENEIDA. *Clin Gastroenterol Hepatol*. 2024;22(11):2280-2290. doi:10.1016/j.cgh.2024.05.013

24. Noor NM, Lee JC, Bond S, et al. A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): A multicentre, open-label randomised controlled trial. *Lancet Gastroenterol Hepatol*. 2024;9(5):415-427. doi:10.1016/S2468-1253(24)00034-7

599 25. Vestergaard MV, Allin KH, Poulsen GJ, Lee JC, Jess T. Characterizing the pre-
600 clinical phase of inflammatory bowel disease. *Cell Rep Med*. 2023;4(11):101263.
601 doi:10.1016/j.xcrm.2023.101263

602 26. Markowetz F. All models are wrong and yours are useless: Making clinical
603 prediction models impactful for patients. *Npj Precis Oncol*. 2024;8(1):54.
604 doi:10.1038/s41698-024-00553-6

605

Supplemental display items

Table of contents

Supplemental figures

Supplemental tables

Supplemental figures

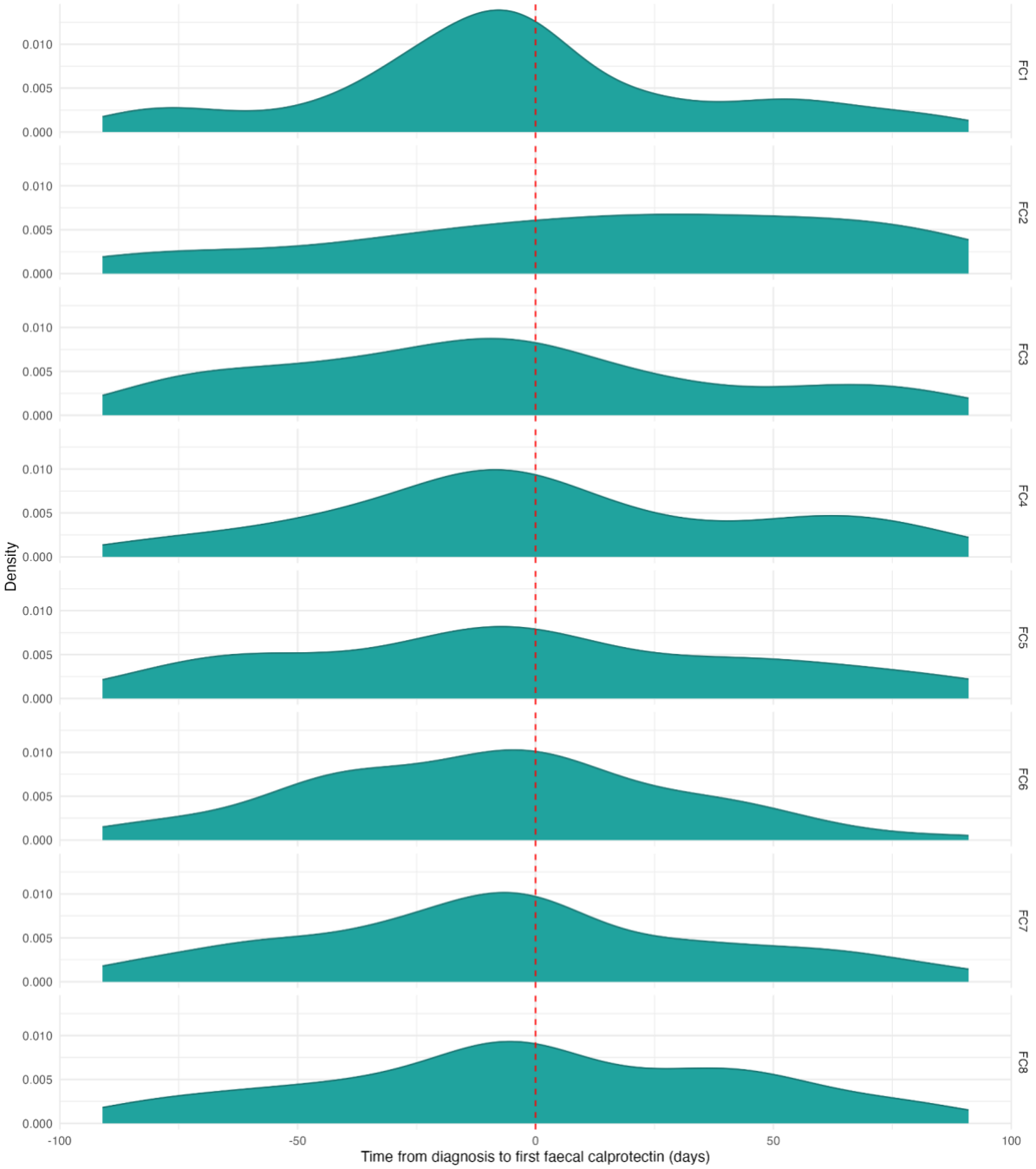

Figure S1. Distribution of observation times for diagnostic faecal calprotectin (FC) relative to reported date of diagnosis for the study cohort. Stratified by FC cluster assignment. Diagnostic FC was defined as the first FC test within ±90 days of diagnosis.

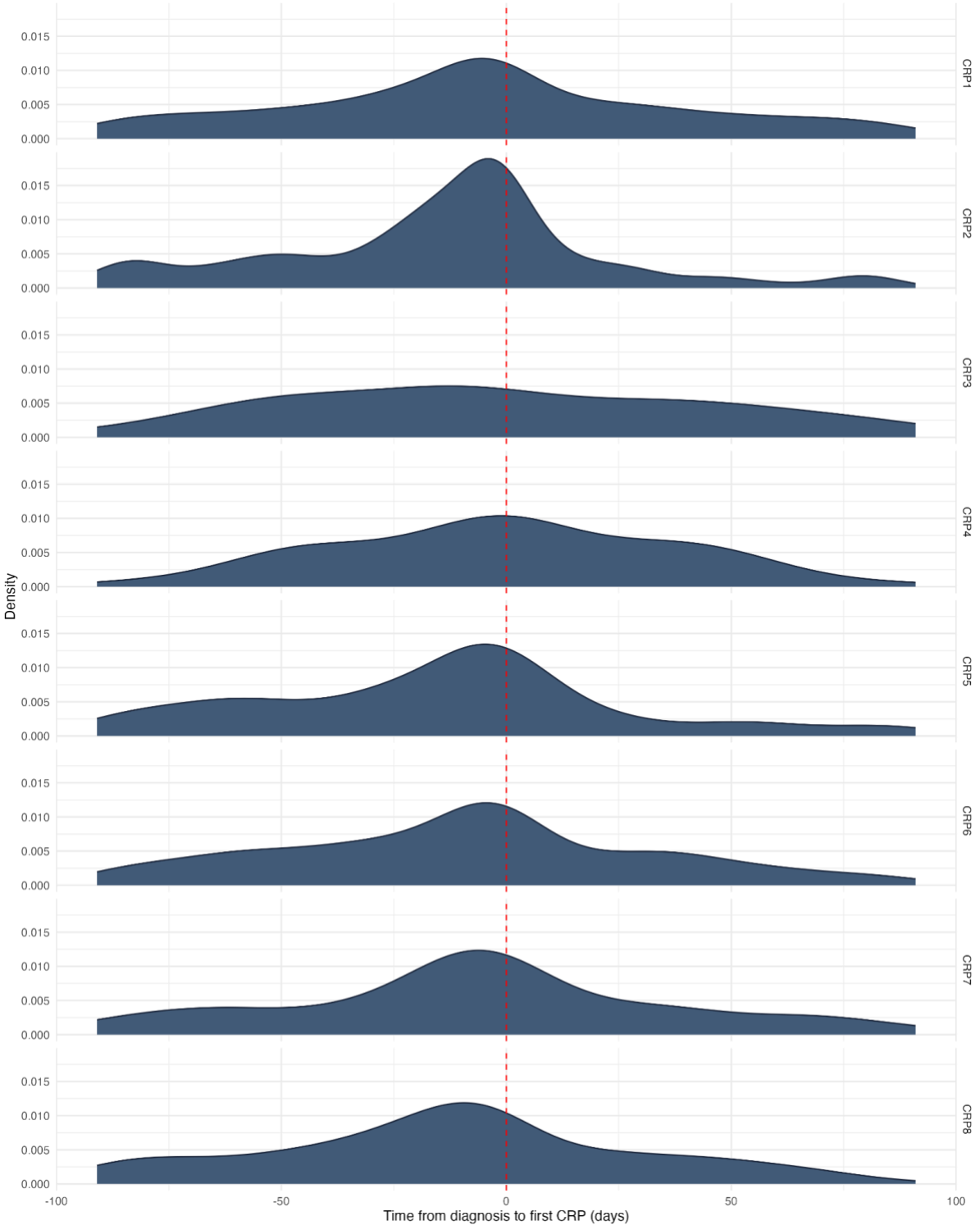

Figure S2. Distribution of observation times for diagnostic C-reactive protein (CRP) relative to reported date of diagnosis for the study cohort. Stratified by CRP cluster assignment. Diagnostic CRP was defined as the first CRP test within ±90 days of diagnosis.

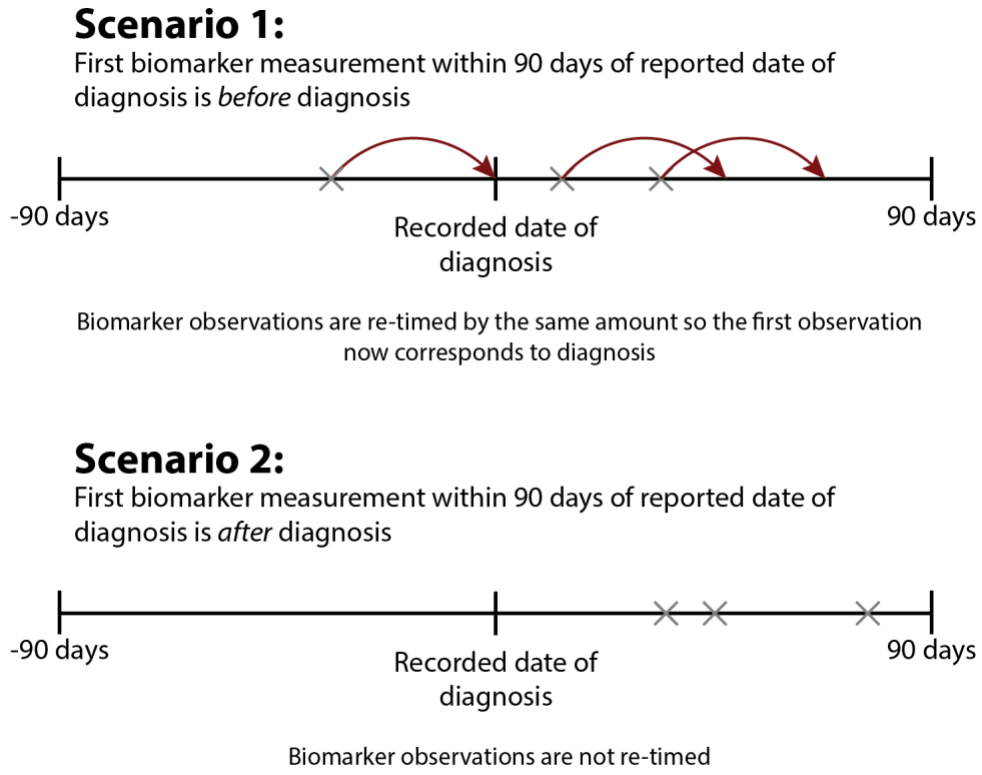

Figure S3. Illustration of how biomarker (faecal calprotectin or CRP) observation times were adjusted depending on if the diagnostic observation was before or after the date of diagnosis recorded in electronic health records.

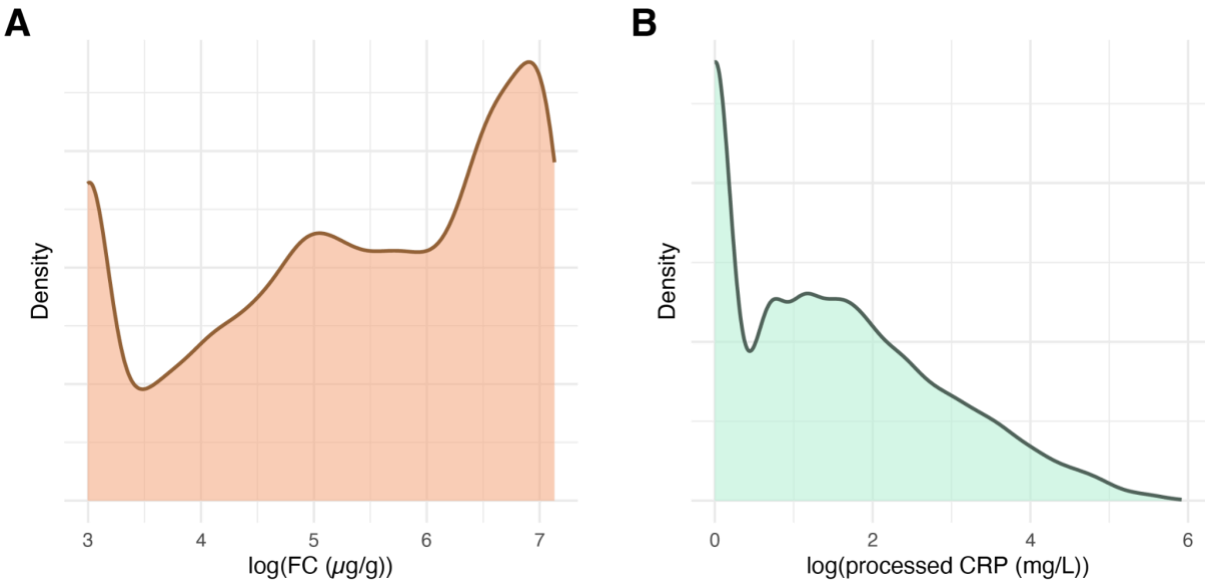

Figure S4. Distribution of diagnostic biomarker measurements across the study cohort. (A) faecal calprotectin after applying a logarithmic transformation; (B) pre-processed (grouped into time intervals with the median used for multiple measurements) CRP after logarithmic transformation.

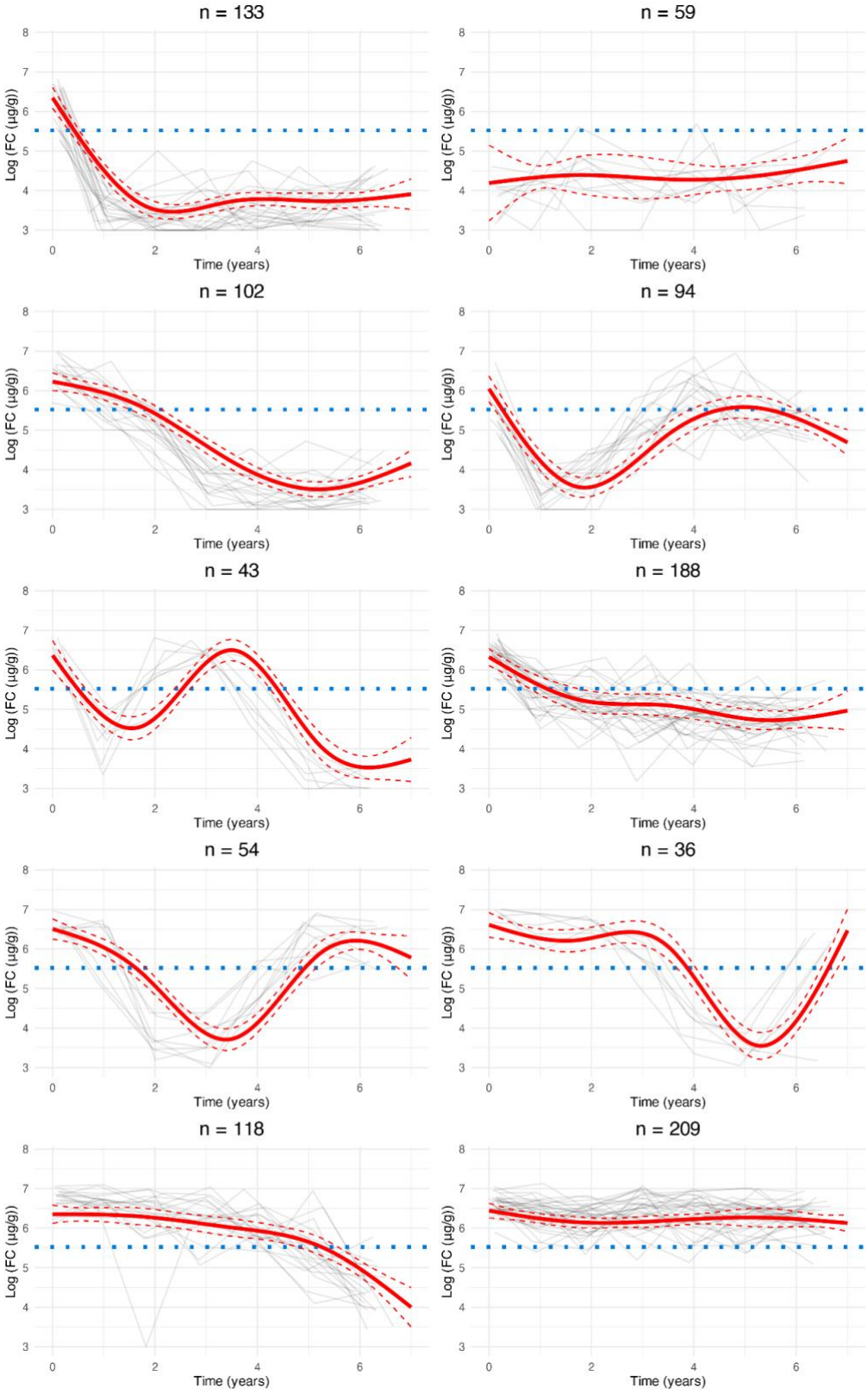

Figure S5. Cluster trajectories obtained from LCMM assuming ten clusters fitted to faecal calprotectin data. Red lines indicate predicted mean cluster profiles with 95% confidence intervals. Dotted horizontal lines indicate log(250µg/g). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from groups of six subjects.

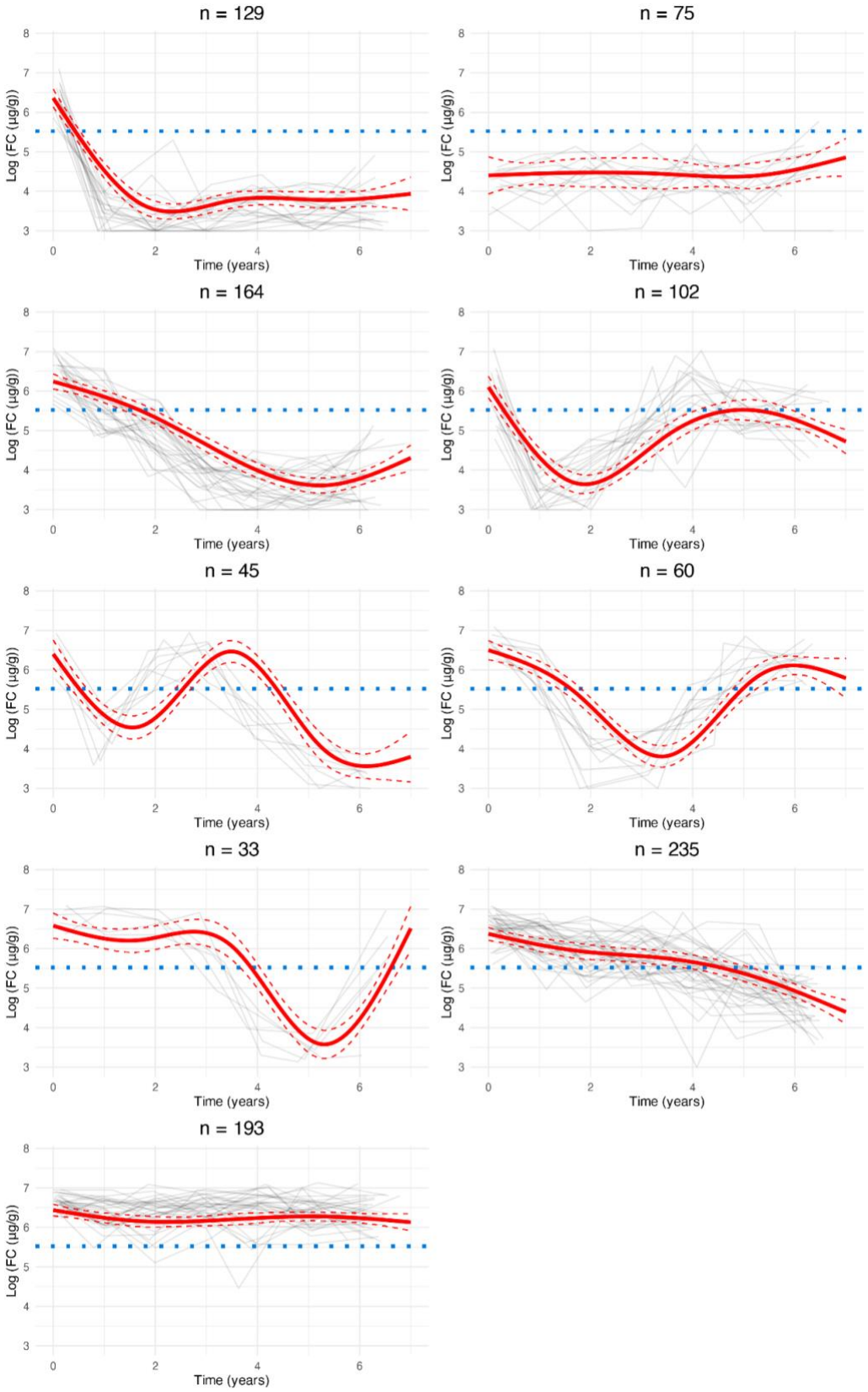

Figure S6. Cluster trajectories obtained from LCMM assuming nine clusters fitted to faecal calprotectin data. Red lines indicate predicted mean cluster profiles with 95% confidence intervals. Dotted horizontal lines indicate log(250µg/g). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from groups of six subjects.

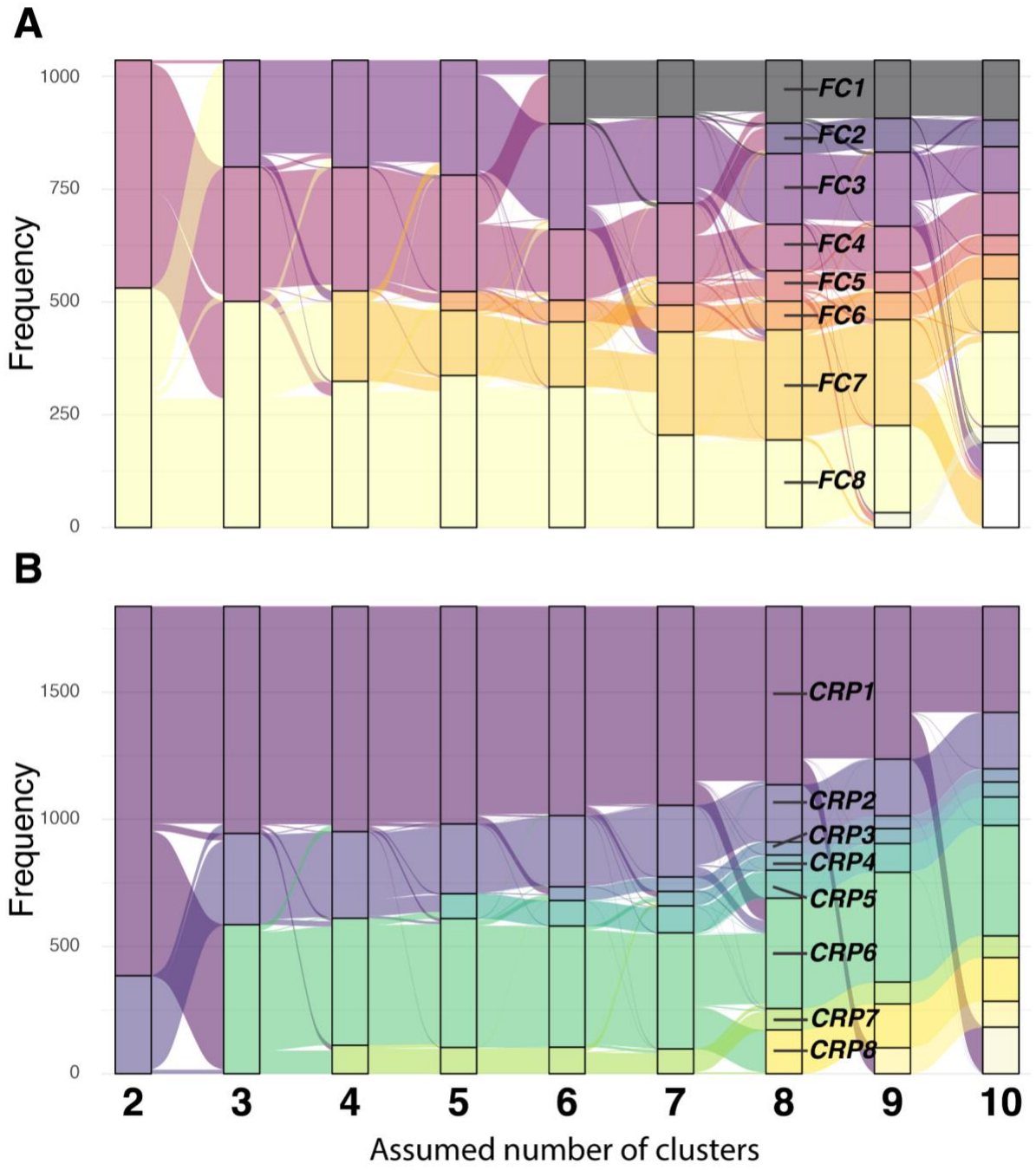

Figure S7. Alluvial plot demonstrating how cluster assignment changes as the number of assumed clusters increases for the chosen models for (A) faecal calprotectin and (B) C-reactive protein. The clusters found by the 8-cluster models are labelled.

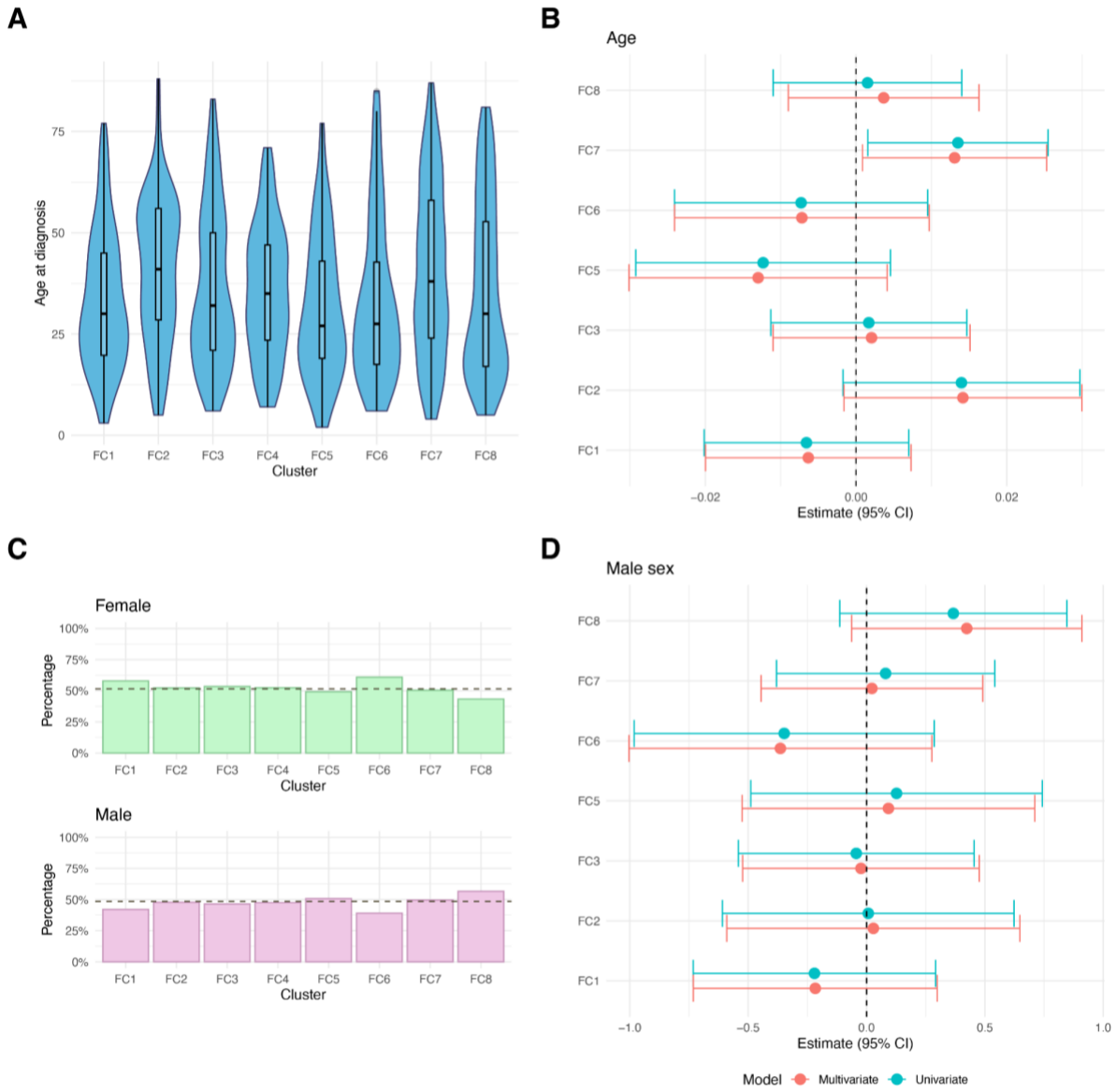

Figure S8. (A) For each FC cluster, violin plots show the distribution of age at diagnosis across subjects, highlighting median and interquartile ranges. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for age in a multinomial logistic regression model that uses FC cluster assignment as outcome. (C) For each FC cluster, panels show the proportion of individuals with female and male sex. The dashed horizontal line represents overall proportions across the entire FC cohort. (D) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for male sex versus females (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. In (B) and (C), effect sizes are with respect to the reference cluster (in this case FC4). In both cases, the multivariate model includes age, sex and IBD type as covariates. The dashed vertical lines are used as a reference to indicate no effect.

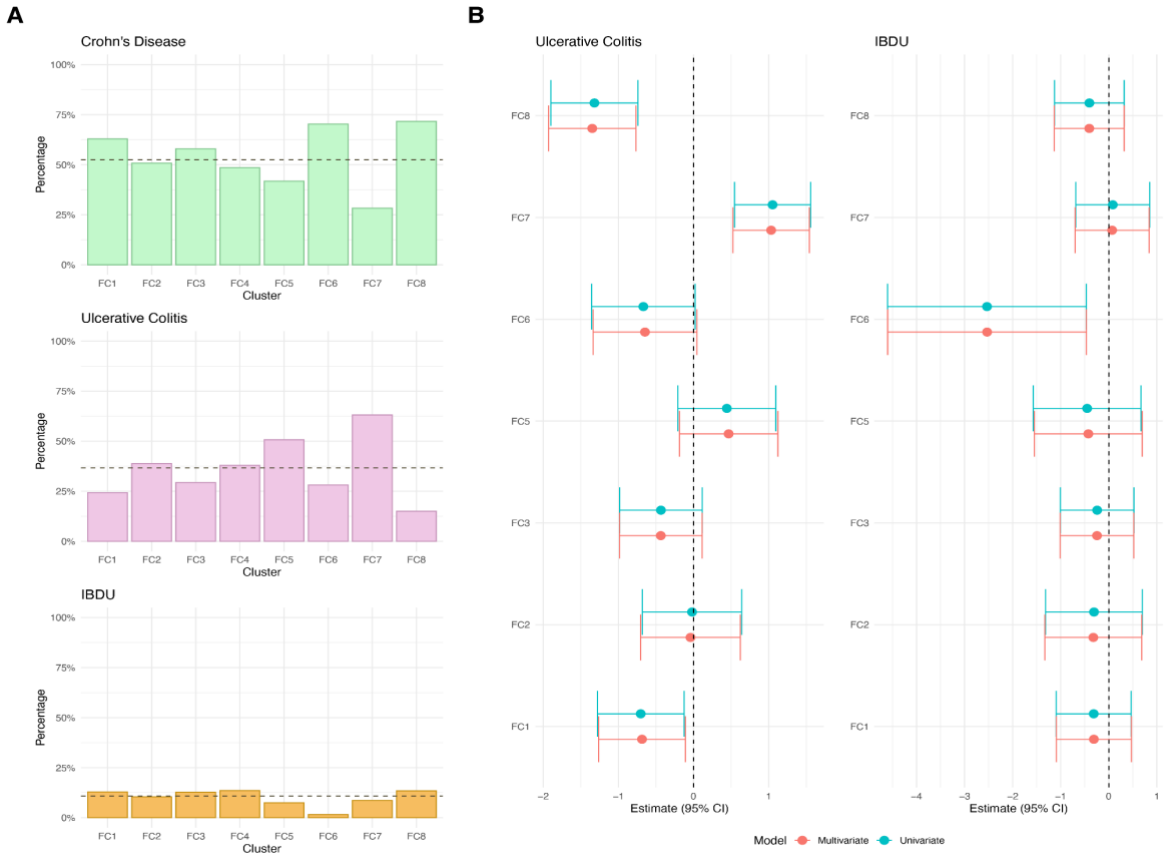

Figure S9. (A) For each FC cluster, panels show the proportion of individuals with Crohn's disease, ulcerative colitis and IBDU respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for IBD type: ulcerative colitis and IBDU versus Crohn's disease (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). In both cases, the multivariate model includes age, sex and IBD type as covariates. The dashed vertical lines are used as a reference to indicate no effect.

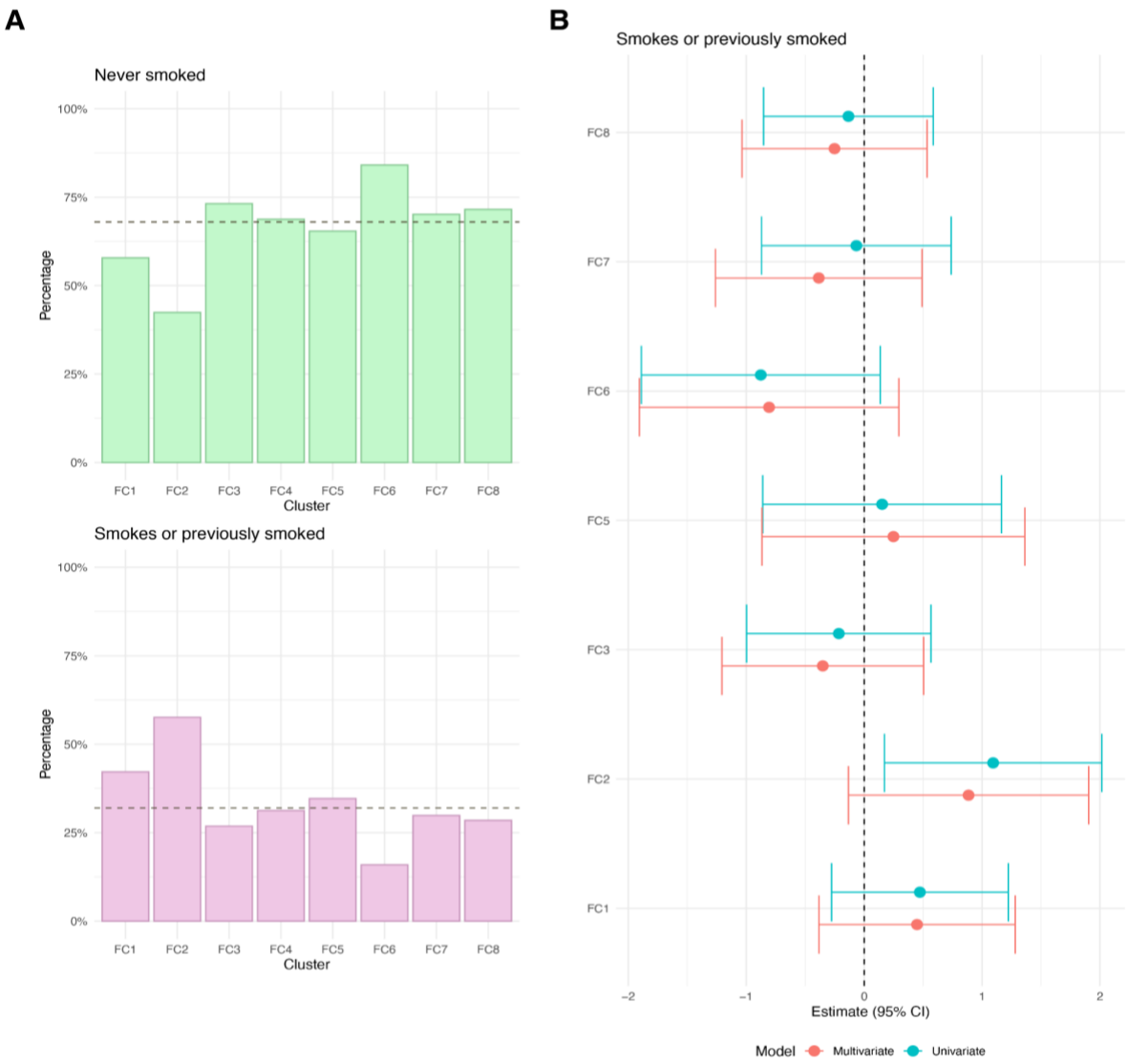

Figure S10. Crohn's disease patients only. (A) For each FC cluster, panels show the proportion of individuals with smoking behaviour recorded as "no" (no and never) and "yes" (current or previously smoked) at diagnosis respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for smoking: yes versus no (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), perianal disease (yes, no) and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

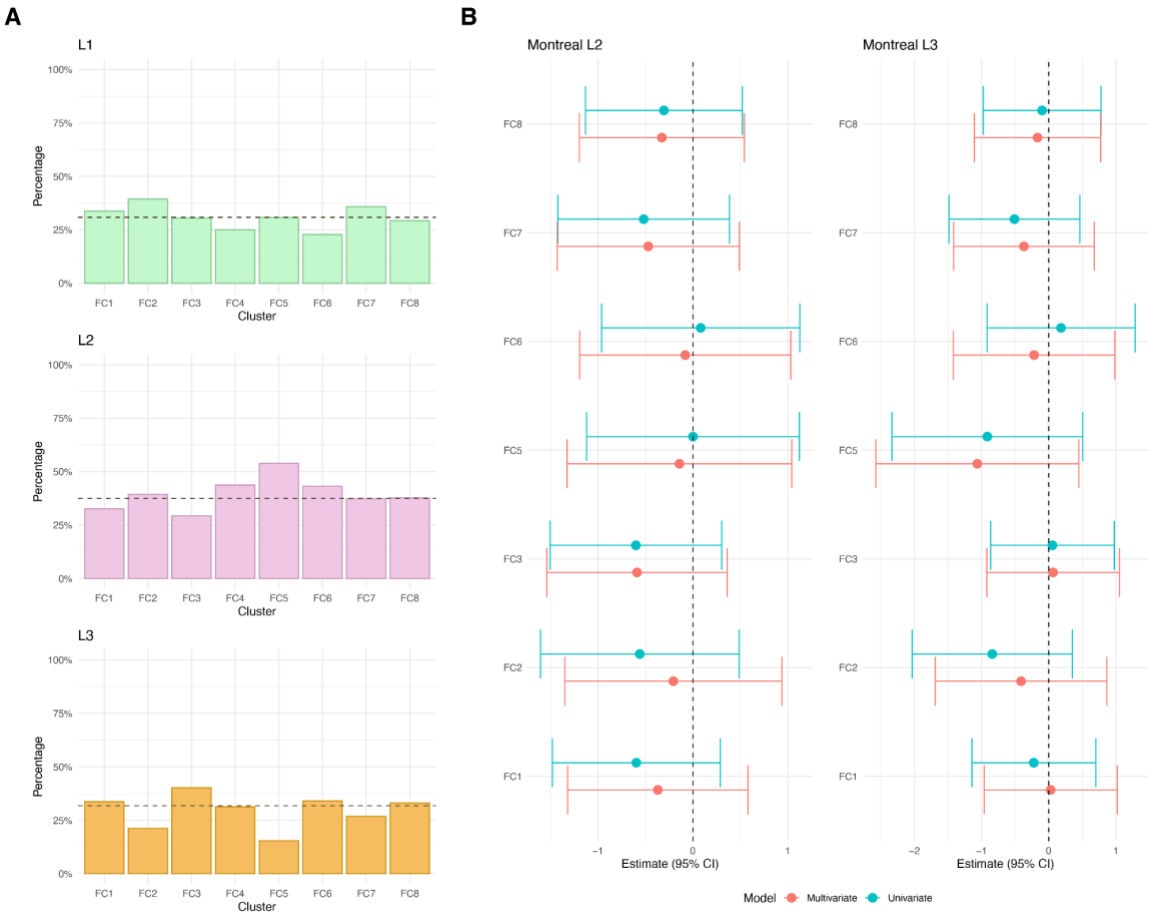

Figure S11. Crohn's disease patients only. (A) For each FC cluster, panels show the proportion of individuals with Montreal Location recorded as L1, L2 and L3 respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for Montreal Location: L2 and L3 versus L1 (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), perianal disease (yes, no) and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

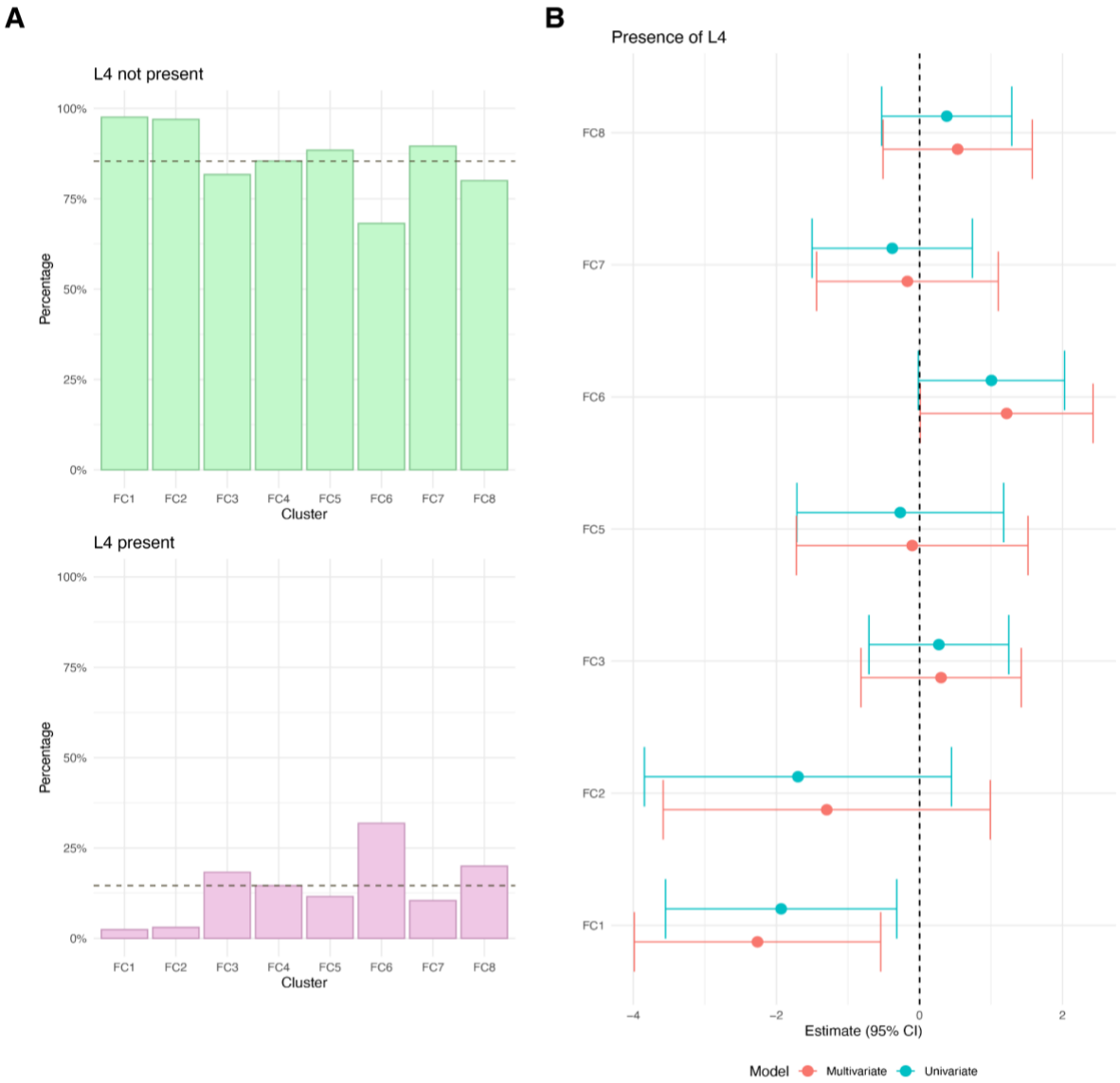

Figure S12 Crohn's disease patients only. (A) For each FC cluster, panels show the proportion of individuals with Montreal L4 (upper gastrointestinal inflammation), recorded as present or non present. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for L4: present versus not present (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), perianal disease (yes, no) and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

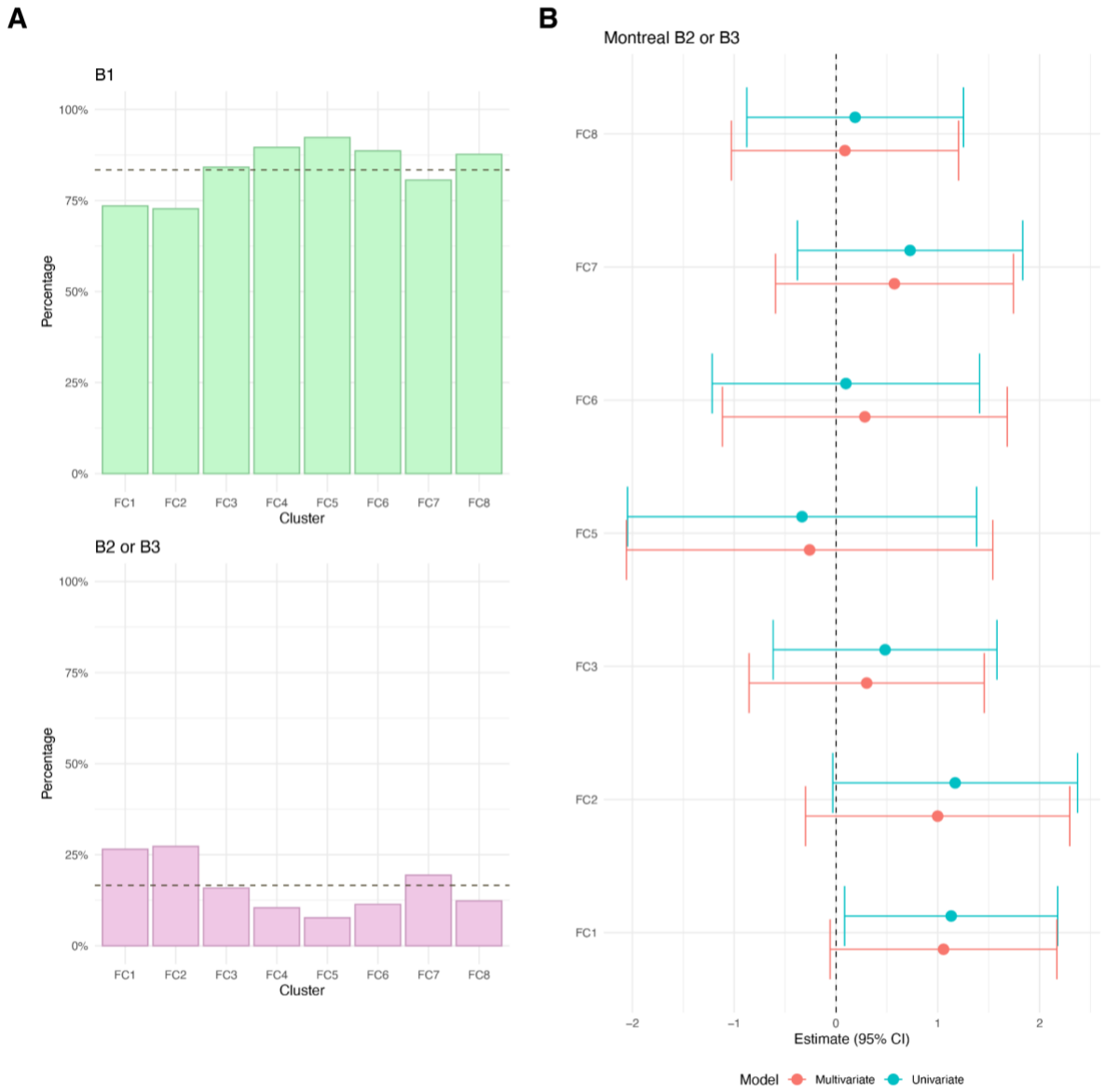

Figure S13. Crohn's disease patients only. (A) For each FC cluster, panels show the proportion of individuals with Montreal behaviour recorded as B1 or B2/B3 respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for Montreal behaviour: B2/B3 versus B1 (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), perianal disease (yes, no) and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

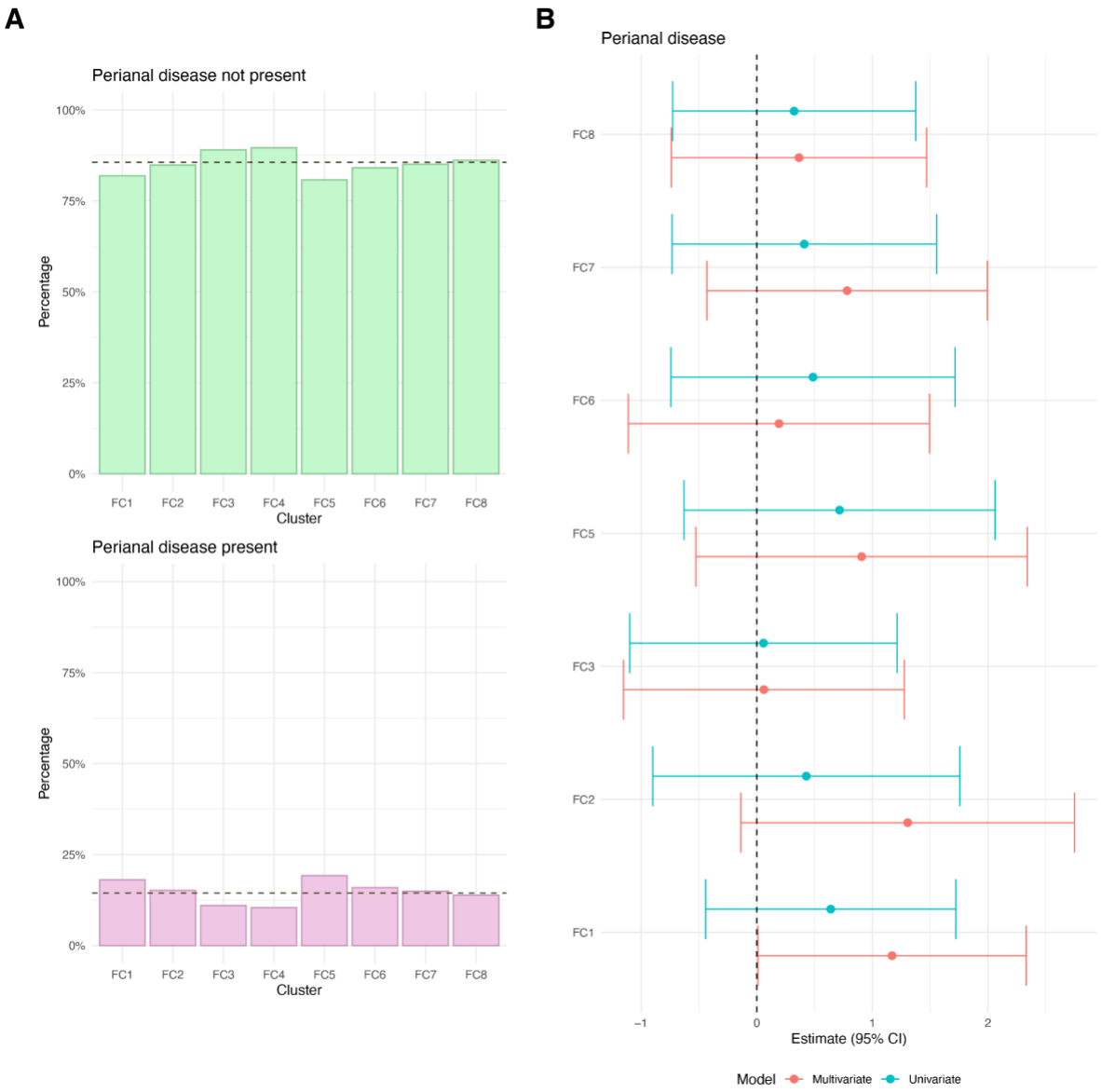

Figure S14 Crohn's disease patients only. (A) For each FC cluster, panels show the proportion of individuals with perianal disease recorded as present or not present. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for perianal disease: present versus not present (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), perianal disease (yes, no) and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

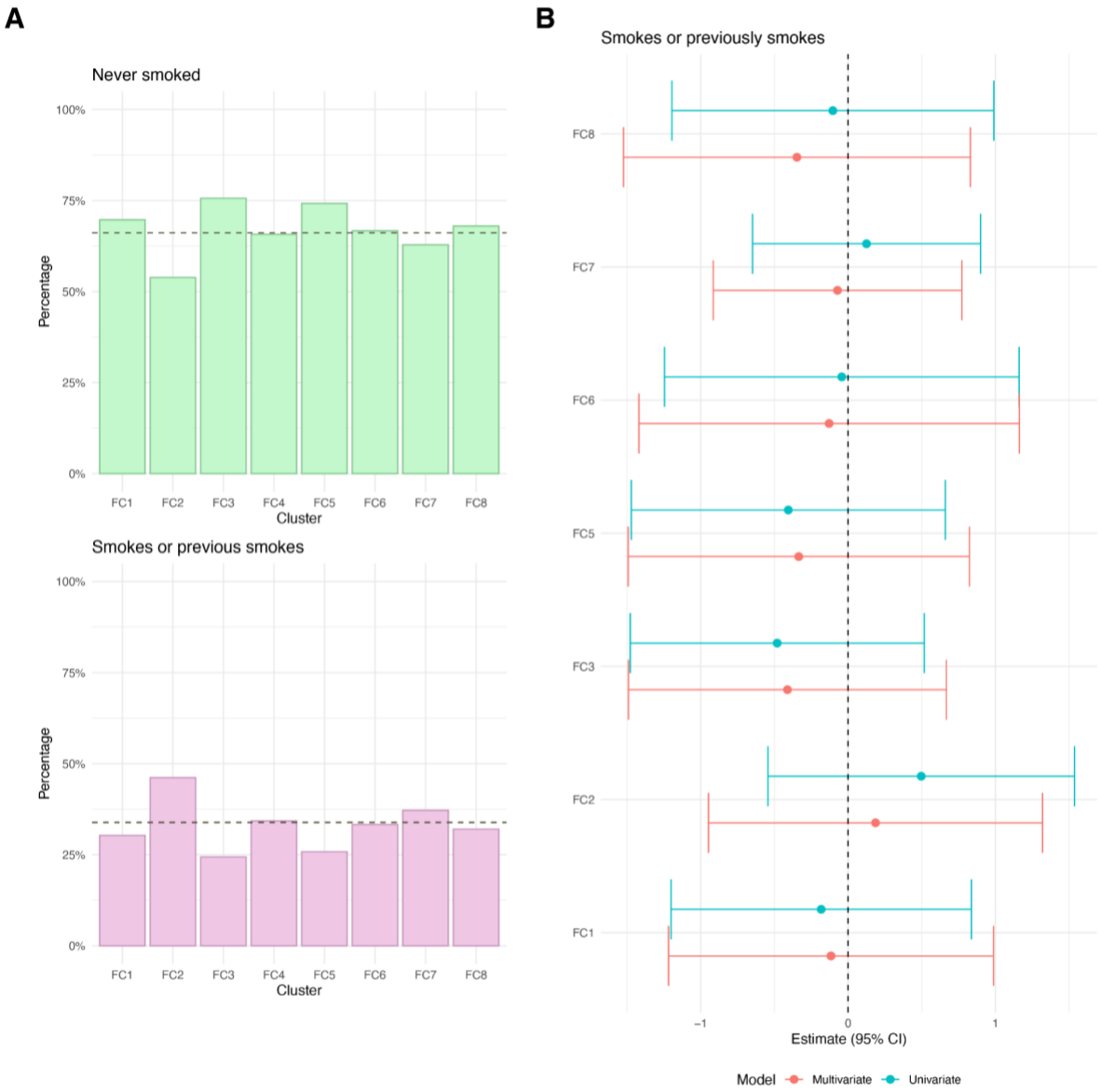

Figure S15. Ulcerative colitis patients only. (A) For each FC cluster, panels show the proportion of individuals with smoking behaviour recorded as "no" (no and never) and "yes" (current or previously smoked) at diagnosis respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for smoking: yes versus no (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking and Montreal extent (E1, E2, E3). The dashed vertical lines are used as a reference to indicate no effect.

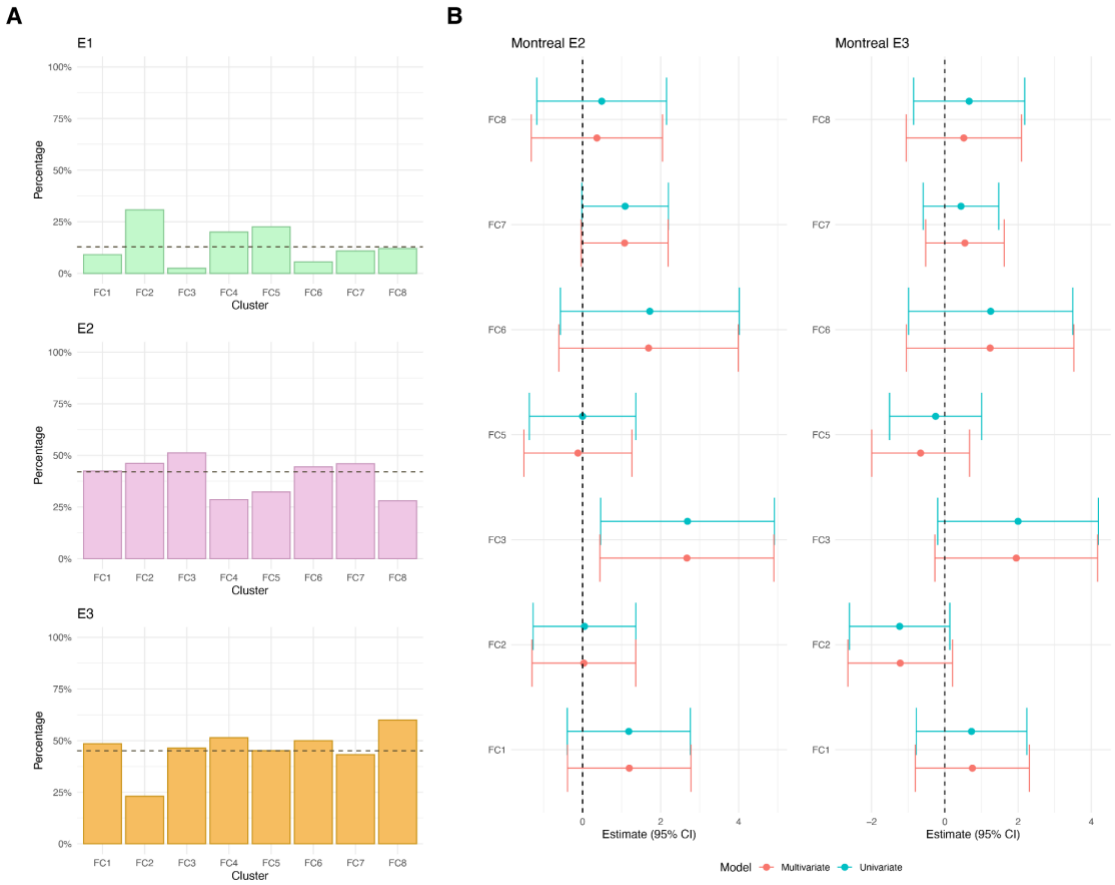

Figure S16. Ulcerative colitis patients only. (A) For each cluster, panels show the proportion of individuals with Montreal extent recorded as E1, E2 and E3 respectively. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for Montreal extent: E2 and E3 versus E1 (baseline category) in a multinomial logistic regression model that uses FC cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case FC4). The multivariate model includes age, sex, smoking and Montreal extent (E1, E2, E3). The dashed vertical lines are used as a reference to indicate no effect.

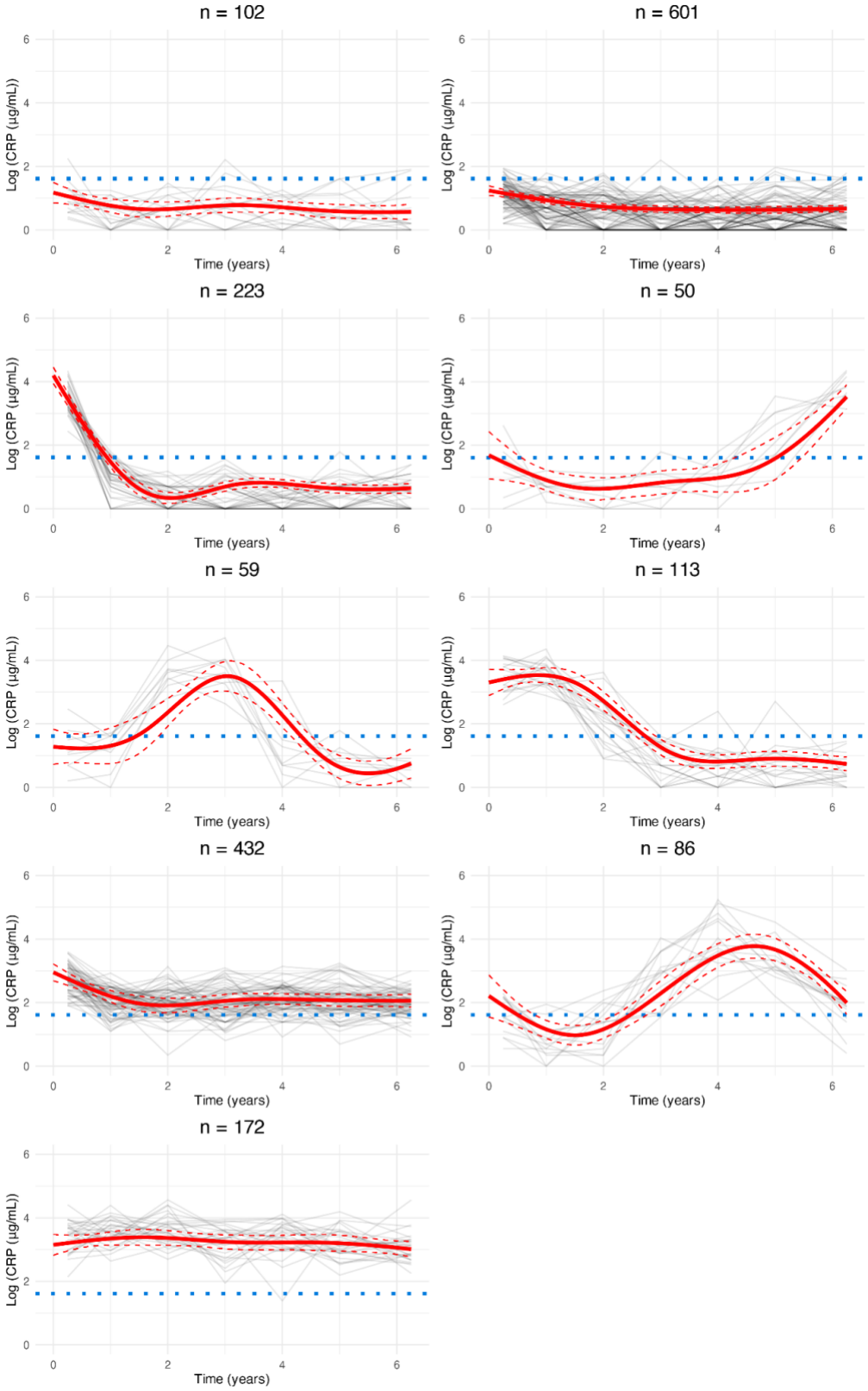

Figure S17. Cluster trajectories obtained from LCMM assuming nine clusters fitted to C-reactive protein data. Red lines indicate predicted mean cluster profiles with 95% confidence intervals. Dotted horizontal lines indicate log(5µg/mL). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from groups of six subjects.

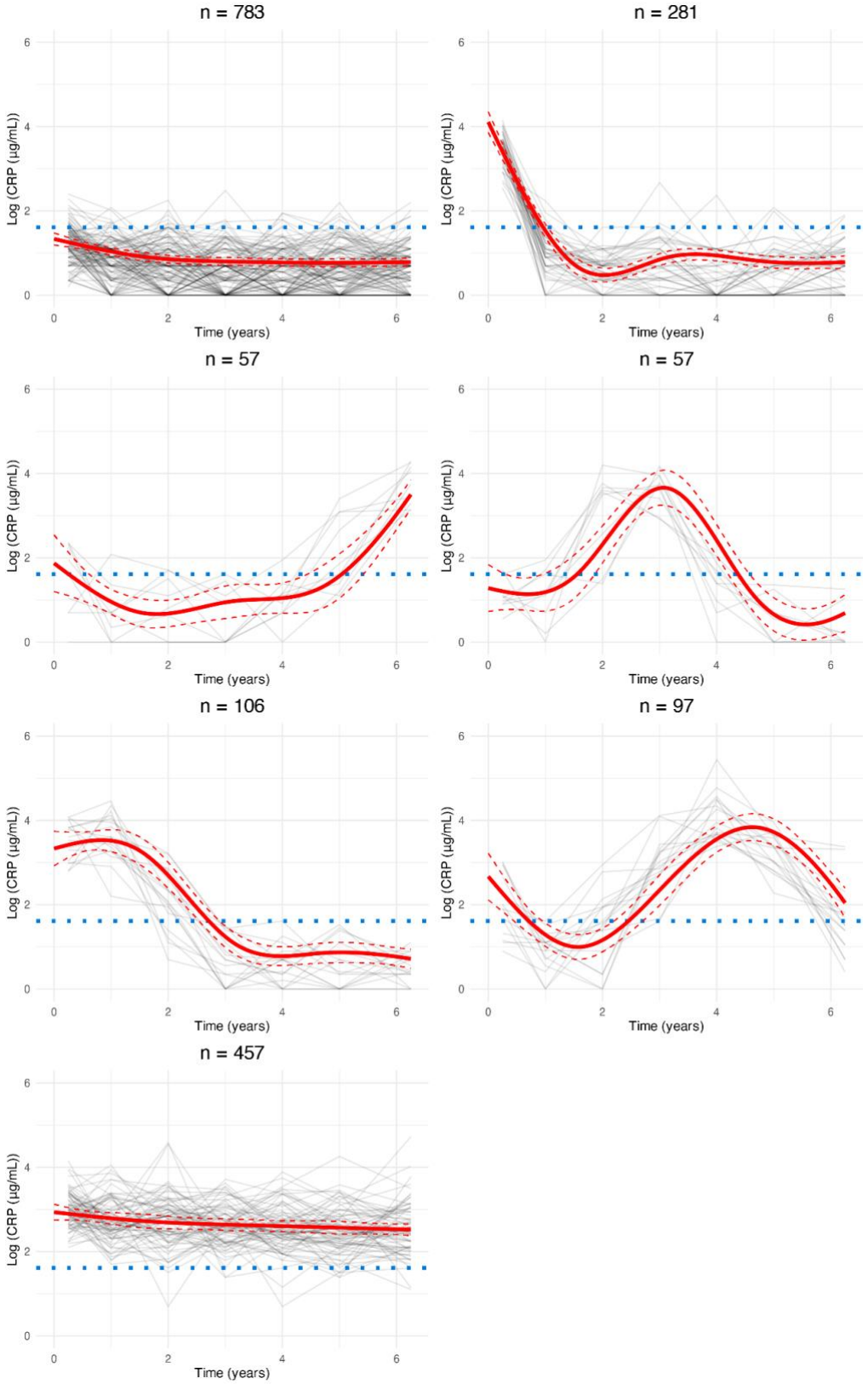

Figure S18. Cluster trajectories obtained from LCMM assuming seven clusters fitted to C-reactive protein data. Red lines indicate predicted mean cluster profiles with 95% confidence intervals. Dotted horizontal lines indicate log(5µg/mL). For visualisation purposes, pseudo subject-specific trajectories have been generated by amalgamating observations from groups of six subjects.

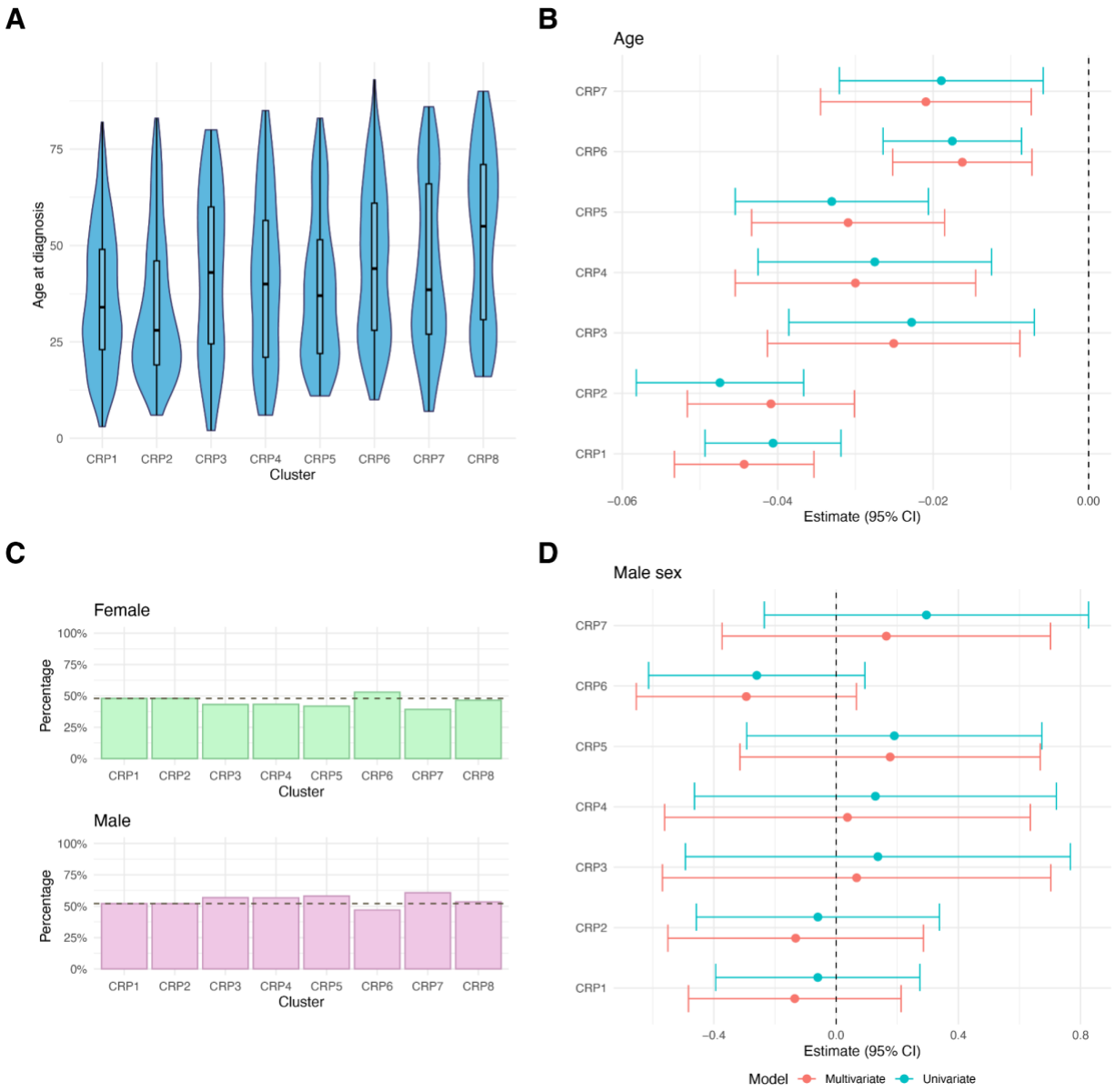

Figure S19. (A) For each CRP cluster, violin plots show the distribution of age at diagnosis across subjects, highlighting median and interquartile ranges. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for age in a multinomial logistic regression model that uses CRP cluster assignment as outcome. (C) For each CRP cluster, panels show the proportion of individuals with female and male sex. The dashed horizontal line represents overall proportions across the entire CRP cohort. (D) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for male sex versus females (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. In (B) and (C), effect sizes are with respect to the reference cluster (in this case CRP8). In both cases, the multivariate model includes age, sex and IBD type as covariates. The dashed vertical lines are used as a reference to indicate no effect.

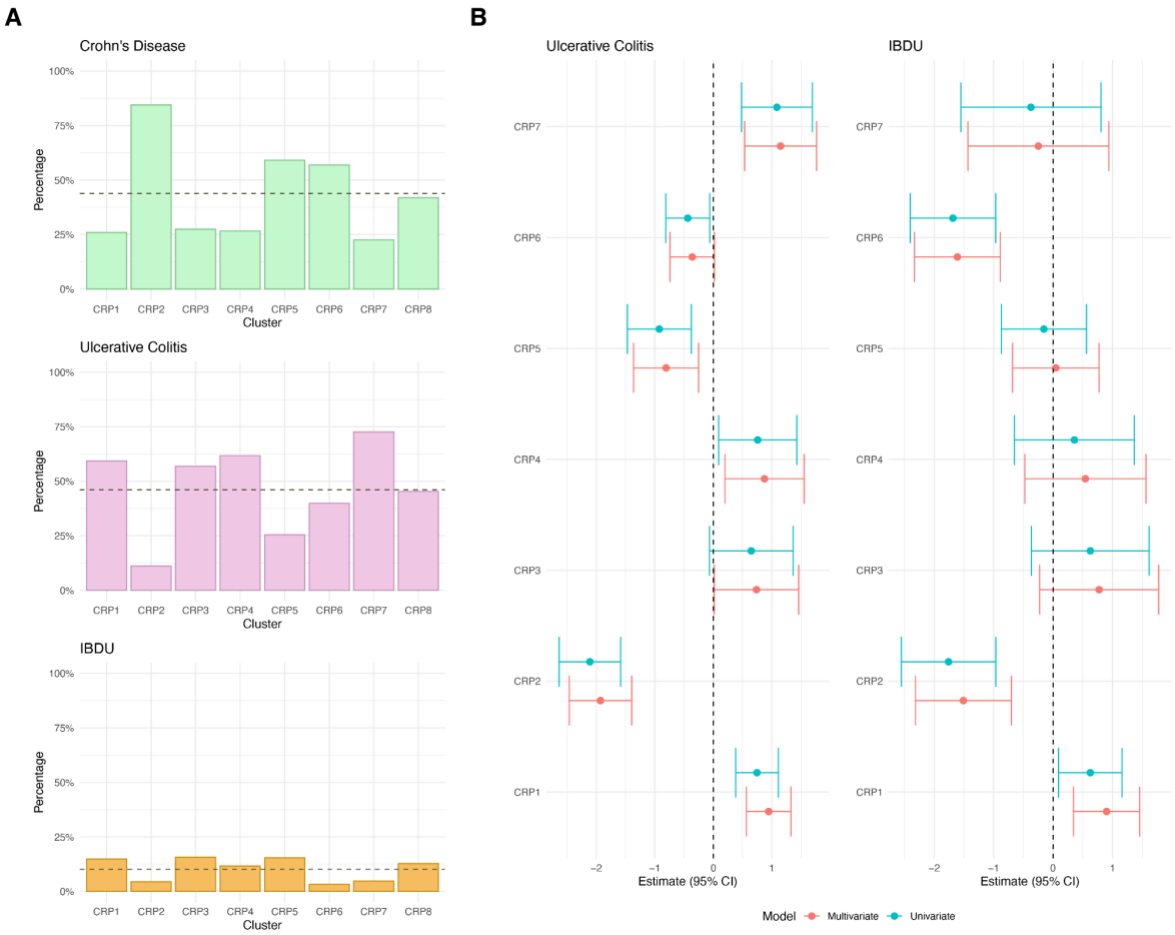

Figure S20. (A) For each CRP cluster, panels show the proportion of individuals with Crohn's disease, ulcerative colitis and IBDU respectively. The dashed horizontal line represents overall proportions across the entire CRP cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for IBD type: ulcerative colitis and IBDU versus Crohn's disease (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case CRP8). In both cases, the multivariate model includes age, sex and IBD type as covariates. The dashed vertical lines are used as a reference to indicate no effect.

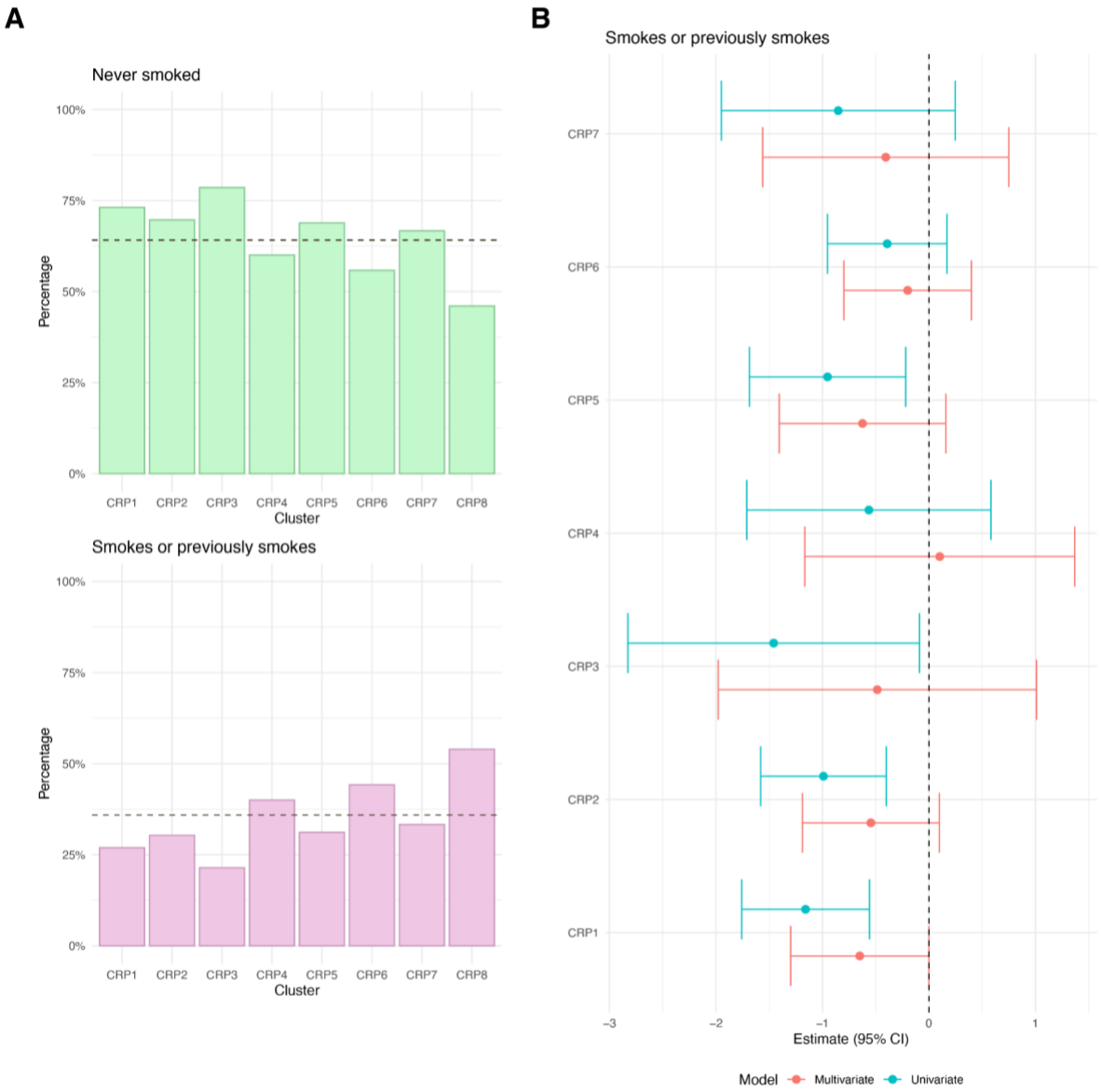

Figure S21. Crohn's disease patients only. (A) For each CRP cluster, panels show the proportion of individuals with smoking behaviour recorded as "no" (no and never) and "yes" (current or previously smoked) at diagnosis respectively. The dashed horizontal line represents overall proportions across the entire CRP cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for smoking: yes versus no (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case CRP8). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

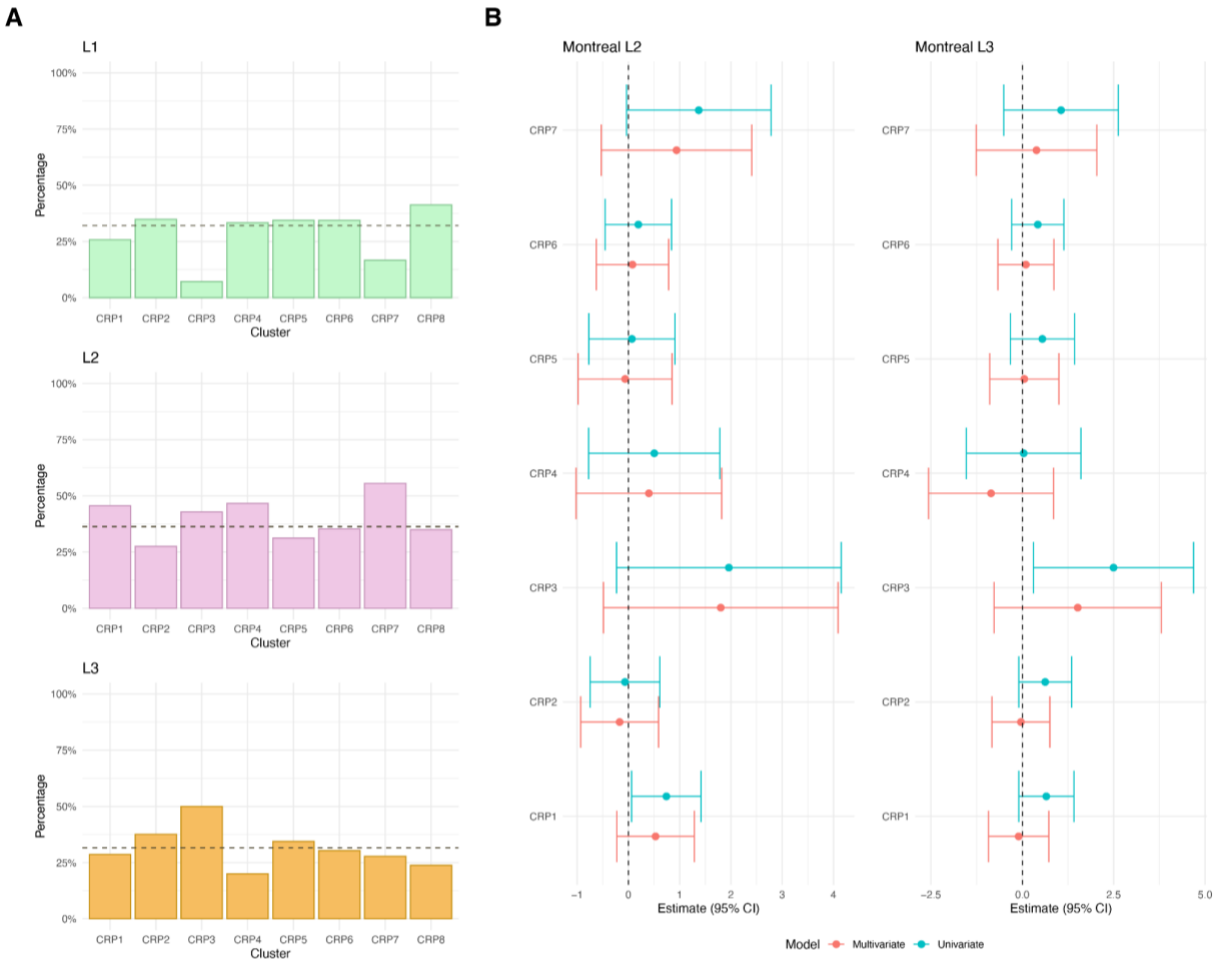

Figure S22. Crohn's disease patients only. (A) For each CRP cluster, panels show the proportion of individuals with Montreal Location recorded as L1, L2 and L3 respectively. The dashed horizontal line represents overall proportions across the entire CRP cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for Montreal Location: L2 and L3 versus L1 (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case CRP8). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

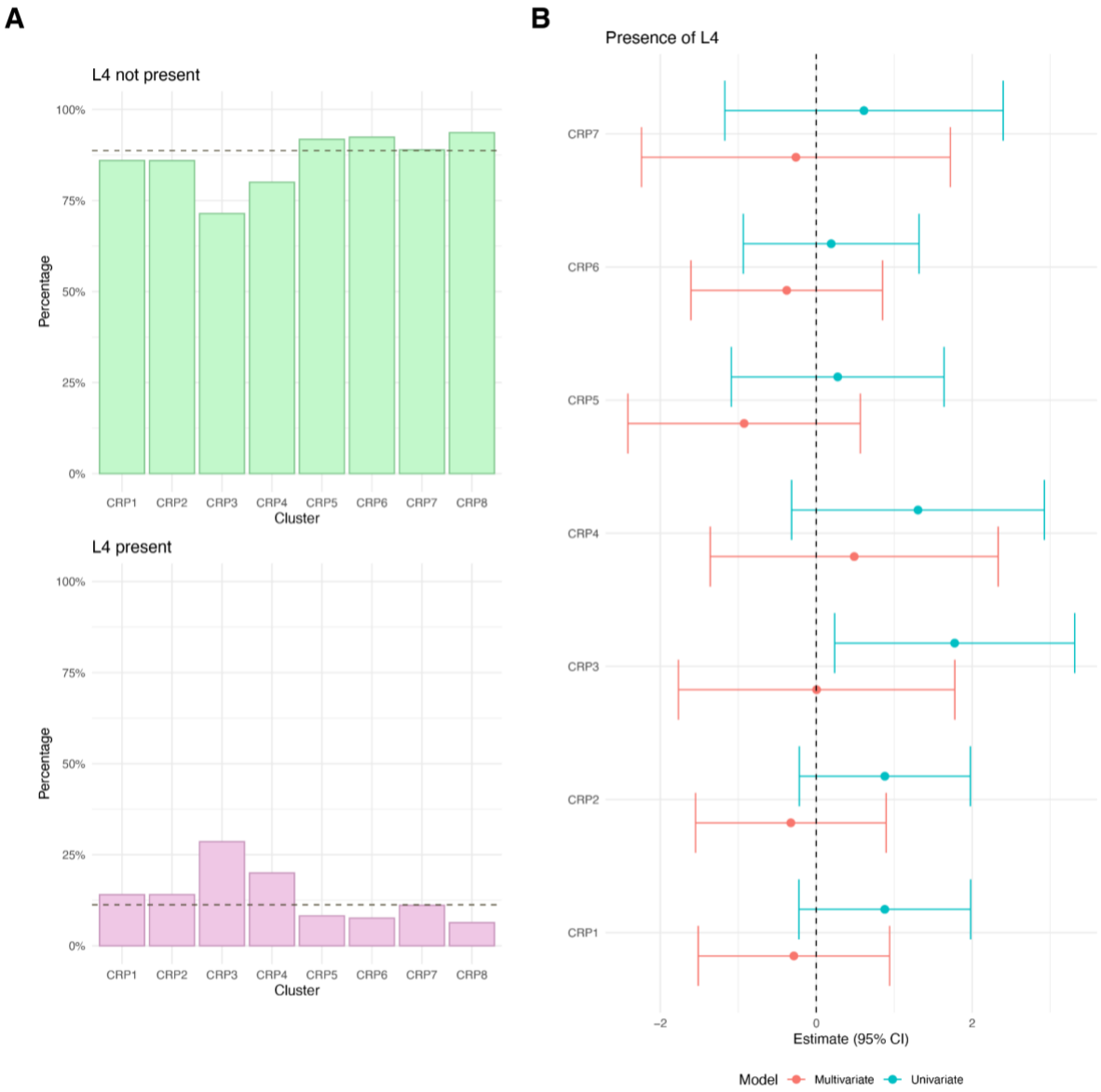

Figure S23. Crohn's disease patients only. (A) For each CRP cluster, panels show the proportion of individuals with Montreal L4 (upper gastrointestinal inflammation), recorded as present or non present. The dashed horizontal line represents overall proportions across the entire CRP cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for L4: present versus not present (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case CRP8). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

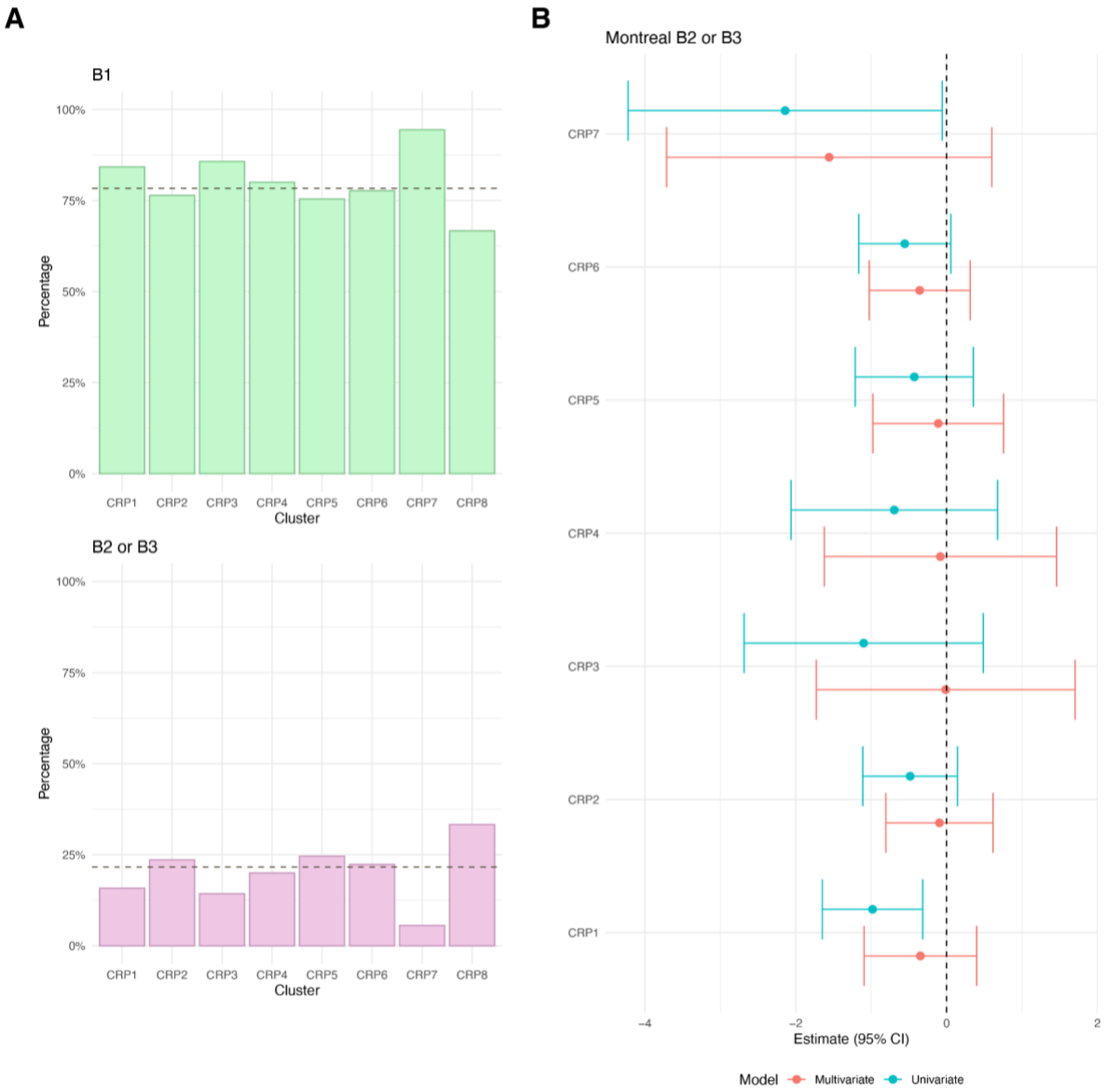

Figure S24. Crohn's disease patients only. (A) For each CRP cluster, panels show the proportion of individuals with Montreal behaviour recorded as B1 or B2/B3 respectively. The dashed horizontal line represents overall proportions across the entire CRP cohort. (B) Forest plot showing the estimated effect sizes and associated 95% confidence intervals for Montreal behaviour: B2/B3 versus B1 (baseline category) in a multinomial logistic regression model that uses CRP cluster assignment as outcome. Effect sizes are with respect to the reference cluster (in this case CRP8). The multivariate model includes age, sex, smoking, Montreal location (L1, L2, L3), upper gastrointestinal inflammation (L4), and Montreal behaviour (B1, B2/B3) as covariates. The dashed vertical lines are used as a reference to indicate no effect.

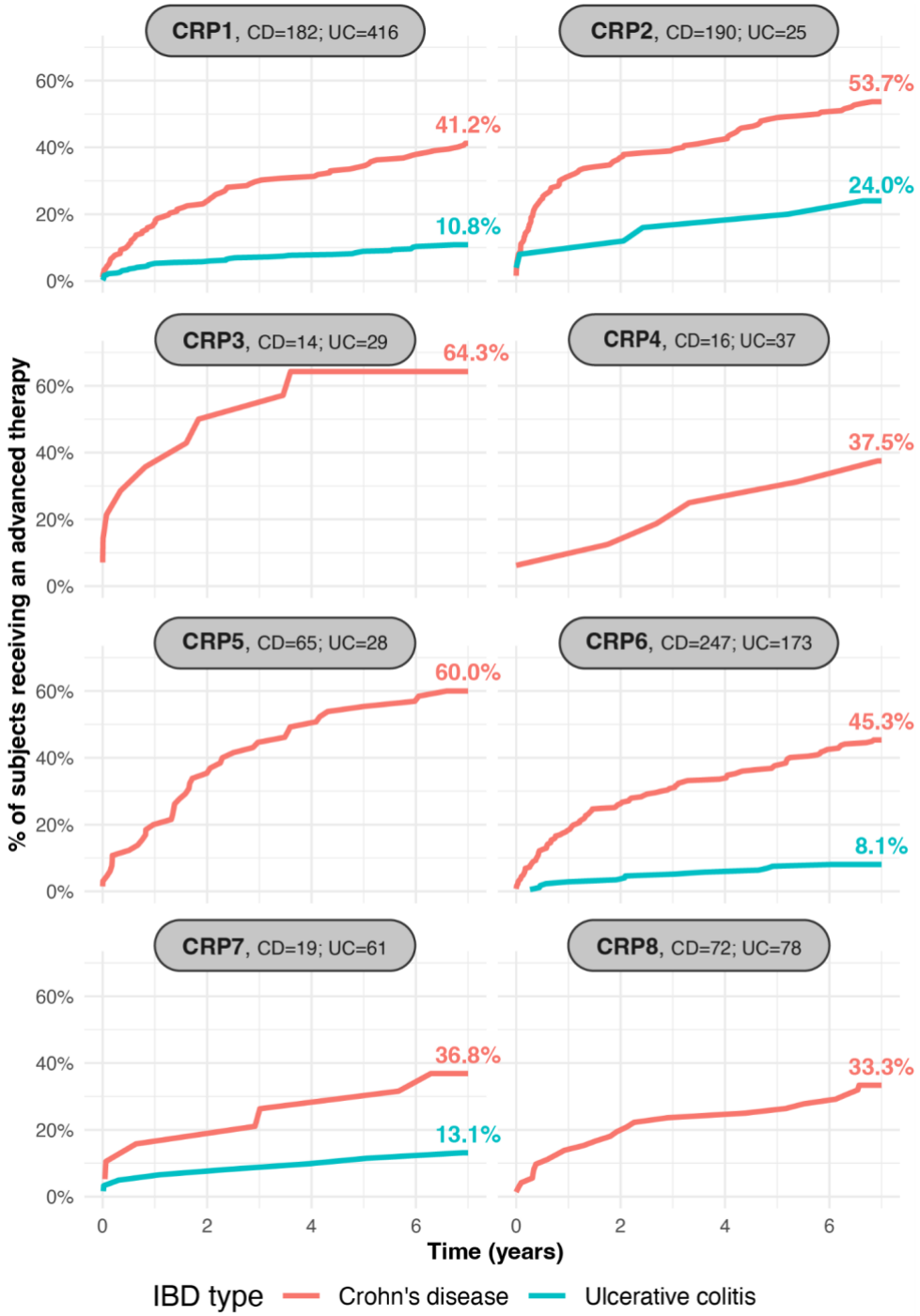

Figure S25. CRP cluster-specific cumulative distribution for first-line advanced therapy prescribing for Crohn's disease (red) and ulcerative colitis (teal) subjects. Clusters are ordered from lowest (CRP1) to highest (CRP8) cumulative inflammatory burden. The number of CD and UC subjects present in each cluster is displayed as panel titles. Curves which would describe fewer than five subjects are not shown.

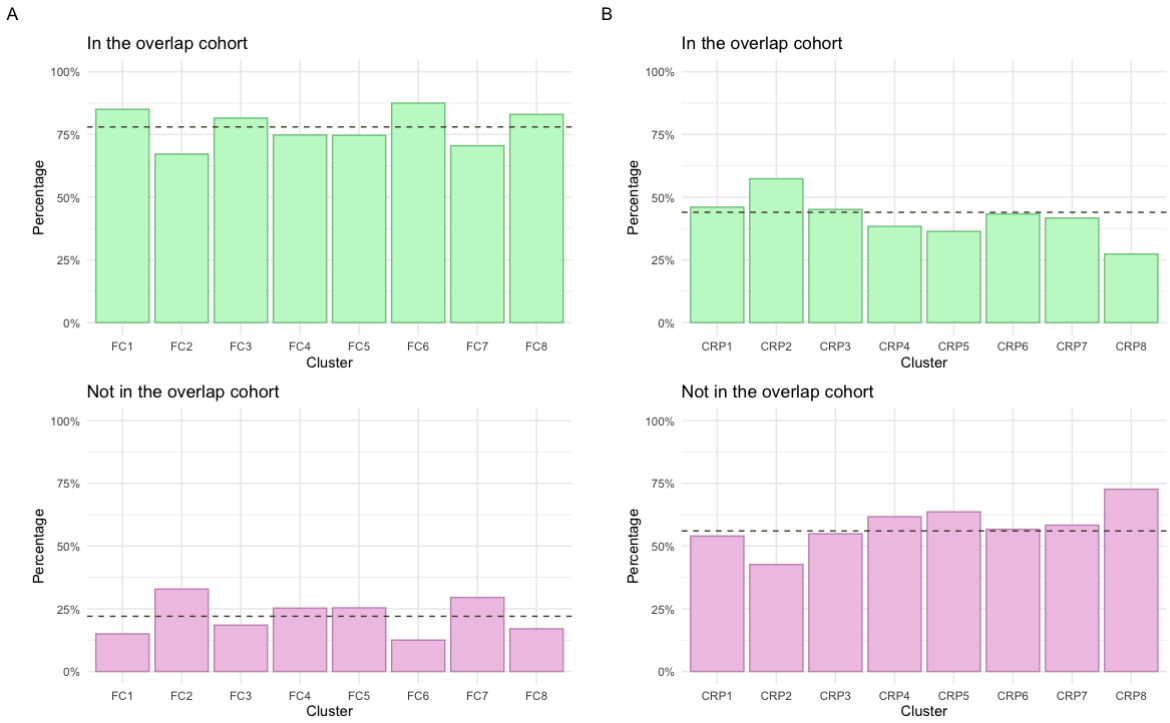

Figure S26. (A) For each FC cluster, panels show the proportion of individuals included in the overlap cohort, which consists of subjects included in both the FC and CRP analysis. The dashed horizontal line represents overall proportions across the entire FC cohort. (B) As in (A), but focusing on CRP clusters instead. The dashed horizontal line represents overall proportions across the entire CRP cohort.

Supplemental tables

Clusters	Maximum log-likelihood	AIC	BIC
2	-16532·10	33094·20	33168·35
3	-16354·39	32754·78	32868·47
4	-16260·21	32582·42	32735·66
5	-16201·86	32481·71	32674·49
6	-16141·82	32377·63	32609·96
7	-16109·77	32329·54	32601·42
8	-16076·89	32279·77	32591·19
9	-16039·66	32221·33	**32572·29**
10	**-16015·29**	**32188·59**	32579·09

Table S1. Likelihood-based model statistics for LCMMs fitted to faecal calprotectin data across 2–10 clusters using the chosen model specification. Boldface font indicates optimal values.

Clusters	Maximum log-likelihood	AIC	BIC
2	-15451·41	30932·82	31015·57
3	-15314·76	30675·52	30802·40
4	-15240·15	30542·29	30713·30
5	-15173·04	30424·09	30639·23
6	-15106·25	30306·50	30565·78
7	-15051·43	30212·86	30516·27
8	-15016·96	**30159·91**	**30507·45**
9	-15014·76	30171·53	30563·20
10	**-15014·40**	30186·81	30622·61

Table S2. Likelihood-based model statistics for LCMMs fitted to processed CRP data across 2–10 clusters using the chosen model specification. Boldface font indicates optimal values.