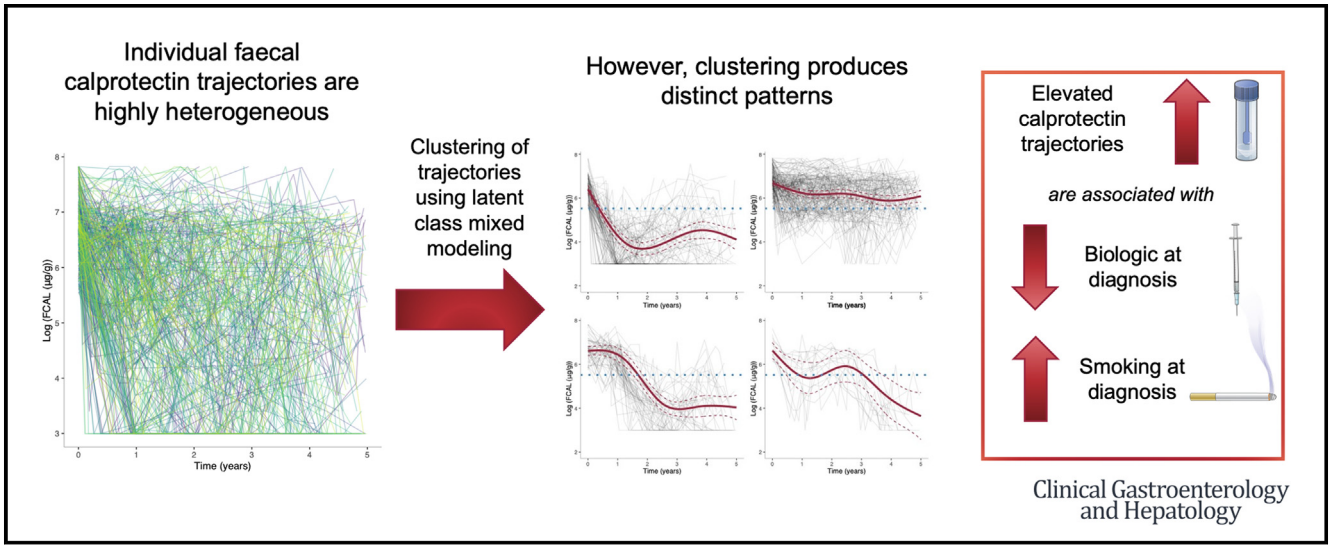
Longitudinal Fecal Calprotectin Profiles Characterize Disease Course Heterogeneity in Crohn's Disease

Nathan Constantine-Cooke,[1,2] Karla Monterrubio-Gómez,[1] Nikolas Plevris,[2,3] Lauranne A. A. P. Derikx,[4] Beatriz Gros,[3] Gareth-Rhys Jones,[3,5] Riccardo E. Marioni,[2] Charlie W. Lees,[2,3,*] and Catalina A. Vallejos[1,6,*]

[1]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom; [2]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom; [3]Edinburgh IBD Unit, Western General Hospital, Edinburgh, United Kingdom; [4]Inflammatory Bowel Disease Center, Radboud University Medical Center, Nijmegen, the Netherlands; [5]Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom; and [6]Alan Turing Institute, British Library, London, United Kingdom



BACKGROUND AND AIMS: The progressive nature of Crohn's disease is highly variable and hard to predict. In addition, symptoms correlate poorly with mucosal inflammation. There is therefore an urgent need to better characterize the heterogeneity of disease trajectories in Crohn's disease by utilizing objective markers of inflammation. We aimed to better understand this heterogeneity by clustering Crohn's disease patients with similar longitudinal fecal calprotectin profiles.

METHODS: We performed a retrospective cohort study at the Edinburgh IBD Unit, a tertiary referral center, and used latent class mixed models to cluster Crohn's disease subjects using fecal calprotectin observed within 5 years of diagnosis. Information criteria, alluvial plots, and cluster trajectories were used to decide the optimal number of clusters. Chi-square test, Fisher's exact test, and analysis of variance were used to test for associations with variables commonly assessed at diagnosis.

RESULTS: Our study cohort comprised 356 patients with newly diagnosed Crohn's disease and 2856 fecal calprotectin measurements taken within 5 years of diagnosis (median 7 per subject). Four distinct clusters were identified by characteristic calprotectin profiles: a cluster with consistently high fecal calprotectin and 3 clusters characterized by different downward longitudinal trends. Cluster membership was significantly associated with smoking ($P = .015$), upper gastrointestinal involvement ($P < .001$), and early biologic therapy ($P < .001$).

*Authors share co-senior authorship.

Abbreviations used in this paper: AIC, Akaike information criterion; BIC, Bayesian information criterion; CD, Crohn's disease; FCAL, fecal calprotectin; IBD, inflammatory bowel disease; IQR, interquartile range; LCMM, latent class mixed model.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2023.03.026

CONCLUSIONS: **Our analysis demonstrates a novel approach to characterizing the heterogeneity of Crohn's disease by using fecal calprotectin. The group profiles do not simply reflect different treatment regimens and do not mirror classical disease progression endpoints.**

Keywords: Biomarker; Epidemiology; Monitoring.

Crohn's disease (CD) affects around 1 in 350 people in the United Kingdom,[1,2] with substantial variation in phenotypes and disease outcomes. Historically, 30% follow a quiescent disease course,[3] while many will require surgery due to strictures, fistulas, or lack of response to medical therapy. Despite this heterogeneity, our ability to characterize disease variability remains poor and, in the case of Montreal location and behavior, involves invasive examinations, which limits the suitability of frequent longitudinal measurements.

Endoscopy remains the gold standard for monitoring inflammatory bowel disease (IBD); however, it is costly, invasive, and not without risk. As such, noninvasive stool markers like fecal calprotectin (FCAL) are increasingly used to objectively monitor inflammation in IBD.[4,5] FCAL is one of the most well characterized noninvasive biomarkers in IBD.[6] Multiple studies have shown that it accurately correlates with mucosal inflammation, in both ulcerative colitis and CD. Furthermore, the Effect of Tight Control Management on Crohn's Disease (CALM) study[7] demonstrated that a treat-to-target approach based on FCAL results in superior outcomes when compared with a treatment strategy based on symptoms alone. It is therefore sensible to consider using FCAL to characterize heterogeneity found in intestinal inflammation. By incorporating all FCAL data, instead of only FCAL measurements that can be dichotomized into specific time points, FCAL can be modeled as a continuous longitudinal process. While FCAL has previously been modeled in this way, no published research has attempted to cluster CD patients by longitudinal FCAL profiles, instead capturing heterogeneity across patients through a priori selected covariates (such subjects in endoscopic or clinical remission and those who have relapsed).[8,9]

Disease heterogeneity in CD has previously been described longitudinally by the Inflammatory Bowel Disease in South-Eastern Norway (IBSEN) study.[3] In the IBSEN study, subjects with CD chose which profile they believed best described their disease activity out of 4 profiles specified a priori. We aimed to perform a modernized iteration of this work by instead using FCAL profiles to characterize patient heterogeneity. We hypothesize that an unsupervised analysis to uncover latent patient subgroups with distinct longitudinal FCAL patterns can lead to better disease characterization.

Materials and Methods

Study Design

We performed a retrospective cohort study at the Edinburgh IBD Unit to determine if there were subgroups within the CD patient population identifiable from FCAL measurements which had been collected within 5 years of diagnosis. We modeled longitudinal FCAL profiles using latent class mixed models (LCMMs),[10] an extension of linear mixed effects models, which enable the identification of distinct subgroups with shared longitudinal patterns. LCMMs have been used to model biomarker trajectories in many contexts.[11,12]

The data were obtained from a cohort study by Plevris et al,[13] which identified all incident CD cases between 2005 and 2017 at the Edinburgh IBD Unit that fulfilled set inclusion criteria. For all patients, electronic health records (TrakCare; InterSystems, Cambridge, MA) were used to extract demographic as well as outcomes and FCAL values (both up to June 2019). Data for drug treatments and disease location were also extracted.

Criteria and Definitions

First, the inclusion criteria from Plevris et al[13] were applied: (1) a CD diagnosis between 2005 and 2017, (2) an initial FCAL measurement at diagnosis (or within 2 months) and prior to treatment, (3) an initial diagnostic FCAL result $\geq$250 μg/g, (4) an accurate date of diagnosis, (5) at least 1 additional FCAL measurement within 12 months of diagnosis, (6) at least 12 months of follow-up, and (7) neither having surgery nor a Montreal disease progression/new perianal disease within 12 months of diagnosis. Second, the following additional criterion was applied in this study: at least 3 FCAL measurements within 5 years of diagnosis.

The following information was available at diagnosis: sex, age, smoking status, FCAL, and Montreal location and behavior. Treatments prescribed within 1 year of diagnosis were also recorded: mesalamine, corticosteroids, thiopurines, methotrexate, exclusive enteral nutrition, and biologic therapies.

FCAL Assay

The Edinburgh IBD Unit has been using FCAL for diagnostic and monitoring purposes since 2005. Stool samples have been routinely collected at all healthcare interactions.[13] Samples are stored at −20$\underline{o}$C and FCAL is measured using a standard enzyme-linked immunosorbent assay technique (Calpro AS, Lysaker, Norway). All FCAL measurements in this study were performed using the same protocol and assay.

Statistical Analysis

Descriptive statistics are presented as median (interquartile range) for continuous variables. Frequencies with percentages are provided for categorical variables.

FCAL measurements >2500 μg/g were set to 2500 μg/g, the upper range for the assay. Likewise, measurements reported as less than the lower range for the assay, 20 μg/g, were set to 20 μg/g. FCAL values were log transformed before the models were fitted. To model the FCAL trajectories and find clusters, we used LCMMs with longitudinal patterns captured using natural cubic splines.[10] Natural cubic splines provide a flexible framework to model FCAL trajectories while remaining stable at either end of the study follow-up period.[14] Using natural cubic splines results in fewer parameters needing to be estimated compared with polynomial regression which requires a high-degree polynomial to achieve the same level of flexibility.[15] Between 2 and 5 knots were considered for the splines, and their performance was compared using Akaike information criterion (AIC). Three knots were found to produce the optimal AIC within this range and were placed at the quartiles of the FCAL measurement times. A full model description is provided in the Supplementary Appendix.

LCMMs assuming 2–6 clusters were fitted. For each number of clusters, the optimal model was found via a grid search approach (50 runs with 10 maximum iterations) following the vignette provided as part of the lcmm R package. Models converged based on parameter and likelihood stability, and on the negativity of the second derivatives. After each optimal model was found, the log likelihood, AIC, and Bayesian information criterion (BIC) were calculated. An alluvial plot was produced to provide intuition of how additional clusters are formed as the number of assumed clusters increases. These findings were used to decide on the appropriate number of clusters in our study population. As suggested in the lcmm package vignette, goodness of fit for the selected model was assessed by exploring whether model residuals were normally distributed. Uncertainty in cluster assignments was quantified using posterior classification probabilities. To visualize overall trajectories within each cluster, point estimates for each of the model parameters were used and statistical uncertainty was visualized using 95% confidence intervals.

Marginal associations between cluster membership and information available at the time of diagnosis were explored. Chi-square tests and Fisher's exact tests, dependent on suitability, were used for categorical variables. Analysis of variance was used for continuous variables. Upper gastrointestinal inflammation (L4) and perianal disease (P) were tested separately to Montreal location (L1–L3) and Montreal behavior (B1–B3), respectively.

Potential evidence of treatment effects was garnered by testing for associations between cluster membership and whether each treatment was prescribed within 1 year of diagnosis using Fisher's exact test. Biologic prescriptions within 3 months of diagnosis were also considered to study potential earlier treatment effects.

A 5% significance level was used for all statistical tests. Bonferroni adjustments have also been used to provide adjusted *P* values.

As an exploratory analysis, a multinomial logistic regression model,[16] and a random forest classifier[17] were used to predict cluster allocations using information available at the time of diagnosis and biologic prescriptions. For this purpose, a 75:25 train–test split with 4-fold cross-validation was used.[18] Classification performance was assessed via area under the curve extended to multiple classes.[19]

R (version 4.2.1; R Foundation for Statistical Computing, Vienna, Austria) was used for all statistical analyses using the lcmm (v.1.9.5),[20] nnet (v.7.3-17),[21] ranger (v.0.13.1),[22] datefixR (v.0.1.4),[23] tidyverse (v.1.3.1),[24] tidymodels (v.0.2.0),[25] vip (v0.3.2),[26] and ggalluvial (v.0.12.3)[27] R libraries. The analytical reports generated for this study and corresponding source code are hosted online (https://vallejosgroup.github.io/lcmm-site/).

Ethics

As this study was considered a retrospective audit due to all data having been collected as part of routine clinical care, no ethical approval or consent was required

What You Need to Know

Background

Our current ability to characterize heterogeneity in Crohn's disease is based on surgical outcomes, hospitalizations, and Montreal disease progression. We could possibly instead use longitudinal fecal calprotectin trajectories to describe heterogeneity in Crohn's disease using inflammation over time.

Findings

We found 4 clusters in a Crohn's disease patient cohort. Cluster membership was associated with early biologic treatment, smoking status at diagnosis, and upper gastrointestinal involvement. The largest of these clusters demonstrated consistently high fecal calprotectin and were less likely to have received early biologic treatment.

Implications for patient care

Crohn's disease patients appear to belong to distinct subgroups based on fecal calprotectin profiles. Instead of only utilizing clinical endpoints, considering fecal calprotectin over time allows an inflammation-driven approach to classification and may enable risk stratification in the future.

as per UK Health Research Authority guidance. Caldicott guardian approval (NHS Lothian) was granted (Project ID: 18002).

Results

FCAL Measurements

The study by Plevris et al[13] found 1390 incident CD cases. After removing individuals without an accurate diagnostic date or diagnostic FCAL (±60 days of diagnosis), 50 patients had diagnostic FCAL <250 μg/g. Once the additional inclusion criterion of at least 3 FCAL measurements was also applied, 356 subjects met the inclusion criteria for this study (Figure 1, Table 1). Across these patients, 2856 FCAL measurements were recorded within 5 years of diagnosis. The median frequency of FCAL measurements for a subject within this period was 7 (interquartile range, 5–10). The distributions of all FCAL measurements and the number of measurements per subject are presented as supplemental material (Supplementary Figures 1 and 2).

Modeling FCAL Trajectories

LCMMs fitted with 2–6 assumed clusters all converged as per default convergence criteria. As seen in Figure 2, cluster assignments were largely stable across differing assumed clusters, particularly when comparing the 3-, 4-, and 5-cluster models. Performance metrics for each model considered are provided in Supplementary Table 1. BIC suggested that the 2-cluster model was most appropriate, but this model was discarded, as visual inspection of the inferred trajectories suggested a larger number of distinct clusters (Supplementary Figures 3 and 4). AIC favored the 5-cluster model. However, this model was found to overfit the data, as some of the inferred trajectories were similar (Supplementary Figure 5). As a parsimonious choice, we selected the 4-cluster model.

The distribution of the model residuals for the 4-cluster model is bell shaped, but the quantile-quantile plot suggested some deviations from normality in the tails of the distribution within and across clusters (Supplementary Figures 6 and 7). Figure 3 presents the log mean profiles for the 4-cluster model alongside subject-specific observed FCAL trajectories. The model identified 3 main groups of patients: clusters 1, 2, and 3 (92, 191, and 58 subjects, respectively) and a small cluster 4 with 15 subjects. Clusters 1 and 3 display similar profiles—both showing a sharp decrease in FCAL, which then remains low. However, cluster 1 is differentiated by the decrease occurring immediately after diagnosis, while this decrease does not occur until around a year after diagnosis for cluster 3. In contrast, cluster 2 is characterized by a mean profile that remains consistently high: never dropping below the 250 μg/g clinical threshold for disease activity. Finally, the mean profile for cluster 4 exhibits an initial decrease, but this is not sustained during the first 3 years.

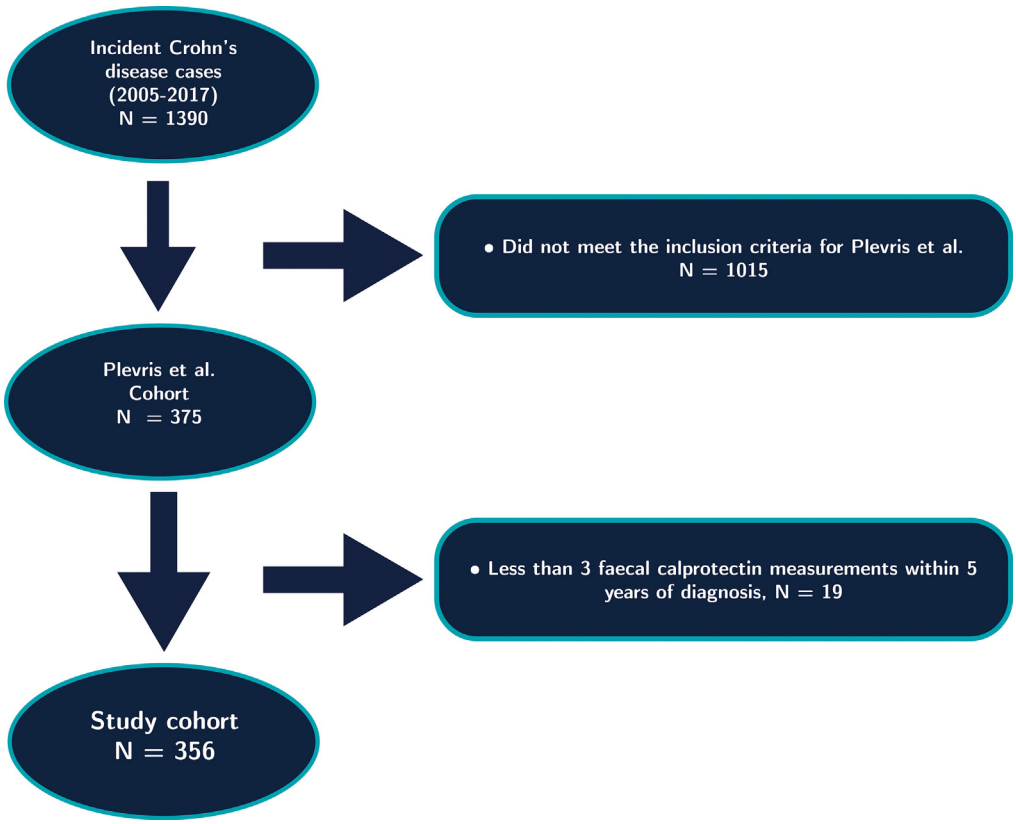


Figure 1. Flowchart demonstrating data processing steps.

Table 1. Cohort Characteristics and Treatments Prescribed to the Cohort

	Overall Population	Cluster				P	P_{adj}
		1	2	3	4		
Sex						.157	1
Male	183 (51)	43 (47)	107 (56)	24 (41)	9 (60)		
Female	173 (49)	49 (53)	84 (44)	34 (59)	6 (40)		
Age at diagnosis						.851	1
First quartile (q_1)	16.0	20.4	15.3	15.1	14.5		
Median (q_2)	27.3	29.6	26.4	26.8	21.7		
Third quartile (q_3)	48.7	45.3	49.9	50.9	51.2		
Smoking status						.015[a]	.116
Smoker	70 (20)	22 (24)	43 (23)	4 (7)	1 (7)		
Nonsmoker	286 (80)	70 (76)	148 (77)	54 (93)	14 (93)		
Diagnostic FCAL, μg/g						.131	1
First quartile (q_1)	590	500	610	630	592		
Median (q_2)	820	725	900	825	660		
Third quartile (q_3)	1140	986	1270	1180	1160		
Montreal behavior						.494	1
Inflammatory (B1)	323 (91)	80 (87)	175 (92)	55 (95)	13 (87)		
Stricturing (B2)	30 (8)	10 (11)	15 (8)	3 (5)	2 (13)		
Penetrating (B3)	3 (1)	2 (2)	1 (1)	0 (0)	0 (0)		
Perianal disease (P)[b]						.776	1
Present	57 (16)	17 (18)	28 (15)	9 (16)	3 (20)		
Not present	299 (84)	75 (82)	163 (85)	49 (84)	12 (80)		
Montreal location						.125	1
Ileal (L1)	95 (27)	22 (24)	58 (30)	9 (16)	6 (40)		
Colonic (L2)	140 (39)	39 (42)	68 (36)	30 (52)	3 (20)		
Ileocolonic (L3)	121 (34)	31 (34)	65 (34)	19 (33)	6 (40)		
Upper gastrointestinal (L4)[c]						<.001[d]	.002[e]
Present	84 (24)	8 (9)	51 (27)	20 (34)	5 (33)		
Not present	272 (76)	84 (91)	140 (73)	38 (66)	10 (67)		
Mesalamine						.326	1
Yes	76 (21)	15 (16)	48 (25)	11 (19)	2 (13)		
No	280 (79)	77 (84)	143 (75)	47 (81)	13 (87)		
Thiopurine						.023[a]	.161
Yes	250 (70)	62 (67)	127 (66)	50 (86)	11 (73)		
No	106 (30)	30 (33)	64 (34)	8 (14)	4 (27)		
Corticosteroids						.983	1
Yes	298 (84)	78 (85)	159 (83)	48 (83)	13 (87)		
No	58 (16)	14 (15)	32 (17)	10 (17)	2 (13)		
Methotrexate						.139	.975
Yes	15 (4)	8 (9)	6 (3)	1 (2)	0 (0)		
No	341 (96)	84 (91)	185 (97)	57 (98)	15 (100)		
Exclusive enteral nutrition						.779	1
Yes	80 (22)	22 (24)	39 (20)	14 (24)	5 (33)		
No	213 (60)	54 (59)	115 (60)	35 (60)	9 (60)		
Not known	63 (18)	16 (17)	37 (19)	9 (16)	1 (7)		
Biologic within 3 mo						<.001[d]	.004[e]
Yes	44 (12)	23 (25)	14 (7)	5 (9)	2 (13)		
No	312 (88)	69 (75)	177 (93)	53 (91)	13 (87)		

Table 1. Continued

	Overall Population	Cluster 1	Cluster 2	Cluster 3	Cluster 4	*P*	P_{adj}
Biologic						<.001[d]	<.001[d]
Yes	94 (26)	42 (46)	35 (18)	12 (21)	5 (33)		
No	262 (74)	50 (54)	156 (82)	46 (79)	10 (67)		

Values are n (%), unless otherwise indicated. All prescriptions were prescribed within 1 year of diagnosis unless otherwise stated. Percentages when stratified across clusters are out of the total number of subjects in the cluster. Biologic is defined as infliximab, adalimumab, ustekinumab, or vedolizumab prescription. FCAL, fecal calprotectin.
[a]Significant at a 5% significance level.
[b]Perianal disease may be present concomitantly to B1, B2, or B3 disease behavior or separately.
[c]Upper gastrointestinal inflammation (L4) may be present in addition to ileal (L1), colonic (L2), or ileocolonic (L3) inflammation.
[d]Significant at a 0.1% significance level.
[e]Significant at a 1% significance level.

Association With Variables Available at Diagnosis

Out of the 8 variables typically available at diagnosis, we tested for association with class membership, 2 variables were found to be significant at the 5% significance level before applying a Bonferroni adjustment: smoking status ($P = .01$, $P_{adj} = .08$) and the presence of upper gastrointestinal inflammation ($P < .001$, $P_{adj} = .002$). A total of 24% and 23% of clusters 1 and 2, respectively, were smokers when they were diagnosed, whereas only 7% of clusters 3 and 4 smoked during this period. Only 9% of cluster 1 had upper gastrointestinal involvement at diagnosis in comparison with the 27%, 34%, and 33% in clusters 2, 3, and 4, respectively.

Association With Treatments

A difference in the percentage of subjects prescribed a biologic therapy within 1 year of diagnosis was observed across classes (Table 1): 46% of cluster 1 were prescribed one of these treatments, compared with 18% and 21% for clusters 2 and 3, respectively.

Out of the prescriptions considered, being prescribed a thiopurine within 1 year of diagnosis ($P = .023$, $P_{adj} = .16$) and being prescribed a biologic within either 3 months ($P < .001$, $P_{adj} = .004$) or 1 year of diagnosis ($P < .001$, $P_{adj} < .001$) were found to be significant before Bonferroni adjustment. However, class membership could not be predicted from demographic data and biologic prescriptions (area under the curve of 0.68 for the multinomial logistic regression model and 0.66 for the random forest classifier).

Discussion

In this study, 4 patient clusters in the CD population with distinct FCAL trajectories have been identified and described (Figure 3). To the best of our knowledge, we are the first to apply LCMMs to characterize latent patient heterogeneity using FCAL data, although others have applied linear mixed models to FCAL data[8,9] or have applied LCMMs in other disease contexts.[28,29]

We have demonstrated that cluster membership is associated with smoking and upper gastrointestinal

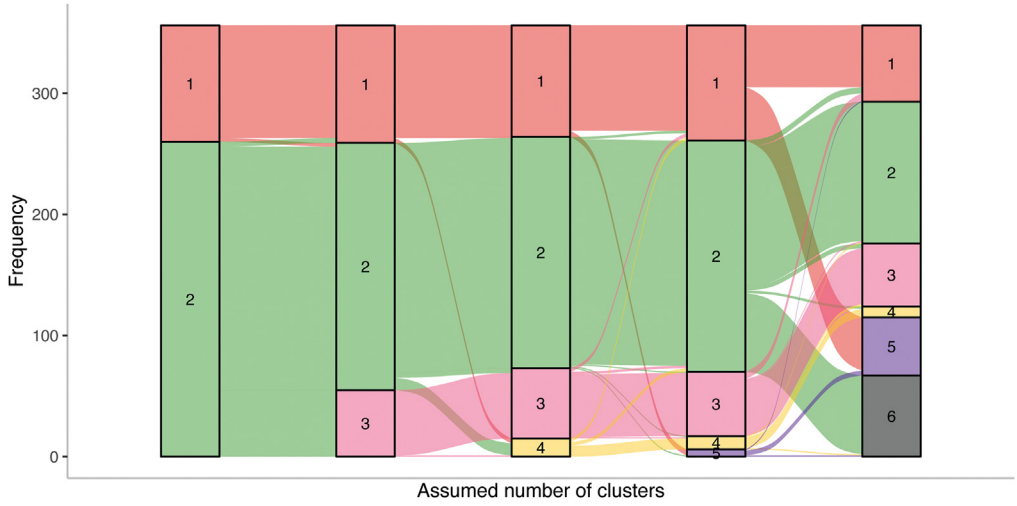


Figure 2. Alluvial plot demonstrating how cluster membership obtained from the FCAL profiles of CD patients changes as the assumed number of clusters increases. The height of each band indicates the size of each cluster.

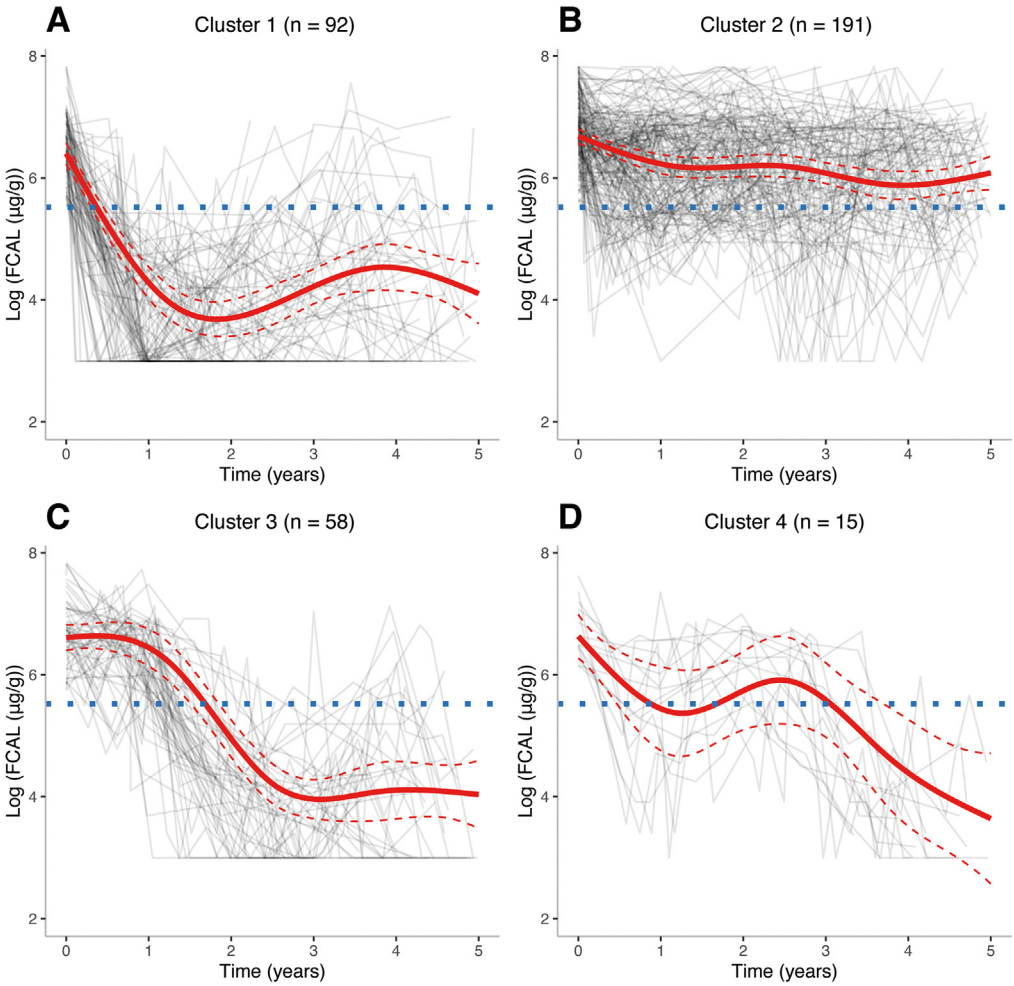


Figure 3. Log-transformed subject-specific 5-year FCAL profiles for the study cohort for (*A*) cluster 1, (*B*) cluster 2, (*C*) cluster 3, and (*D*) cluster 4. The red solid line represents the predicted mean trajectory for each cluster, while the red dotted lines represent 95% confidence intervals. The gray lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log (250 μg/g): the common threshold for biochemical remission in CD. See Supplementary Figure 8 for the fits in the original measurement scale.

inflammation. A comparatively high number of subjects who smoked at diagnosis were found in both clusters 1 and 2 despite cluster 1 being characterized by an overall decrease in FCAL and cluster 2 being characterized by a consistently high profile. The interpretation of this finding is not clear from our data. Previous research has found smoking to be associated with low drug concentrations for infliximab and adalimumab, mediating low remission rates in CD patients[30] in addition to being associated with undergoing surgery and disease progression.[31] Upper gastrointestinal involvement is likely a proxy for a more severe CD subphenotype. We also observed cluster membership to be associated with early biologic treatment. This is reasonable given the often reported association between FCAL and endoscopic activity and an association between biologic treatments and endoscopic healing for CD patients.[32,33]

While examining the residuals for the 4-cluster model, we found evidence against normality in the tails of the distribution, where some outliers can be observed. In most cases, this is driven by FCAL observations being truncated to be within the limits of detection of the assay (20–2500 μg/g). As subjects were required to have a FCAL above 250 μg/g at diagnosis to be included in the study, only the upper truncation applies for diagnostic FCAL. We believe that the impact of any violation of the assumption of normally distributed residuals on our findings is minimal. If this assumption is violated, then an inappropriate number of classes can be found when solely relying on model selection metrics such as AIC and BIC.[34] Instead, we also considered visual inspection (alluvial plots, mean cluster profiles vs subject-specific trajectories) in addition to AIC and BIC as a more rigorous approach to determining the number of latent classes, avoiding the identification of spurious clusters.

The approach demonstrated here has notable advantages over the methodology used by the IBSEN study, which required participants to choose which diagram they believed best described their disease activity out of 4 possible options.[3] Using FCAL profiles allows the quantification of inflammation in an objective manner, rather than using patient-reported symptom activity, which may be influenced by recency bias and the tendency for patient-reported data to exhibit extreme responses.[35] Furthermore, using FCAL allows longitudinal profiles to be generated in a data-driven manner. Instead of profiles needing to be generated based on prior beliefs and opinion, we can allow these profiles to be formed

naturally. Finally, FCAL profiles can be readily generated for many CD patients from electronic healthcare records without requiring active involvement from patients.

Some similarities can be observed between the clinically derived profiles in the IBSEN study cohort patterns and the cluster-specific mean profiles uncovered in this study. Both studies identified a large group of patients that exhibit a decline in severity of symptoms (clusters 1 and 3 in our study) and a group with chronic continuous symptoms (cluster 2 in our study). However, the IBSEN study identified a group with increasing intensity of symptoms that was not found by our analysis. Such differences may be due to the disconnect between symptoms and inflammation, which is commonly seen when using endoscopic activity scores.[36] Moreover, the IBSEN study findings were gathered before the widespread emergence of biologic therapies for CD and may not represent more modern trends that may not be well known a priori: demonstrating the advantage of being able to infer subgroup profiles in a data-driven manner.

In this study, 8 potential associations with variables typically available at diagnosis, and 7 potential associations with treatments have been explored. As such, we potentially invite criticism due to multiple testing. Indeed, some associations reported here (eg, between-cluster membership and smoking) fail to be significant after applying Bonferroni corrections. However, we believe that our findings here are biologically plausible and in line with other published literature.

The retrospective design of this study remains a limitation, and the results reported may be due to observational biases and should not be assigned a causal interpretation. Quantifying causal treatment effects from such observational data is an active area of research and beyond the scope of this study.[37,38] The data gathering process is observational, and while FCAL is collected routinely at all clinical interactions, subjects with more complicated disease are likely to have more measurements available. The retrospective study design also means that all subjects did not have the same treatment options at the same stage in their disease trajectories, as subjects may have been diagnosed at any time between 2005 and 2017. However, the date of diagnosis, converted to the number of days the subject was diagnosed after January 1, 2001, was considered for potential association with cluster membership, and no significant association was found ($P = .12$). We also acknowledge the potential for inclusion bias in this study. The study by Plevris et al[13] required subjects to have an FCAL of at least 250 μg/g at diagnosis and excluded subjects that met 1 of the endpoints within a year of diagnosis. The former potentially excludes subjects with milder disease, while the latter potentially excludes subjects with more aggressive disease. Extreme trajectories may therefore have been underrepresented in our analysis. While we are confident in the classes described here, additional classes may be found if the inclusion criteria were relaxed.

The clusters reported here are intended purely for exploring heterogeneity in CD and are not intended for use as predictors in a risk score. Indeed, some FCAL measurements were taken after typical outcomes of interest (eg, surgery); hence, cluster membership information is not a suitable risk factor. However, our approach provides an objective way to characterize disease trajectory heterogeneity using a routinely collected inflammation marker, providing a proof of concept for novel longitudinal patient stratification in the context CD. It should also be noted that only a single cohort was examined in this study, and generalizing these results to other populations requires caution.

Conclusions

We have demonstrated the suitability and utility of LCMMs for identifying clusters within the CD population based on FCAL profiles. After we found and described 4 clusters, we reported cluster membership to be significantly associated with smoking and upper gastrointestinal involvement. We believe that our findings are an important first step toward embracing longitudinal FCAL measurements to explain disease heterogeneity in CD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at http://doi.org/10.1016/j.cgh.2023.03.026.

References

1. Jones GR, Lyons M, Plevris N, et al. IBD prevalence in Lothian, Scotland, derived by capture–recapture methodology. Gut 2019;68:1953–1960.
2. Hamilton B, Green H, Heerasing N, et al. Incidence and prevalence of inflammatory bowel disease in Devon, UK. Frontline Gastroenterol 2021;12:461–470.
3. Henriksen M, Jahnsen J, Lygren I, et al. Clinical course in Crohn's disease: results of a five-year population-based follow-up study (the IBSEN study). Scand J Gastroenterol 2007; 42:602–610.
4. D'Amico F, Nancey S, Danese S, et al. A practical guide for faecal calprotectin measurement: myths and realities. J Crohns Colitis 2020;15:152–161.
5. Kennedy NA, Jones GR, Plevris N, et al. Association between level of fecal calprotectin and progression of Crohn's disease. Clin Gastroenterol Hepatol 2019;17:2269–2276.e4.
6. Plevris N, Lees CW. Disease monitoring in inflammatory bowel disease: evolving principles and possibilities. Gastroenterology 2022;162:1456–1475.e1.
7. Colombel JF, Panaccione R, Bossuyt P, et al. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. The Lancet 2017; 390:2779–2789.
8. De Vos M, Dewit O, D'Haens G, et al. Fast and sharp decrease in calprotectin predicts remission by infliximab in anti-TNF naïve patients with ulcerative colitis. J Crohns Colitis 2012;6:557–562.

9. Zhulina Y, Cao Y, Amcoff K, et al. The prognostic significance of faecal calprotectin in patients with inactive inflammatory bowel disease. Aliment Pharmacol Ther 2016;44:495–504.
10. 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;78:1–56.
11. Courvoisier D, Alpizar-Rodriguez D, Gottenberg J, et al. Rheumatoid arthritis patients after initiation of a new biologic agent: trajectories of disease activity in a large multinational cohort study. EBioMedicine 2016;11:302–306.
12. Jiang G, Luk AOY, Tam CHT, et al. Progression of diabetic kidney disease and trajectory of kidney function decline in Chinese patients with type 2 diabetes. Kidney Int 2019; 95:178–187.
13. 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;19:1835–1844.e6.
14. Elhakeem A, Hughes RA, Tilling K, et al. Using linear and natural cubic splines, SITAR, and latent trajectory models to characterise nonlinear longitudinal growth trajectories in cohort studies. BMC Med Res Methodol 2022;22:68.
15. James G, Witten D, Hastie T, et al, eds. An Introduction to Statistical Learning: With Applications in R. 2nd ed. New York, NY: Springer, 2021.
16. Kwak C, Clayton-Matthews A. Multinomial logistic regression. Nurs Res 2002;51:404–410.
17. Breiman L. Random forests. Mach Learn 2001;45:5–32.
18. Hastie T, Tibshirani R, Friedman J, eds. The Elements of Statistical Learning: Data Mining, Inteference, and Prediction. 2nd ed. New York, NY: Springer, 2009.
19. Hand DJ, Till RJ. A Simple Generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 2001;45:171–186.
20. Proust-Lima C, Philipps V, Diakite A, et al. lcmm: extended mixed models using latent classes and latent processes. R package version: 1.9.3. Available at: https://cran.r-project.org/package=lcmm. Accessed January 18, 2023.
21. Venables WN, Ripley BD. Modern Applied Statistics With S. 4th ed. New York, NY: Springer, 2002.
22. Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 2017;77:1–17.
23. Constantine-Cooke N. datefixR: fix really messy dates in R. R package version 0.1.4. Available at: https://CRAN.R-project.org/package=datefixR. Accessed February 20, 2022.
24. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1686.
25. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Available at: https://CRAN.R-project.org/package=tidymodels. Accessed March 19, 2022.
26. Greenwell BM, Boehmke BC. Variable importance plots—an introduction to the vip package. R Journal 2020;12:343–366.
27. Brunson JC. ggalluvial: layered grammar for alluvial plots. J Open Source Softw 2020;5:2017.
28. Chapuis N, Ibrahimi N, Belmondo T, et al. Dynamics of circulating calprotectin accurately predict the outcome of moderate COVID-19 patients. EBioMedicine 2022;80:104077.
29. Vistisen D, Andersen GS, Hulman A, et al. Progressive decline in estimated glomerular filtration rate in patients with diabetes after moderate loss in kidney function—even without albuminuria. Diabetes Care 2019;42:1886–1894.
30. Kennedy NA, Heap GA, Green HD, et al. Predictors of anti-TNF treatment failure in anti-TNF-naive patients with active luminal Crohn's disease: a prospective, multicentre, cohort study. Lancet Gastroenterol Hepatol 2019;4:341–353.
31. Lawrance IC, Murray K, Batman B, et al. Crohn's disease and smoking: is it ever too late to quit? J Crohns Colitis 2013; 7:e665–e671.
32. Jusué V, Chaparro M, Gisbert JP. Accuracy of fecal calprotectin for the prediction of endoscopic activity in patients with inflammatory bowel disease. Dig Liver Dis 2018; 50:353–359.
33. Narula N, Wong EC, Dulai PS, et al. Comparative effectiveness of biologics for endoscopic healing of the ileum and colon in Crohn's disease. Am J Gastroenterol 2022;117:1106–1117.
34. Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. Psychol Methods 2003;8:338–363.
35. Vaerenbergh YV, Thomas TD. Response styles in survey research: a literature review of antecedents, consequences, and remedies. Int J Public Opin Res 2012;25:195–217.
36. Koutroumpakis E, Katsanos K. Implementation of the simple endoscopic activity score in Crohn's disease. Saudi J Gastroenterol 2016;22:183–191.
37. Nogueira AR, Pugnana A, Ruggieri S, et al. Methods and tools for causal discovery and causal inference. Wiley Interdiscip Rev Data Min Knowl Discov 2022;12:e1449.
38. Hammerton G, Munafò MR. Causal inference with observational data: the need for triangulation of evidence. Psychol Med 2021; 51:563–578.

Correspondence
Address correspondence to: Nathan Constantine-Cooke, MSc, Prof Charlie W. Lees or Dr Catalina A. Vallejos, Institute of Genetics and Cancer, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom EH4 2XU. e-mail: nathan.constantine-cooke@ed.ac.uk.

Acknowledgments
A preprint of this work is hosted on medRxiv: https://doi.org/10.1101/2022.08.16.22278320.

The analytical reports generated for this study and corresponding code are hosted online (https://vallejosgroup.github.io/lcmm-site/).

The data used in this study is not publicly available, as it originates from patients who have not given consent for the data to be publicly shared. For access to the data, please contact Charlie W. Lees.

CRediT Authorship Contributions

Nathan Constantine-Cooke, MSc (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Methodology: Lead; Software: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Karla Monterrubio-Gómez (Conceptualization: Equal; Formal analysis: Supporting; Supervision: Supporting; Validation: Lead; Writing – original draft: Supporting; Writing – review & editing: Equal)

Nikolas Plevris (Conceptualization: Equal; Data curation: Lead; Writing – review & editing: Lead)

Lauranne A.A.P Derikx (Writing – review & editing: Equal)

Beatriz Gros (Data curation: Supporting; Writing – original draft: Supporting; Writing – review & editing: Equal)

Gareth-Rhys Jones (Data curation: Supporting; Writing – review & editing: Equal)

Riccardo E. Marioni (Supervision: Supporting; Writing – review & editing: Equal)

Charlie W. Lees (Conceptualization: Lead; Data curation: Equal; Funding acquisition: Lead; Supervision: Lead; Writing – review & editing: Lead)

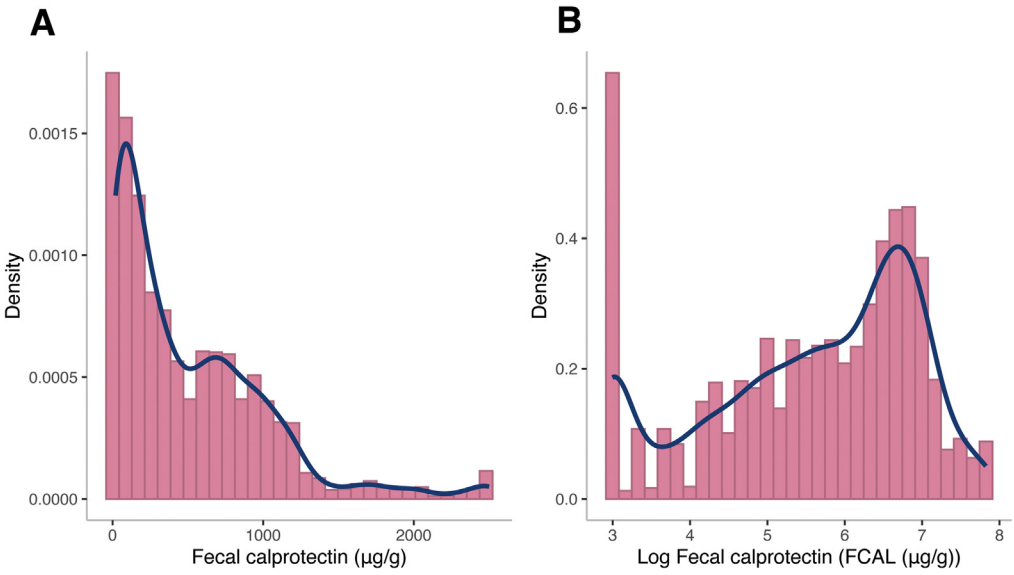
Catalina A. Vallejos (Conceptualization: Lead; Formal analysis: Supporting; Funding acquisition: Lead; Methodology: Equal; Supervision: Lead; Validation: Supporting; Writing – review & editing: Lead)

Conflicts of Interest

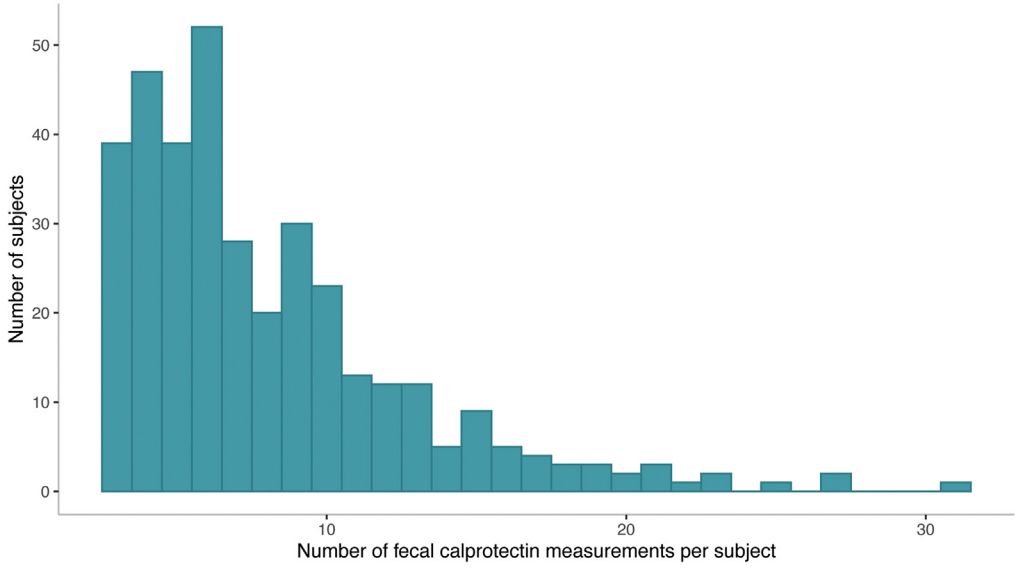
These authors disclose the following: Nikolas Plevris has received consultancy fees from Takeda; and speaker fees and/or travel support from AbbVie, Takeda, and Norgine. Lauranne A.A.P. Derikx has received consultancy fees from Sandoz; and speaking fees from Janssen. Beatriz Gros has received consultancy fees from AbbVie. Gareth-Rhys Jones has received speaker fees from AbbVie, Takeda, Pfizer, Ferring, and Janssen. Charlie W. Lees has received research support from AbbVie and Gilead; consultancy fees from AbbVie, Pfizer, Janssen, Gilead, Celltrion, Pharmacosmos, Takeda, Vifor, Iterative Scopes, Trellus Health, Galapagos, Vifor Pharma, Bristol Meyers Squibb, Boehringer Ingelheim, Sandoz, Novartis, Fresnius Kabi, and Tillotts; and speaker fees and/or travel support from Janssen, AbbVie, Pfizer, Dr. Falk Pharma, Ferring, Hospira, GSK, and Takeda. The remaining authors disclose no conflicts.

Funding

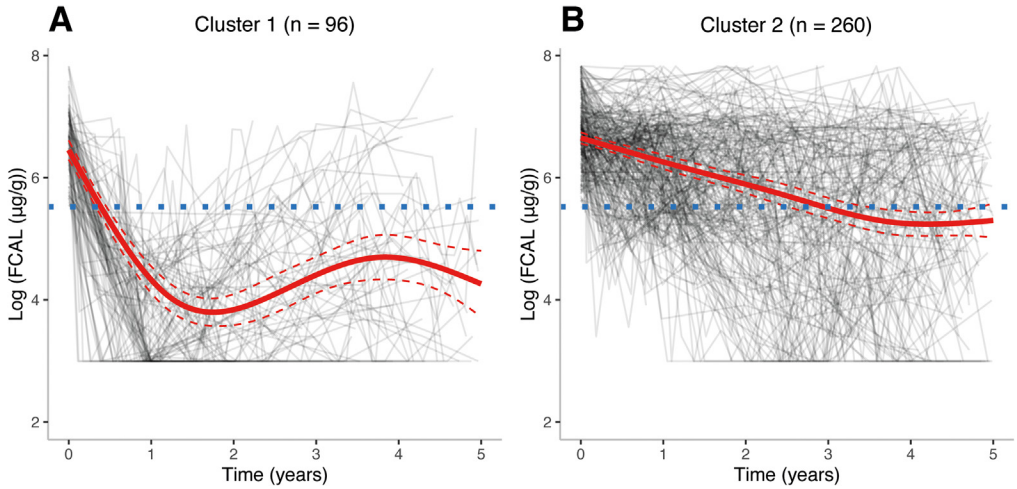
This work was supported by the Medical Research Council and University of Edinburgh Precision Medicine PhD studentship (MR/N013166/1, to Nathan Constantine-Cooke) and the UKRI Future Leaders Fellowship (MR/S034919/1, to Charlie W. Lees). Catalina A. Vallejos was funded by a Chancellor's Fellowship provided by the University of Edinburgh. Karla Monterrubio-Gómez was supported by an MRC University Unit grant to the MRC Human Genetics Unit. Gareth-Rhys Jones is supported by a Wellcome Trust Clinical Research Career Development Fellowship.



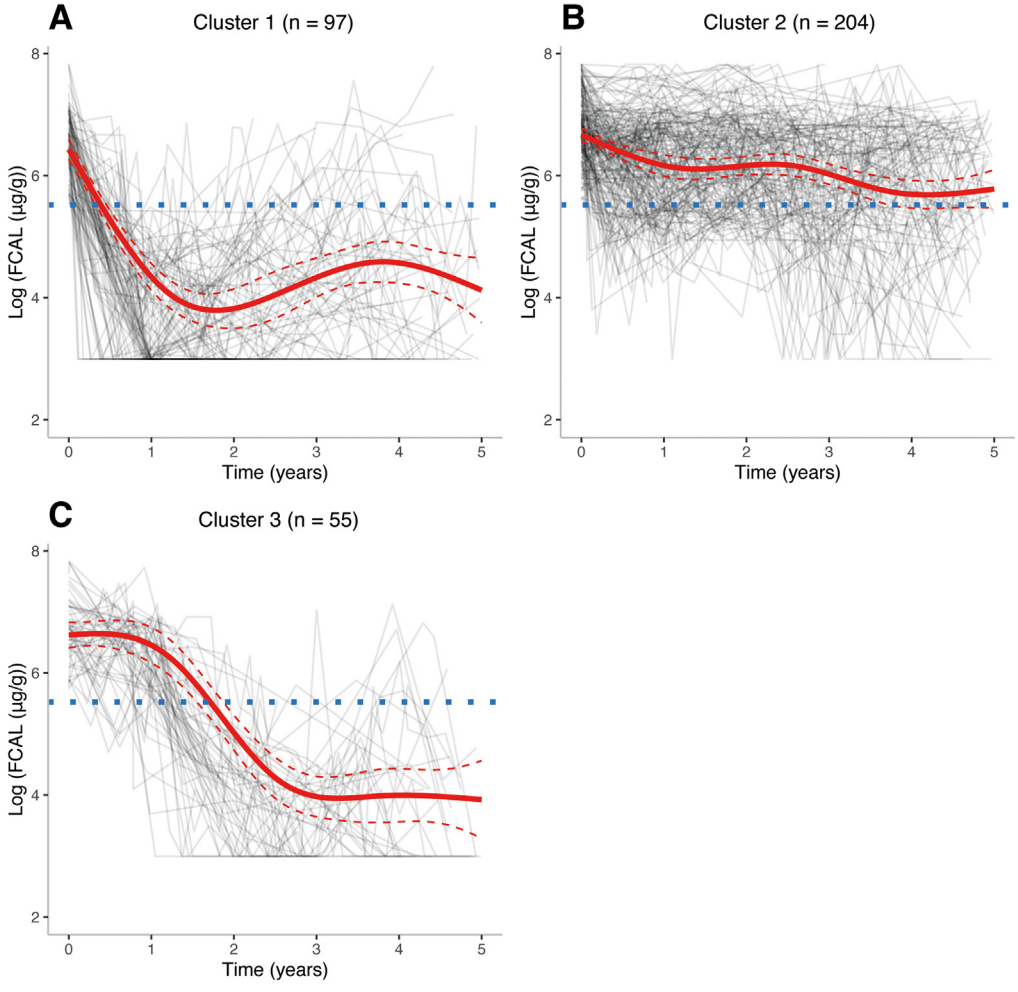
Supplementary Figure 1. Distribution of fecal calprotectin (FCAL) in (*A*) natural scale and (*B*) log scale.



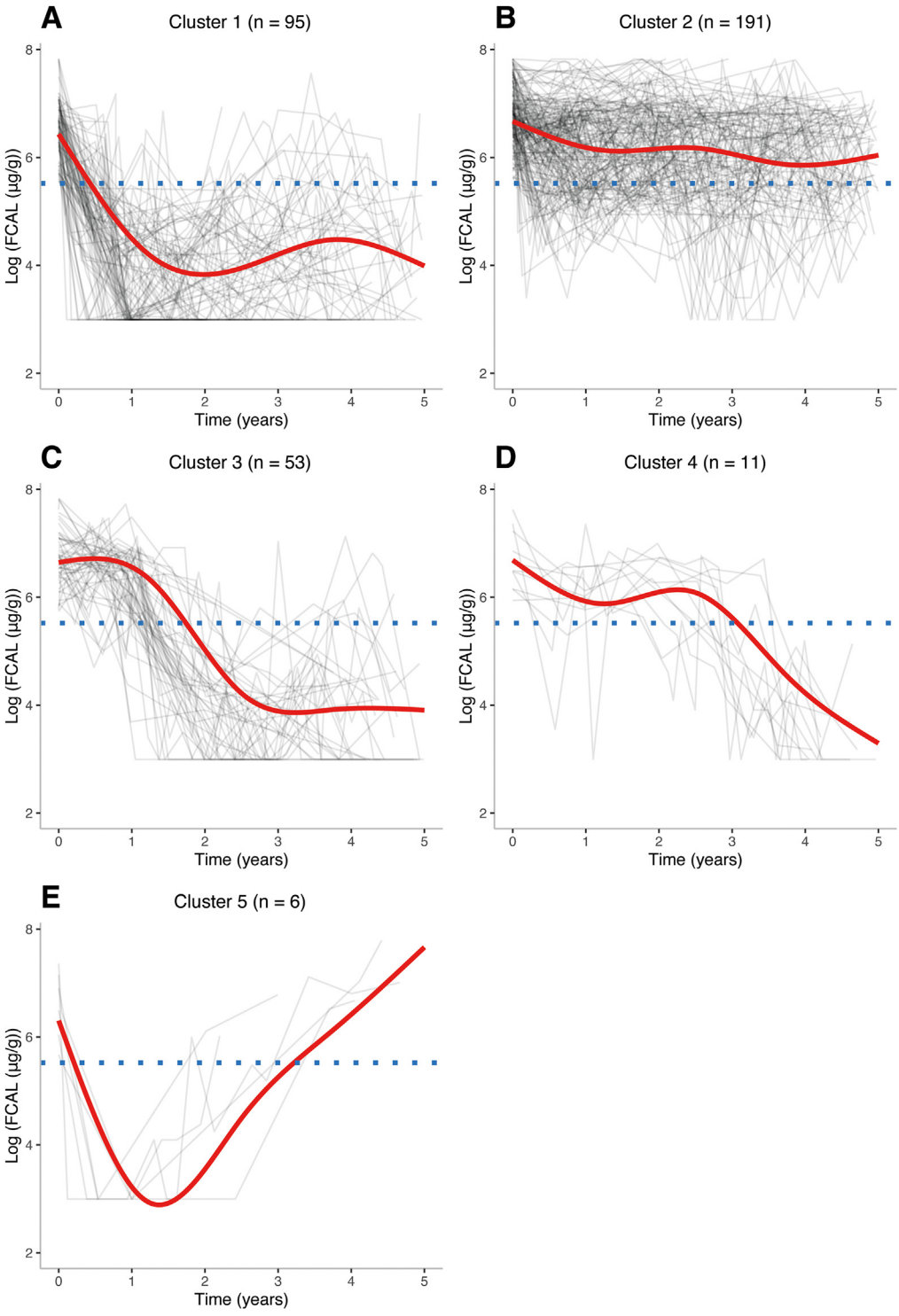
Supplementary Figure 2. Distribution of number of FCAL measurements per subject within 5 years of diagnosis.



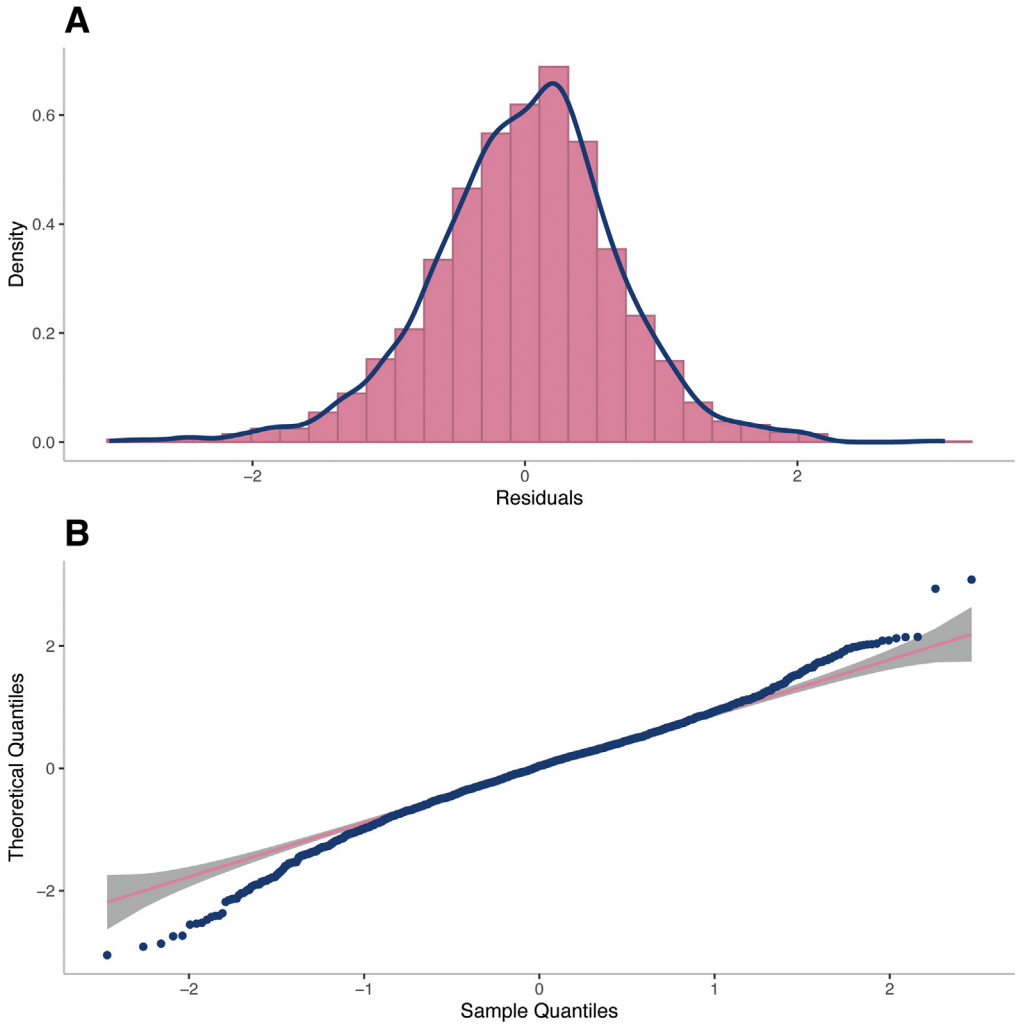
Supplementary Figure 3. Assuming 2 clusters, log-transformed subject-specific 5-year FCAL profiles for the study cohort for (*A*) cluster 1 and (*B*) cluster 2. The red solid line represents the predicted mean trajectory for each group, while the red dotted lines represent 95% confidence intervals. The gray lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 μg/g), the commonly accepted threshold for biochemical remission in Crohn's disease.



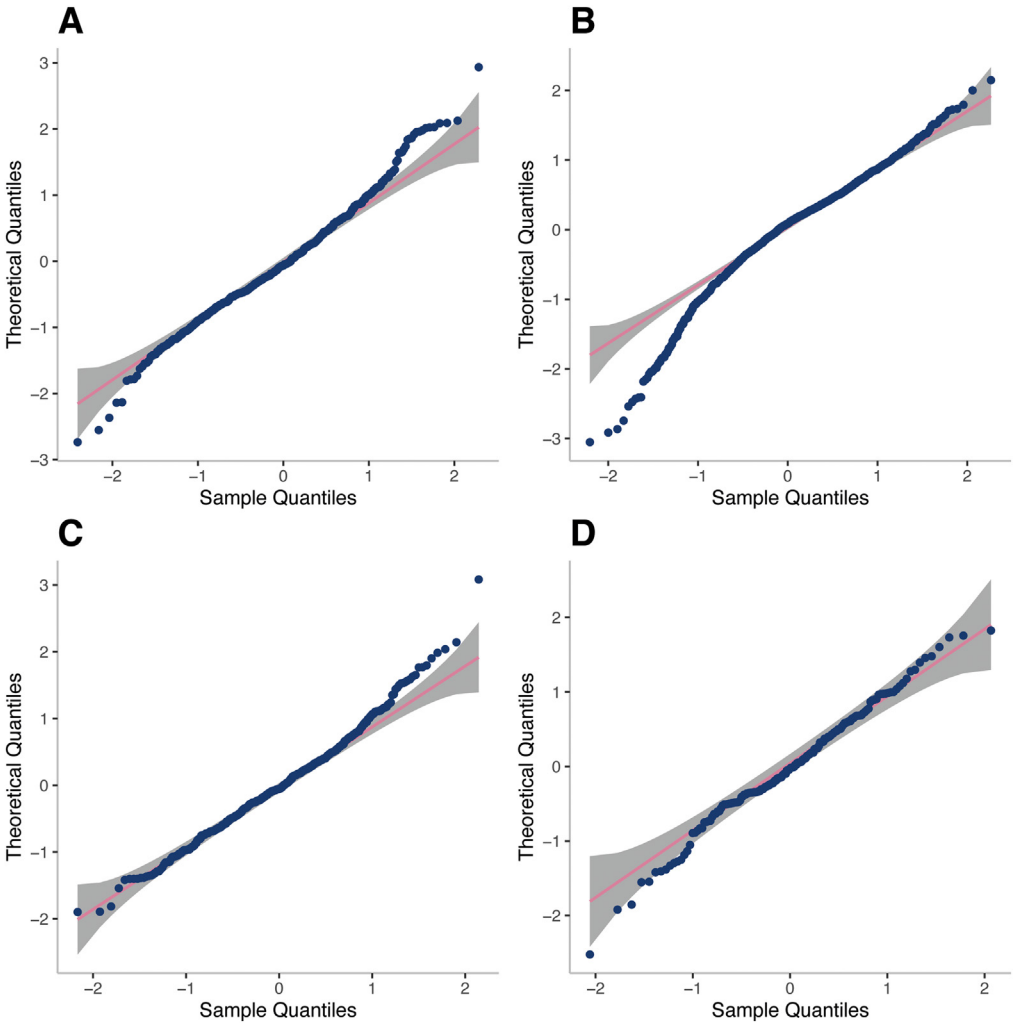
Supplementary Figure 4. Assuming 3 clusters, log-transformed subject-specific 5-year FCAL profiles for the study cohort for (*A*) cluster 1, (*B*) cluster 2, and (*C*) cluster 3. The red solid line represents the predicted mean trajectory for each group, while the red dotted lines represent 95% confidence intervals. The gray lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 μg/g), the commonly accepted threshold for biochemical remission in Crohn's disease.



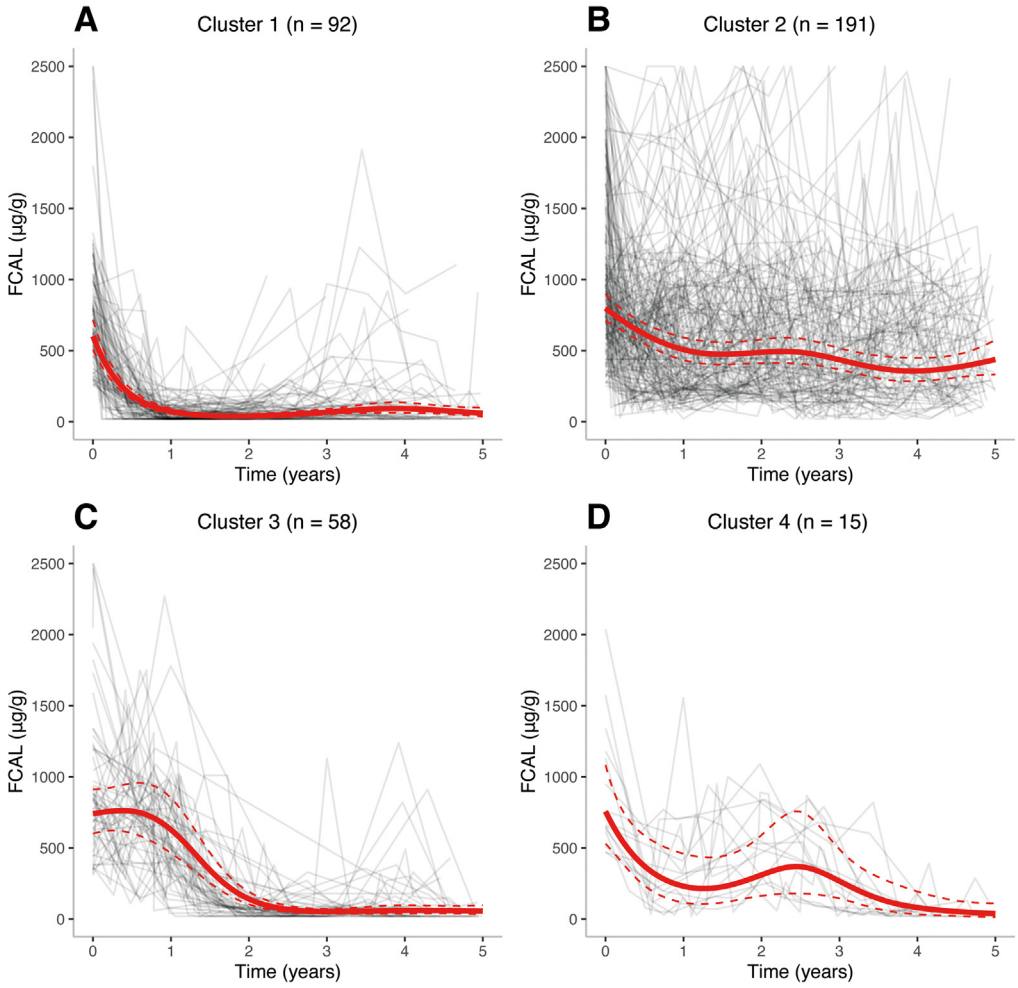
Supplementary Figure 5. Assuming 5 clusters, log-transformed subject-specific 5-year FCAL profiles for the study cohort for (*A*) cluster 1, (*B*) cluster 2, (*C*) cluster 3, (*D*) cluster 4, and (*E*) cluster 5. The red solid line represents the predicted mean trajectory for each group, while the red dotted lines represent 95% confidence intervals. The gray lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 μg/g), the commonly accepted threshold for biochemical remission in Crohn's disease.



Supplementary Figure 6. (*A*) Density plot of residuals for the 4-cluster model. (*B*) Quantile-quantile plot of residuals for the 4-cluster model.



Supplementary Figure 7. Quantile-quantile plots of residuals for the 4-cluster latent class mixed model stratified by (*A*) cluster 1, (*B*) cluster 2, (*C*) cluster 3, and (*D*) cluster 4.



Supplementary Figure 8. Five-year mean FCAL trajectories for clusters obtained by fitting latent class mixed models for (*A*) cluster 1, (*B*) cluster 2, (*C*) cluster 3, and (*D*) cluster 4. The red solid line represents the predicted mean trajectory for each group, while the red dotted lines represent 95% confidence intervals. The gray lines indicate the trajectory of each subject.

Supplementary Table 1. Model Fit Statistics for Latent Class Models Fitted to the Fecal Calprotectin Data for Different Numbers of Latent Subgroups

G	Log Likelihood	AIC	BIC
1	−4030.5	8103	8184.4
2	−4005.3	8064.7	8169.3[a]
3	−3990.5	8064.9	8174.8
4	−3983.5	8045	8196.1
5	−3970.2	8030.3[a]	8204.7
6	−3968.4[a]	8038.8	8236.5

AIC, Akaike information criterion; BIC, Bayesian information criterion; G, number of assumed clusters.
[a]Optimal found value.