Latent Crohn's Disease Subgroups are Identified by Longitudinal Faecal Calprotectin Profiles

PREPRINT

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 Marioni [2], Charlie W. Lees [† 2, 3], and Catalina A. Vallejos [† 1, 6]

[†] Shared senior authorship

[1] MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[2] Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[3] Edinburgh IBD Unit, Western General Hospital, Edinburgh, UK

[4] Inflammatory Bowel Disease Center, Radboud University Medical Center, Nijmegen, The Netherlands

[5] Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK

[6] The Alan Turing Institute, British Library, London, UK

August 16, 2022

ABSTRACT

Background: High faecal calprotectin is associated with poor outcomes in Crohn's disease. Monitoring of faecal calprotectin trajectories could characterise disease progression before severe complications occur.

Aims: We undertook an unbiased assessment of a retrospective incident Crohn's disease cohort to assess for inter-individual variability in faecal calprotectin levels over time. We aimed to explore whether latent classes of such profiles are associated with a composite endpoint consisting of surgery, hospitalisation, or Montreal behaviour progression and other clinical information.

Methods: Latent class mixed models were used to model faecal calprotectin trajectories within five years of diagnosis. Akaike information criterion, Bayesian information criterion, alluvial plots, and class-specific trajectories were used to decide the optimal number of classes. Log-rank tests of Kaplan-Meier estimators were used to test for associations between class membership and outcomes.

Results: Our study cohort comprised 365 subjects and 2856 faecal calprotectin measurements (median 7 per subject). Four latent classes were found and broadly described as a class with consistently high faecal calprotectin and three classes characterised by downward trends for calprotectin. Class membership was significantly associated with the composite endpoint, and separately, hospitalisation and Montreal disease progression, but not surgery. Early biologic therapy was strongly associated with class membership.

Conclusions: Our analysis provides a novel stratification approach for Crohn's disease patients based on faecal calprotectin trajectories. Characterising this heterogeneity helps to better understand different patterns of disease progression and to identify those with a higher risk of worse outcomes. Ultimately, this information will assist the design of more targeted interventions.

Keywords ► Crohn's disease · Outcomes research · Statistics

1 Introduction

Crohn's disease (CD) affects around 1 in 350 people in the UK and is a progressive disease [1,2]. CD patients typically present with discontinuous inflammation in the gastrointestinal tract and for many subjects the condition evolves to stricturing and fistulising behaviour. These latter phenotypes typically result in hospitalisation and surgery. However not all patients will progress, with approximately 30% following a quiescent disease course [3]. A significant unresolved question in the clinic is the prediction of future disease behaviour early in the disease course. This is an important problem: early introduction of biologic therapy will limit disease progression, at least in part, but will not be required for all patients.

Faecal calprotectin (FCAL) is routinely used to monitor mucosal inflammation and guide treatment decisions [4,5]. It has been shown that elevated FCAL (> 115µg/g) is associated with progression to complicated CD clinical manifestations including strictures and fistulas which necessitate surgery [6]. Conversely, normalising FCAL (< 250µg/g) within a year of diagnosis reduces the risk of disease progression, indicating the benefit of a treat-to-target approach based upon FCAL values [7].

Longitudinal FCAL monitoring could aid the prediction of disease progression by providing a more granular representation of a patient's condition before severe complications occur. The ability to access linked electronic health records for heterogeneous patient populations opens numerous opportunities to characterise FCAL trajectories throughout the course of the disease as well as to connect this information with other clinically relevant data such as age, sex, lifestyle factors, disease phenotypes and treatment.

Previous studies of FCAL trajectories have only captured heterogeneity across patients through a priori selected covariates by comparing subjects in endoscopic or clinical remission and those who have relapsed [8,9]. Here, we hypothesise that an unsupervised analysis to uncover latent patient subgroups with distinct longitudinal FCAL patterns can lead to better patient stratification and ultimately enable those who are more at risk of poor outcomes being prescribed biologic therapies sooner.

2 Materials and Methods

2.1 Study Design

We performed a retrospective cohort study at the Edinburgh IBD Unit, a tertiary referral centre, to determine if there were subgroups within the CD patient population identifiable from FCAL measurements collected within five years of diagnosis. For this purpose, we modelled longitudinal FCAL profiles using latent class mixed models (LCMMs) [10], an extension of linear mixed effects models which enables the identification of distinct subgroups with shared longitudinal patterns. LCMMs have been used to model biomarker trajectories in many contexts (e.g. modelling disease activity score in rheumatoid arthritis [11] and estimated glomerular filtration rate in type 2 diabetes [12]).

The data were obtained from a retrospective cohort study by Plevris et al. which identified all incident CD cases between 2005 and 2017 at The Edinburgh IBD unit [7]. 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). Information about drug treatments and disease location was also extracted.

2.2 Outcomes

For the time-to-event analysis, the primary endpoint was a composite outcome given by disease progression (defined as a progression in Montreal disease behaviour: B1 to B2/B3 or B2 to B3 or new perianal disease), resectional surgery for CD, or CD-related hospitalisation. These outcomes (disease progression, surgery, or hospitalisation) were also considered separately as secondary endpoints. Dates for disease progression were defined as occurring when the first investigation which found the change was performed, such as an MRI or colonoscopy. Examination under anaesthetic or incision and drainage of a perianal abscess were classified as hospitalisations.

2.3 Criteria & Definitions

First, the inclusion criteria from Plevris et al. were applied: (1) CD diagnosis between 2005 and 2017; (2) an initial FCAL measurement at diagnosis (or within 2 months) and prior to treatment; (3) initial FCAL result $\geq 250\mu g/g$; (4) an accurate date of diagnosis; (5) at least one additional FCAL measurement within 12 months of diagnosis; (6) at least 12 months of followup; (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: (8) at least 3 FCAL measurements within 5 years of diagnosis.

The following information was available at diagnosis: sex, age, smoking status, FCAL, Montreal location (alongside upper gastrointestinal inflammation), and Montreal behaviour (alongside perianal disease). Treatments prescribed within one year of diagnosis were also recorded: 5-ASAs (aminosalicylates), thiopurines, corticosteroids, methotrexate, exclusive enteral nutrition, and biologic therapies (either infliximab, adalimumab, ustekinumab or vedolizumab).

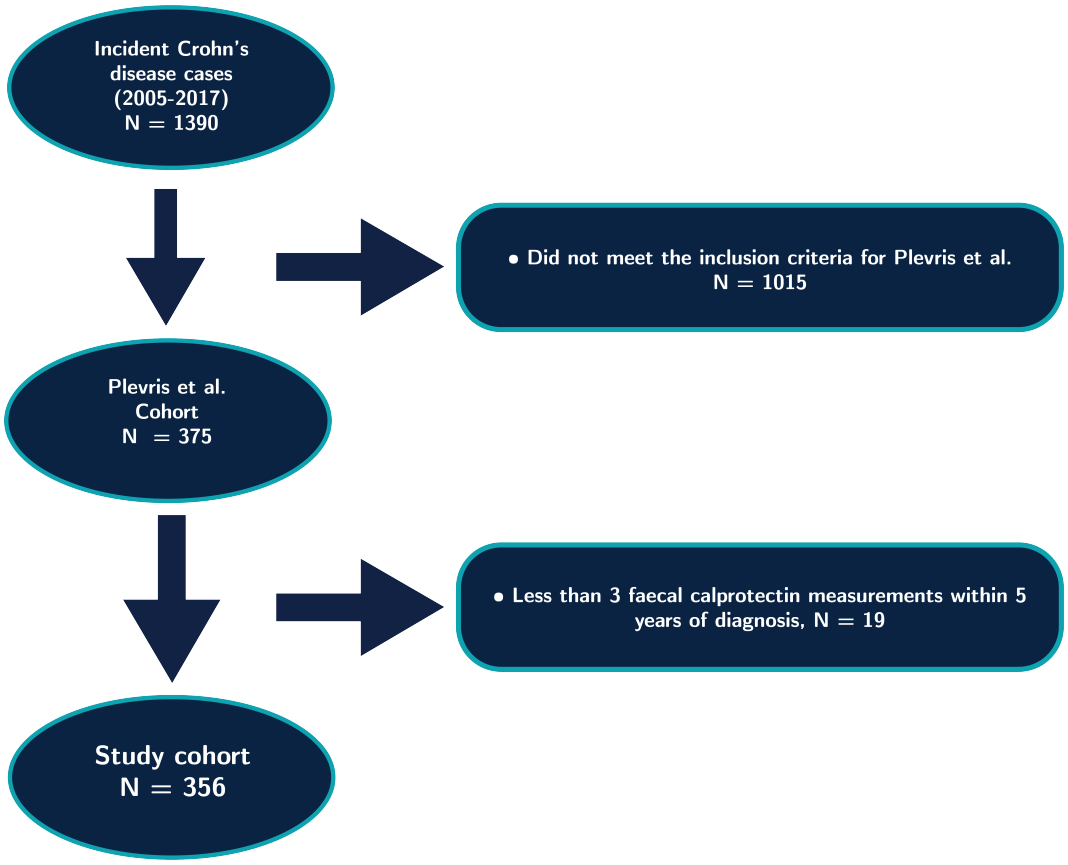

Figure 1: Flowchart demonstrating data processing steps. FCAL: faecal calprotectin.

2.4 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[7]. Patients are also given collection kits in the clinic or sent by post to their home. Samples are stored at -20°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.

Figure 2: Alluvial plot demonstrating how membership of latent classes obtained from the faecal calprotectin profiles of Crohn's disease patients changes as the assumed number of classes increase. The height of each band indicates the size of each class.

2.5 Statistical Analysis

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

FCAL measurements greater than $2500\mu g/g$ were set to 2500 $\mu g/g$, the upper range for the assay. Likewise, measurements reported as less than the lower range for the assay, $20\mu g/g$, were set to $20\mu g/g$. The FCAL data were log-transformed before the models were fitted. To model the FCAL trajectories and find latent classes, we used LCMMs with longitudinal patterns captured using natural cubic splines[10]. Natural cubic splines provide a flexible framework to model FCAL trajectories whilst remaining stable at either end of the study followup period[13]. Using natural cubic splines results in fewer parameters needing to be estimated compared to polynomial regression which requires a high-degree polynomial to achieve the same level of flexibility[14]. Between two and five 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. The knots were placed at the first quartile, median, and third quartile of all FCAL measurement times. A full model description is provided as an Appendix.

LCMMs assuming 2, 3, 4, 5, or 6 latent classes were fitted. For each of these models, 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 were deemed to converge based on parameter and likelihood stability, and on the negativity of the second derivatives. After each optimal model was found, the maximum log-likelihood, AIC, and Bayesian information criterion (BIC) were calculated. An alluvial plot was produced to provide intuition of how additional classes are formed as the number of assumed classes increases. These findings were used to decide on the appropriate number of latent classes in our study population. Uncertainty in class assignments was quantified using posterior classification probabilities. To visualise overall trajectories within each class, point estimates for each of the model parameters were used and statistical uncertainty was visualised using 95% confidence intervals.

For crossing Kaplan-Meier curves, we considered the Peto-Peto test[15] to assess association between class membership and the specified endpoints. The latter test is more resilient to violations of the proportional hazards assumption than log-rank tests. Marginal associations between class membership and information available at the time of diagnosis were also explored. Chi-square tests and Fisher's exact tests, dependent on suitability, were used for categorical variables. ANOVA was used for continuous variables. Upper gastrointestinal inflammation (L4) and perianal disease (P) were tested separately to Montreal location (L1-L3) and Montreal behaviour (B1-B3) respectively.

Potential evidence of treatment effects was garnered by testing for associations between class membership and whether each treatment was prescribed within one year of diagnosis using Fisher's exact test. Biologic prescriptions within three 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 (p_{adj}).

As an exploratory analysis, a multinomial logistic regression model[16] and a random forest classifier[17] were used to predict latent class membership 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 (AUC)[19].

R[20] (v.4.2.0) was used for all statistical analyses using the `lcmm`[21] (v.1.9.5), `survival`[22] (v.3.3-1), `survminer`[23] (v.0.4.9), `nnet`[24] (v.7.3-17), `ranger`[25] (v.0.13.1), `datefixR`[26] (v.0.1.4), `tidyverse`[27] (v.1.3.1), `tidymodels`[28] (v.0.2.0), `vip`[29] (v0.3.2) and `ggalluvial`[30] (v.0.12.3) R libraries. The analytical reports generated for this study and corresponding source code are hosted online*.

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

3 Results

3.1 FCAL Measurements

356 subjects with incident CD met the inclusion criteria for this study (Figure 1, Table 1). Across all patients, 2856 FCAL measurements were recorded within five years of diagnosis. The median frequency of FCAL measurements for a subject within this period was 7 (IQR 5-10). The overall distribution is presented in Figure S1.

3.2 Modelling FCAL Trajectories

LCMMs fitted with 2, 3, 4, 5, or 6 assumed latent classes all converged as per default convergence criteria. As seen in Figure 2, class assignments were largely stable across differing assumed latent classes, particularly when comparing the 3-class, 4-class and 5-class models. Performance metrics for each model considered are provided in Table S1. BIC suggested the 2-class model was most appropriate, but this model was discarded as visual inspection of the inferred trajectories suggested a larger number of distinct latent classes (Figure S2). AIC and the maximum log-likelihood favoured the 5-class and 6-class models, respectively. However, those models were found to overfit the data as some of the inferred trajectories were similar (Figure S4 and Figure S5). Therefore, as a parsimonious choice, we selected the 4-class model.

Figure 3 presents the log mean profiles for 4-class model alongside subject-specific observed FCAL trajectories. The model identified 3 main groups of patients: classes 1, 2 and 3 (92, 191, and 58 subjects, respectively) and a small class 4 with 15 subjects. Classes 1 and 3 display similar profiles — both showing a sharp decrease in FCAL which then remains low. However, class 1 is differentiated by the decrease occurring immediately after diagnosis, whilst this decrease does not occur until around a year after diagnosis for class 3. In contrast, class 2 is characterised by a mean profile which remains consistently high: never dropping below the $250\mu g/g$ clinical threshold for disease activity. Finally, the mean profile for class 4 exhibits an initial decrease, but this is not sustained during the first 3 years.

*https://vallejosgroup.github.io/lcmm-site/

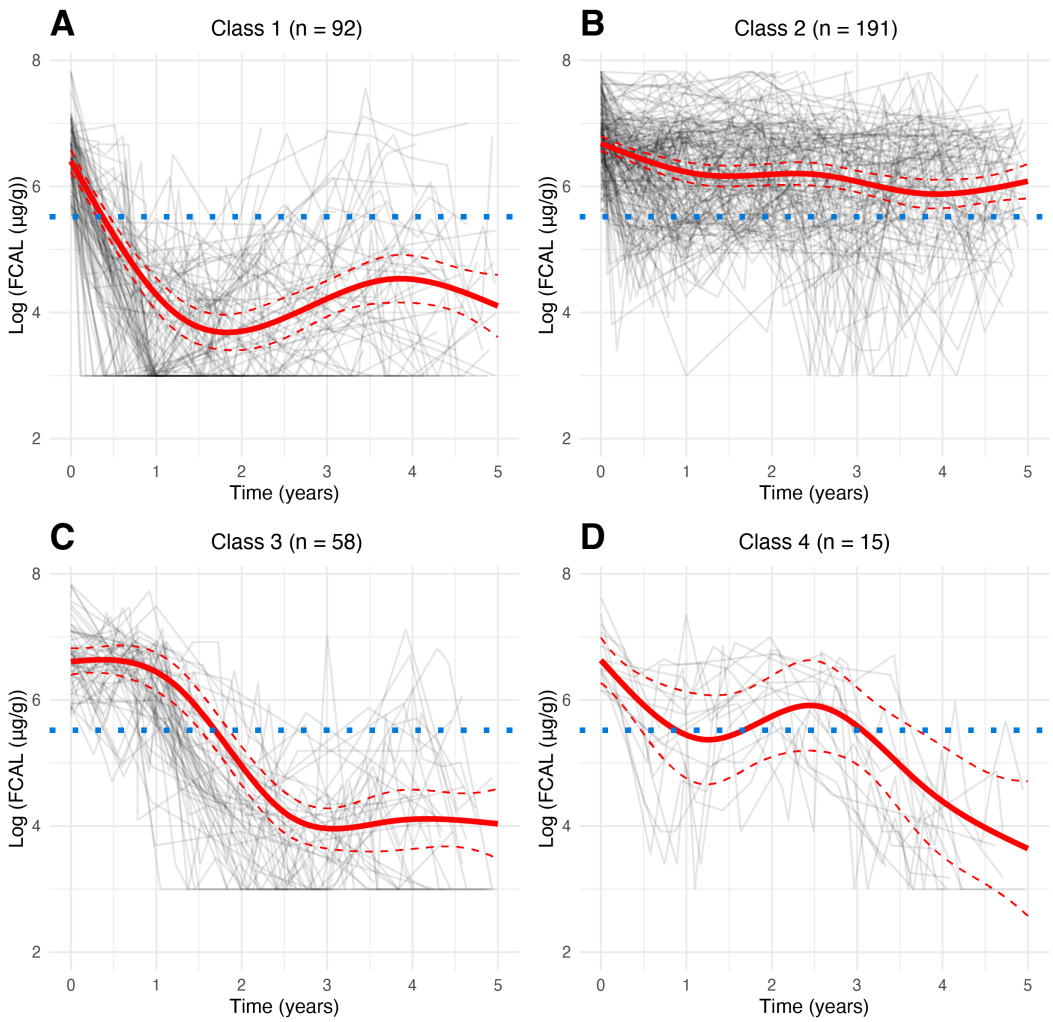

Figure 3: Log-transformed subject-specific five-year faecal calprotectin profiles for the study cohort for **A**, class 1; **B**, class 2; **C**, class 3; **D**, class 4. The red solid line represents the predicted mean trajectory for each group, whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 $\mu g/g$): the commonly accepted threshold for biochemical remission in Crohn's disease. See Figure S6 for the fits in the original measurement scale.

3.3 Association with Outcomes

Figure 4 presents Kaplan-Meier curves for the primary and secondary endpoints, stratified by class assignment. The primary composite endpoint, alongside hospitalisation and disease progression, was found to be significant by Peto-Peto test ($p < 0.001$; $p_{\text{adj}} = 0.002$). However, surgery was not found to be significantly associated with class membership. For all four outcomes, class 2 resulted in the poorest survival distributions across the classes, whilst class 1 typically resulted in the best survival probabilities.

		Latent classes					
	Population	Class 1	Class 2	Class 3	Class 4	p	p_{adj}
Sex						0.157	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						0.851	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						0.015*	0.116
Smoker	70 (20%)	22 (24%)	43 (23%)	4 (7%)	1 (7%)		
Non-smoker	286 (80%)	70 (76%)	148 (77%)	54 (93%)	14 (93%)		
Diagnostic FCAL ($\mu g/g$)						0.131	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 behaviour						0.494	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)†						0.776	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						0.125	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%)		
Ilealcolonic (L3)	121 (34%)	31 (34%)	65 (34%)	19 (33%)	6 (40%)		
Upper gastrointestinal (L4)‡						< 0.001 ***	0.002 **
Present	84 (24%)	8 (9%)	51 (27%)	20 (34%)	5 (33%)		
Not present	272 (76%)	84 (91%)	140 (73%)	38 (66%)	10 (67%)		
5-ASAs (aminosalicylates)						0.326	1
Yes	76 (21%)	15 (16%)	48 (25%)	11 (19%)	2 (13%)		
No	280 (79%)	77 (84%)	143 (75%)	47 (81%)	13 (87%)		
Thiopurine						0.023*	0.161
Yes	250 (70%)	62 (67%)	127 (66%)	50 (86%)	11 (73%)		
No	106 (30%)	30 (33%)	64 (34%)	8 (14%)	4 (27%)		
Corticosteroids						0.983	1
Yes	298 (84%)	78 (85%)	159 (83%)	48 (83%)	13 (87%)		
No	58 (16%)	14 (15%)	32 (17%)	10 (17%)	2 (13%)		
Methotrexate						0.139	0.975
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						0.779	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 months						< 0.001 ***	0.004 **
Yes	44 (12%)	23 (25%)	14 (7%)	5 (9%)	2 (13%)		
No	312 (88%)	69 (75%)	177 (93%)	53 (91%)	13 (87%)		
Biologic						< 0.001 ***	< 0.001 ***
Yes	94 (26%)	42 (46%)	35 (18%)	12 (21%)	5 (33%)		
No	262 (74%)	50 (54%)	156 (82%)	46 (79%)	10 (67%)		

Table 1: Cohort characteristics and treatments prescribed to the cohort. All prescriptions were prescribed within one year of diagnosis unless otherwise stated. Percentages when stratified across latent classes are out of the total number of subjects in the latent class. Biologic is defined as either infliximab, adalimumab, ustekinumab or vedolizumab prescription. † Perianal disease may be present concomitantly to B1, B2 or B3 disease behaviour or separately. ‡ Upper gastrointestinal inflammation may be present in addition to ileal, colonic, or ilealcolonic inflammation. p unadjusted p-value. p_{adj} p-value after Bonferroni correction. * Significant at a 5% significance level. ** Significant at a 1% significance level. *** Significant at a 0.1% significance level.

3.4 Association with Variables Available at Diagnosis

Out of the eight variables typically available at diagnosis we tested for association with class membership, two variables were found to be significant at the 5% significance level before applying a Bonferroni adjustment: smoking status ($p = 0.01; p_{\text{adj}} = 0.08$) and the presence of upper gastrointestinal inflammation ($p < 0.001; p_{\text{adj}} = 0.002$). 24% and 23% of class 1 and class 2 respectively were smokers when they were diagnosed, whereas only 7% of

class 3 and class 4 smoked during this period. Only 9% of class 1 had upper gastrointestinal involvement at diagnosis in comparison to the 27%, 34%, and 33% in class 2, class 3, and class 4 respectively.

3.5 Association with Treatments

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

Out of the prescriptions considered, being prescribed a thiopurine within one year of diagnosis ($p = 0.023$; $p_{\text{adj}} = 0.16$) and being prescribed a biologic either within three months ($p < 0.001$; $p_{\text{adj}} = 0.004$) or one year of diagnosis ($p < 0.001$; $p_{\text{adj}} < 0.0001$) were found to be significant before Bonferroni adjustment. However, class membership could not be predicted from demographic data and biologic prescriptions (AUC of 0.68 for the multinomial logistic regression model and 0.66 for the random forest classifier).

4 Discussion

In this study, four latent subgroups 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 FCAL data to identify latent subgroups, although others have applied linear mixed models to FCAL data[8,9] or have applied LCMMs to other biomarkers[31,32]. We have demonstrated membership of these classes is associated with disease outcomes: namely the composite endpoint, hospitalisation, and Montreal behaviour progression. This builds upon work by Plevris et al. by considering FCAL as a continuous process instead of only a diagnostic and a followup measurement. Of note, Plevris et al. found normalisation of FCAL ($< 250\mu g/g$) to be associated with surgery in addition to the other outcomes found significant here. However, surgery was not found to be associated with class membership in this study.

We observed class membership to be associated with early biologic treatment. It is arguably reasonable to expect an association between biologic treatments and FCAL profiles given the often-reported association between FCAL and endoscopic activity and an association between biologic treatments and endoscopic healing for CD patients[33,34].

We found smoking status and upper gastrointestinal involvement to be associated with class membership. A comparatively high number of subjects who smoked at diagnosis were found in both class 1 and class 2 despite class 1 being characterised by an overall decrease in FCAL and class 2 being characterised 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[35] in addition to being associated with undergoing surgery and disease progression[36]. Upper gastrointestinal involvement is likely a proxy for a more severe CD sub-phenotype.

Some similarities can be observed between the clinically derived profiles in the IBSEN cohort[3] patterns and the class-specific mean profiles uncovered in this study. Both studies identified a large group of patients that exhibit a decline in severity of symptoms (class 1 and class 3 in our study) and a group with chronic continuous symptoms (class 2 in our study). However, the IBSEN study identified a group with increasing intensity of symptoms which 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[37]. Moreover, the IBSEN study findings were gathered before the widespread emergence of biologic therapies for CD and may not represent more modern trends which also may not be well known a priori. IBSEN participants also had to choose within a pre-specified set of patterns. Instead, our unsupervised analysis is able to infer subgroup profiles in a data-driven manner.

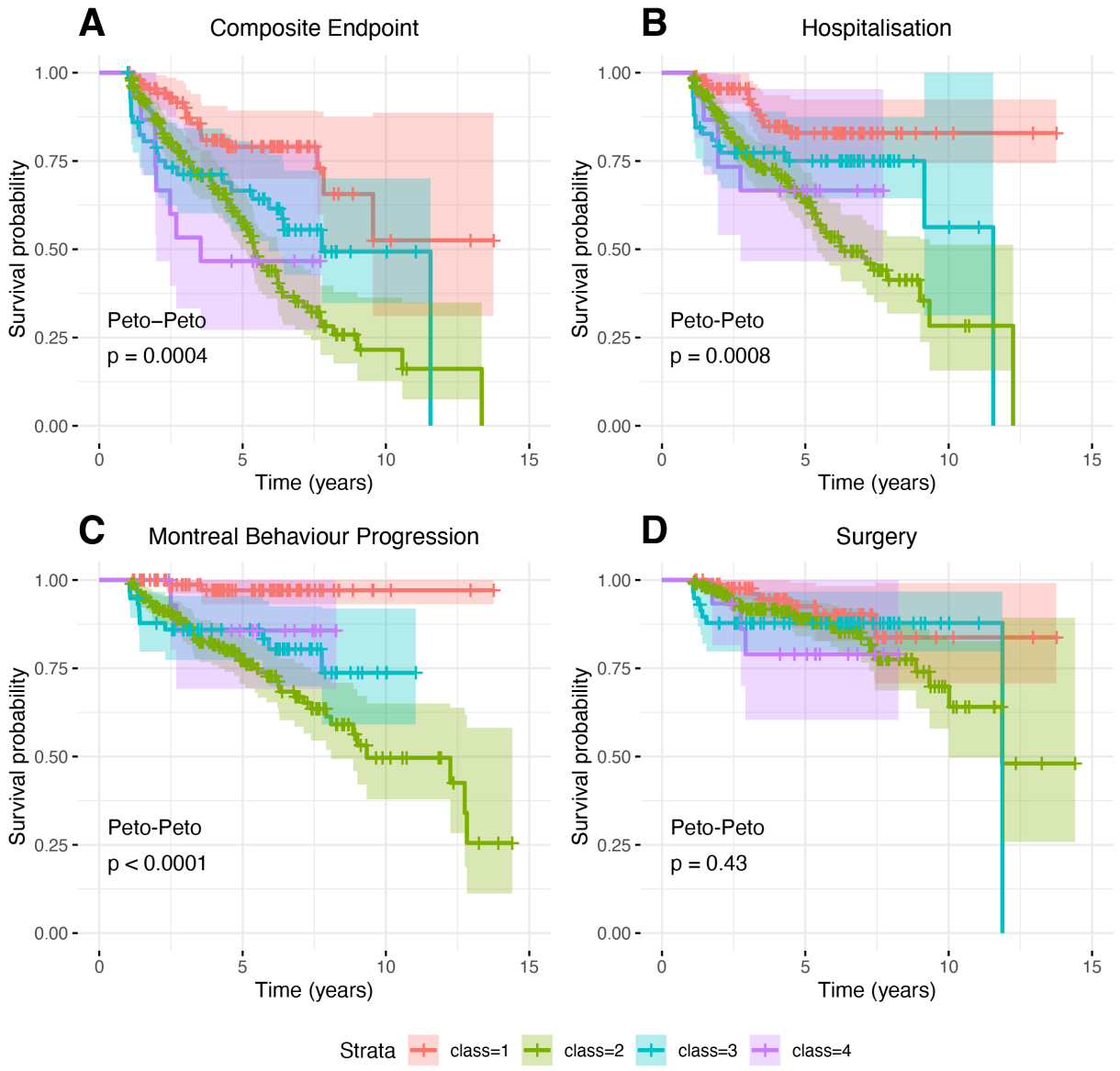

Figure 4: Kaplan-Meier curves stratified by latent class for the following endpoints: **A**, a composite endpoint consisting of hospitalisation, Montreal behaviour progression or surgery; **B**, hospitalisation; **C**, Montreal behaviour progression; **D**, surgery. Peto-Peto: test for the comparison of survival curves.

Many previous studies incorporating FCAL observations are reliant on fixed thresholds to dichotomise FCAL at a limited number of time-points, such as diagnosis and a single subsequent observation. Instead, longitudinal monitoring of continuous FCAL trajectories can provide a more granular representation of disease progression before severe complications occur. This information has been shown to be useful for prognostic purposes. For example, modelling FCAL as a continuous process has been used to evaluate the effect of infliximab induction therapy on FCAL for ulcerative colitis patients[8]. A similar approach has been used to explore the influence of lifestyle factors, disease activity variables, and treatments on the FCAL measurements of inflammatory bowel disease patients in remission across two years or until clinical relapse[9]. In addition to the relatively small sample sizes of these previous studies (53 and 104 subjects respectively), they only captured heterogeneity of FCAL trajectories across patients through a priori selected covariates. Our approach does not limit sources of heterogeneity to only specified covariates and considers a much larger cohort of 356 subjects.

In this study, eight potential associations with variables typically available at diagnosis, four potential associations with disease endpoints, and a further seven potential associations with treatments have been explored. As such, we potentially invite criticism due to multiple testing. Indeed, some associations reported here (e.g. between class membership and smoking) fail to be significant after applying Bonferroni corrections. However, we believe 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. In particular, quantifying causal treatment effects from such observational data is an active area of research and such analysis is beyond the scope of this study[38,39]. The data gathering process is observational and whilst FCAL is collected routinely at all clinical interactions, subjects with more complicated disease are still likely to have more measurements available. The retrospective study design also means all subjects did not have the same treatment options at the same stage in their disease trajectories, as subjects may have been diagnosed any time between 2005 and 2017. However, the date of diagnosis, converted to the number of days the subject was diagnosed after 01/01/2001, was considered for potential association with class membership and no significant association was found ($p = 0.12$). We also acknowledge the potential for inclusion bias in this study. The study by Plevris et al. required subjects to have an FCAL of at least $250\mu g/g$ at diagnosis and excluded subjects which met one of the endpoints within a year of diagnosis. The former potentially excludes subjects with milder disease, whilst the latter potentially excludes subjects with more aggressive disease.

Our unsupervised analysis provides a proof-of-concept for novel patient stratification approaches in CD that we envisage translating to the clinic as a dynamic risk prediction tool: updating an individuals' risk of disease progression in real-time as additional datapoints become available. Supporting this intention, LCMMs can be extended under a joint model framework to jointly model both longitudinal FCAL trajectories and time-to-event outcomes and make dynamic predictions. This tool could assist clinicians with making informed decisions. Similar software has been developed to predict outcomes of moderate COVID-19 patients from circulating calprotectin profiles[31]. The tool could also be used to encourage a patient to change modifiable lifestyle factors, such as smoking, by showing how changing these lifestyle factors changes their individual risk of experiencing poor disease outcomes. However, extensions to the model may ultimately lead to impractically high computational requirements as sample size and complexity of the model increases. Improving the scalability of joint models is an active and evolving area of research[40,41].

5 Conclusion

We have demonstrated the suitability and utility of latent class mixed modelling for identifying latent classes within the CD population based on FCAL profiles. These models are easily implemented and can assist in the monitoring of FCAL in CD. After we found and described four latent classes, we reported class membership to be significantly associated with disease outcomes. We believe our findings are an important first step towards risk prediction tools which incorporate longitudinal FCAL being available in clinical settings: informing both clinicians and patients. Before such a tool is used in clinics, however the observed treatment associations should be investigated via causal inference approaches.

6 *Authorship*

NC-C, **KM-G**, **NP**, **REM**, **CWL**, and **CV** contributed to the conception and study design for the manuscript. **NC-C**, **NP**, **LD**, **BG**, and **CWL** collected the data for this study. All authors except **REM** had access to the study data. **NC-C** performed all statistical analysis. **NC-C**, **BG**, and **KM-G** drafted the manuscript. All authors were involved with critical revision of the manuscript, and all authors reviewed and approved the final manuscript prior to submission.

7 *Data Availability*

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 **CWL**.

8 *Acknowledgements*

Declaration of funding interests: This work has been supported by the Medical Research Council & University of Edinburgh Precision Medicine PhD studentship (MR/N013166/1, to **NC-C**) and the UKRI Future Leaders Fellowship (MR/S034919/1, to **CWL**. **KM-G** was supported by an MRC University Unit grant to the MRC Human Genetics Unit. **GRJ** is supported by a Wellcome Trust Clinical Research Career Development Fellowship.

Declaration of personal interests: **NC-C**: none declared; **KM-G**: none declared; **NP** has received consultancy fees from Takeda, speaker fees and/or travel support from Abbvie, Takeda, Norgine; **LAAPD** has received consultancy fees from Sandoz, speaking fees from Janssen; **BG** has received consultancy fees from Abbvie; **GRJ** has received speaker fees from Abbvie, Takeda, Pfizer, Ferring and Janssen; **REM**: none declared; **CWL** 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, and Kabi Tillotts; speaker fees and/or travel support from Janssen, Abbvie, Pfizer, Dr Falk, Ferring, Hospira, GSK, and Takeda; **CAV**: none declared.

9 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 Gastroenterology* 2021; 12(6): 461–470. doi: 10.1136/flgastro-2019-101369

[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). *Scandinavian Journal of Gastroenterology* 2007; 42(5): 602-610. PMID: 17454881doi: 10.1080/00365520601076124

[4] D'Amico F, Nancey S, Danese S, Peyrin-Biroulet L. A practical guide for faecal calprotectin measurement: myths and realities. *Journal of Crohn's and Colitis* 2020; 15(1): 152–161. doi: 10.1093/ecco-jcc/jjaa093

[5] Plevris N, Lees CW. Disease Monitoring in Inflammatory Bowel Disease: Evolving Principles and Possibilities. *Gastroenterology* 2022; 162(5): 1456-1475.e1. doi: 10.1053/j.gastro.2022.01.024

[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. *Clinical Gastroenterology and Hepatology* 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. *Clinical Gastroenterology and Hepatology* 2021; 19(9): 1835–1844.e6. doi: 10.1016/j.cgh.2020.08.022

[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. *Journal of Crohn's and Colitis* 2012; 6(5): 557–562. doi: 10.1016/j.crohns.2011.11.002

[9] Zhulina Y, Cao Y, Amcoff K, Carlson M, Tysk C, Halfvarson J. The prognostic significance of faecal calprotectin in patients with inactive inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics* 2016; 44(5): 495–504. doi: 10.1111/apt.13731

[10] Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *Journal of Statistical Software* 2017; 78(2): 1–56. doi: 10.18637/jss.v078.i02

[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. doi: 10.1016/j.ebiom.2016.08.024

[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 International* 2019; 95(1): 178–187. doi: 10.1016/j.kint.2018.08.026

[13] 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 Medical Research Methodology* 2022; 22(1): 68. doi: 10.1186/s12874-022-01542-8

[14] James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning* ch. 7: 297-317; Springer Texts in Statistics. Springer US. 2nd ed. 2021

[15] Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society. Series A (General)* 1972; 135(2): 185. doi: 10.2307/2344317

[16] Kwak C, Clayton-Matthews A. Multinomial logistic regression. *Nursing Research* 2002; 51(6).

[17] Breiman L. Random forests. *Machine Learning* 2001; 45(1): 5–32. doi: 10.1023/A:1010933404324

[18] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*: 241-247; Springer Series in Statistics. Springer New York. 2nd ed. 2009

[19] Melo F. *Area under the ROC Curve*: 38–39; New York, NY: Springer New York . 2013

[20] R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2022.

[21] Proust-Lima C, Philipps V, Diakite A, Liquet B. lcmm: extended mixed models using latent classes and latent processes. https://cran.r-project.org/package=lcmm; 2021. R package version: 1.9.3.

[22] Therneau TM. A Package for Survival Analysis in R. https://CRAN.R-project.org/package=survival; 2021. R package version 3.2-13.

[23] Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. https://CRAN.R-project.org/package=survminer; 2021. R package version 0.4.9.

[24] Venables WN, Ripley BD. *Modern applied statistics with S*. New York: Springer. fourth ed. 2002. ISBN 0-387-95457-0.

[25] Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 2017; 77(1): 1–17. doi: 10.18637/jss.v077.i01

[26] Constantine-Cooke N. datefixR: fix really messy dates in R. https://CRAN.R-project.org/package=datefixR; 2022. R package version 0.1.4

[27] Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software* 2019; 4(43): 1686. doi: 10.21105/joss.01686

[28] Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.. https://CRAN.R-project.org/package=tidymodels; .

[29] Greenwell BM, Boehmke BC. Variable importance plots—An introduction to the vip package. *The R Journal* 2020; 12(1): 343–366. doi: 10.32614/RJ-2020-013

[30] Brunson JC. ggalluvial: layered grammar for alluvial plots. *Journal of Open Source Software* 2020; 5(49): 2017. doi: 10.21105/joss.02017

[31] Chapuis N, Ibrahimi N, Belmondo T, et al. Dynamics of circulating calprotectin accurately predict the outcome of moderate COVID-19 patients. *eBioMedicine* 2022; 80. doi: 10.1016/j.ebiom.2022.104077

[32] Vistisen D, Andersen GS, Hulman A, Persson F, Rossing P, Jørgensen ME. Progressive Decline in Estimated Glomerular Filtration Rate in Patients With Diabetes After Moderate Loss in Kidney Function—Even Without Albuminuria. *Diabetes Care* 2019; 42(10): 1886–1894. doi: 10.2337/dc19-0349

[33] Jusué V, Chaparro M, Gisbert JP. Accuracy of fecal calprotectin for the prediction of endoscopic activity in patients with inflammatory bowel disease. *Digestive and Liver Disease* 2018; 50(4): 353–359. doi: 10.1016/j.dld.2017.12.022

[34] Narula N, Wong EC, Dulai PS, Marshall JK, Jairath V, Reinisch W. Comparative effectiveness of biologics for endoscopic healing of the ileum and colon in Crohn's disease. *American Journal of Gastroenterology* 2022; Publish Ahead of Print. doi: 10.14309/ajg.0000000000001795

[35] 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. *The Lancet Gastroenterology & Hepatology* 2019; 4(5): 341–353. doi: 10.1016/s2468-1253(19)30012-3

[36] Lawrance IC, Murray K, Batman B, et al. Crohn's disease and smoking: Is it ever too late to quit?. *Journal of Crohn's and Colitis* 2013; 7(12): e665–e671. doi: 10.1016/j.crohns.2013.05.007

[37] Koutroumpakis E, Katsanos K. Implementation of the simple endoscopic activity score in Crohn's disease. *Saudi Journal of Gastroenterology* 2016; 22(3): 183. doi: 10.4103/1319-3767.182455

[38] Nogueira AR, Pugnana A, Ruggieri S, Pedreschi D, Gama J. Methods and tools for causal discovery and causal inference. *WIREs Data Mining and Knowledge Discovery* 2022; 12(2). doi: 10.1002/widm.1449

[39] Hammerton G, Munafò MR. Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine* 2021; 51(4): 563–578. doi: 10.1017/s0033291720005127

[40] Murray J, Philipson P. A fast approximate EM algorithm for joint models of survival and multivariate longitudinal data. *Computational Statistics & Data Analysis* 2022; 170: 107438. doi: 10.1016/j.csda.2022.107438

[41] Li S, Li N, Wang H, Zhou J, Zhou H, Li G. Efficient algorithms and implementation of a semiparametric joint model for longitudinal and competing risk data: with applications to massive biobank data. *Computational and Mathematical Methods in Medicine* 2022; 2022: 1362913. doi: 10.1155/2022/1362913

Supplemental Materials for Constantine-Cooke et al.

Appendix A: Figures and Tables

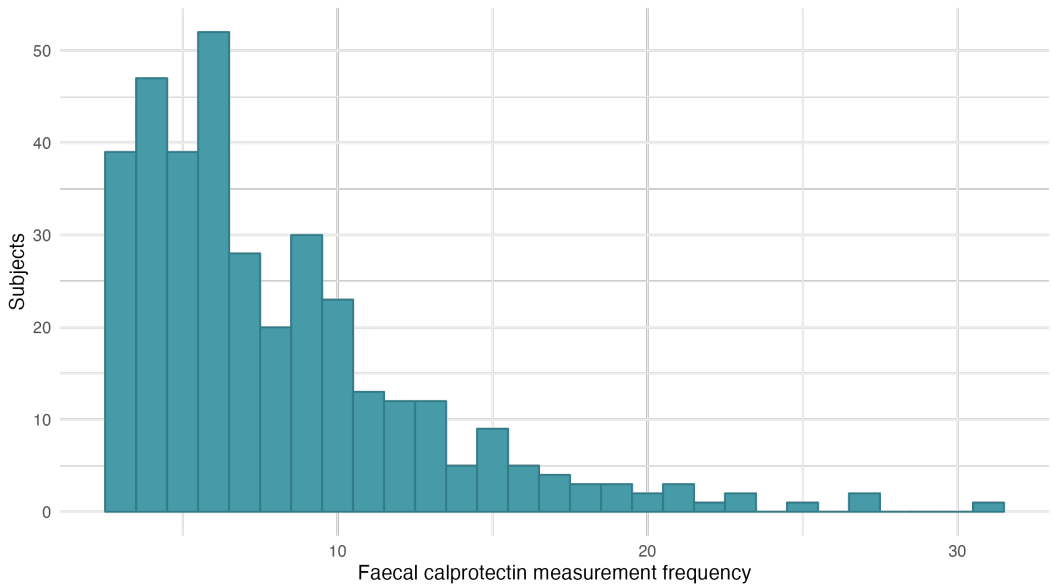

Figure S1: Distribution of number of FCAL measurements within five years of diagnosis per subject.

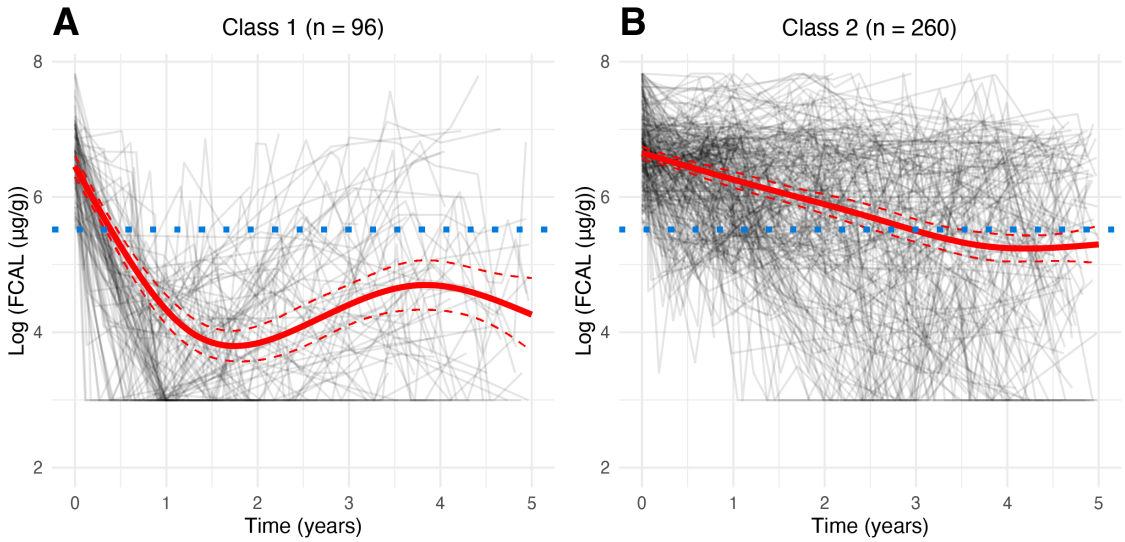

Figure S2: Assuming two latent classes, log-transformed subject-specific five-year faecal calprotectin profiles for the study cohort for **A**, class 1; **B**, class 2. The red solid line represents the predicted mean trajectory for each group whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 $\mu g/g$): the commonly accepted threshold for biochemical remission in Crohn's disease.

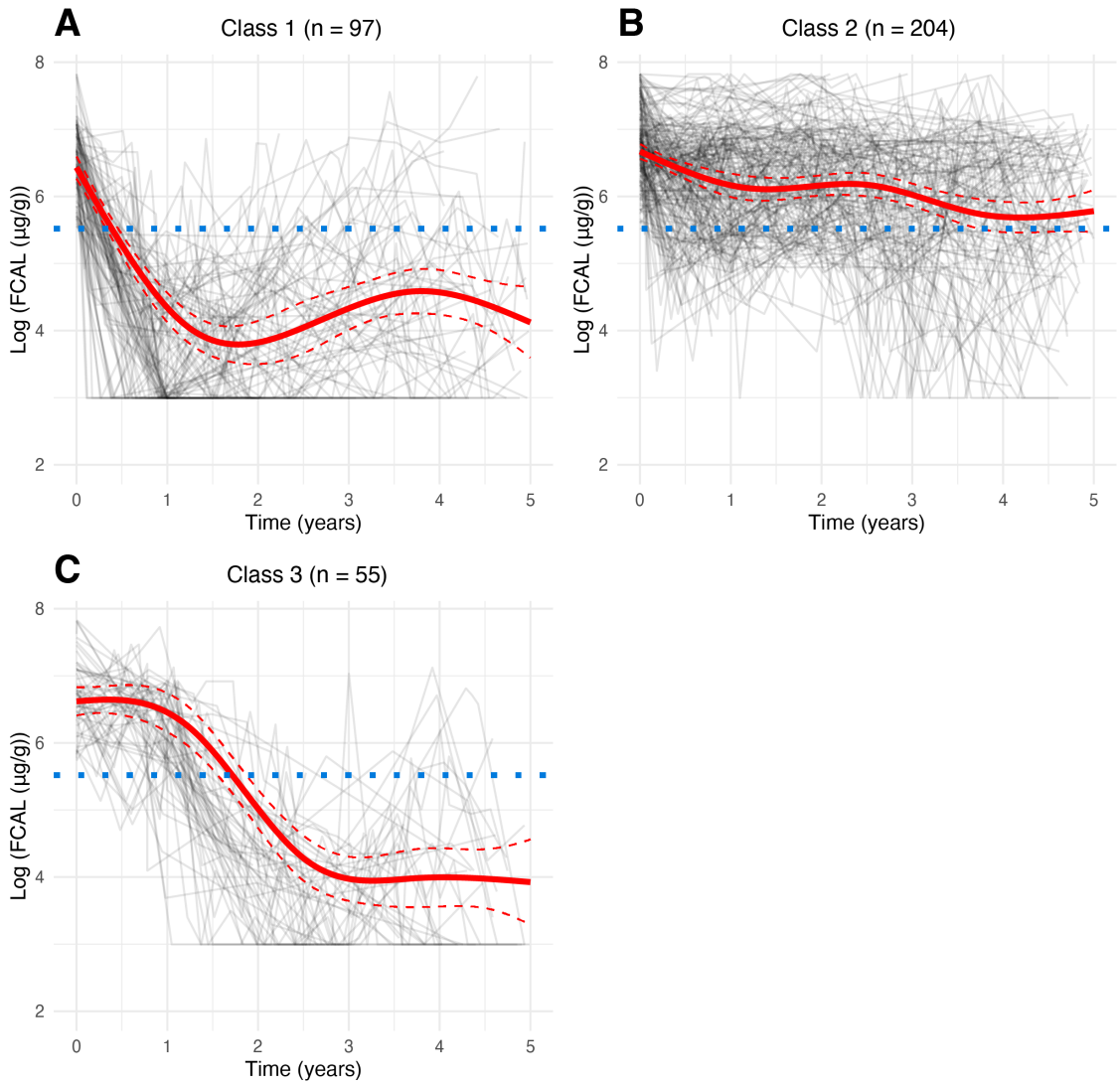

Figure S3: Assuming three latent classes, log-transformed subject-specific five-year faecal calprotectin profiles for the study cohort for **A**, class 1; **B**, class 2; **C**, class 3. The red solid line represents the predicted mean trajectory for each group, whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 $\mu g/g$): the commonly accepted threshold for biochemical remission in Crohn's disease.

G	Maximum log-likelihood	AIC	BIC
1	-4030.5	8103	8184.4
2	-4005.3	8064.7	**8169.3**
3	-3990.5	8064.9	8174.8
4	-3983.5	8045	8196.1
5	-3970.2	**8030.3**	8204.7
6	**-3968.4**	8038.8	8236.5

Table S1: Model fit statistics for latent class models fitted to the faecal calprotectin data for different numbers of latent subgroups. G: number of assumed latent classes; AIC: Akaike information criterion; BIC: Bayesian information criterion.

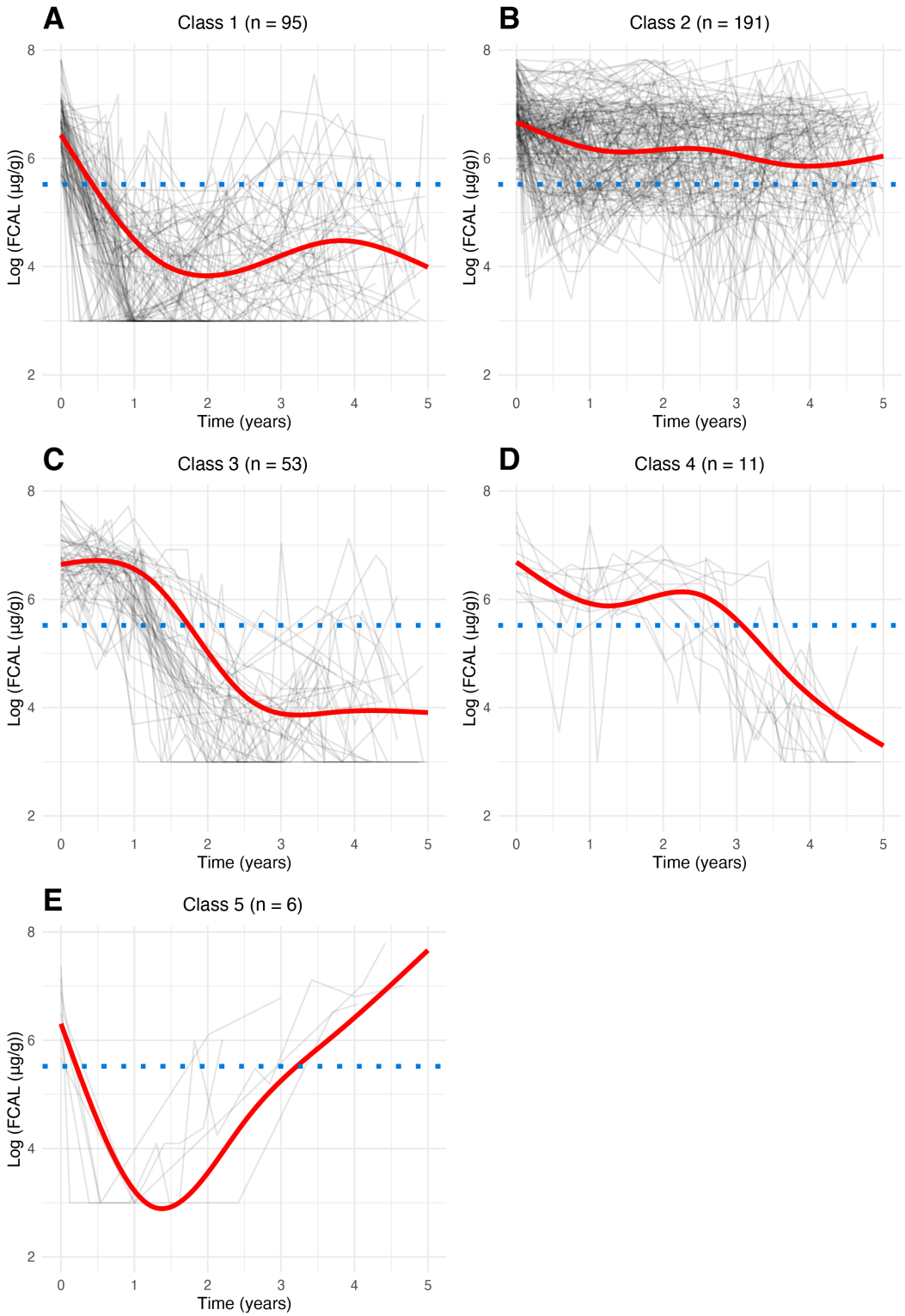

Figure S4: Assuming five latent classes, log-transformed subject-specific five-year faecal calprotectin profiles for the study cohort for **A**, class 1; **B**, class 2; **C**, class 3; **D**, class 4; **E**, class 5. The red solid line represents the predicted mean trajectory for each group, whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 $\mu g/g$): the commonly accepted threshold for biochemical remission in Crohn's disease.

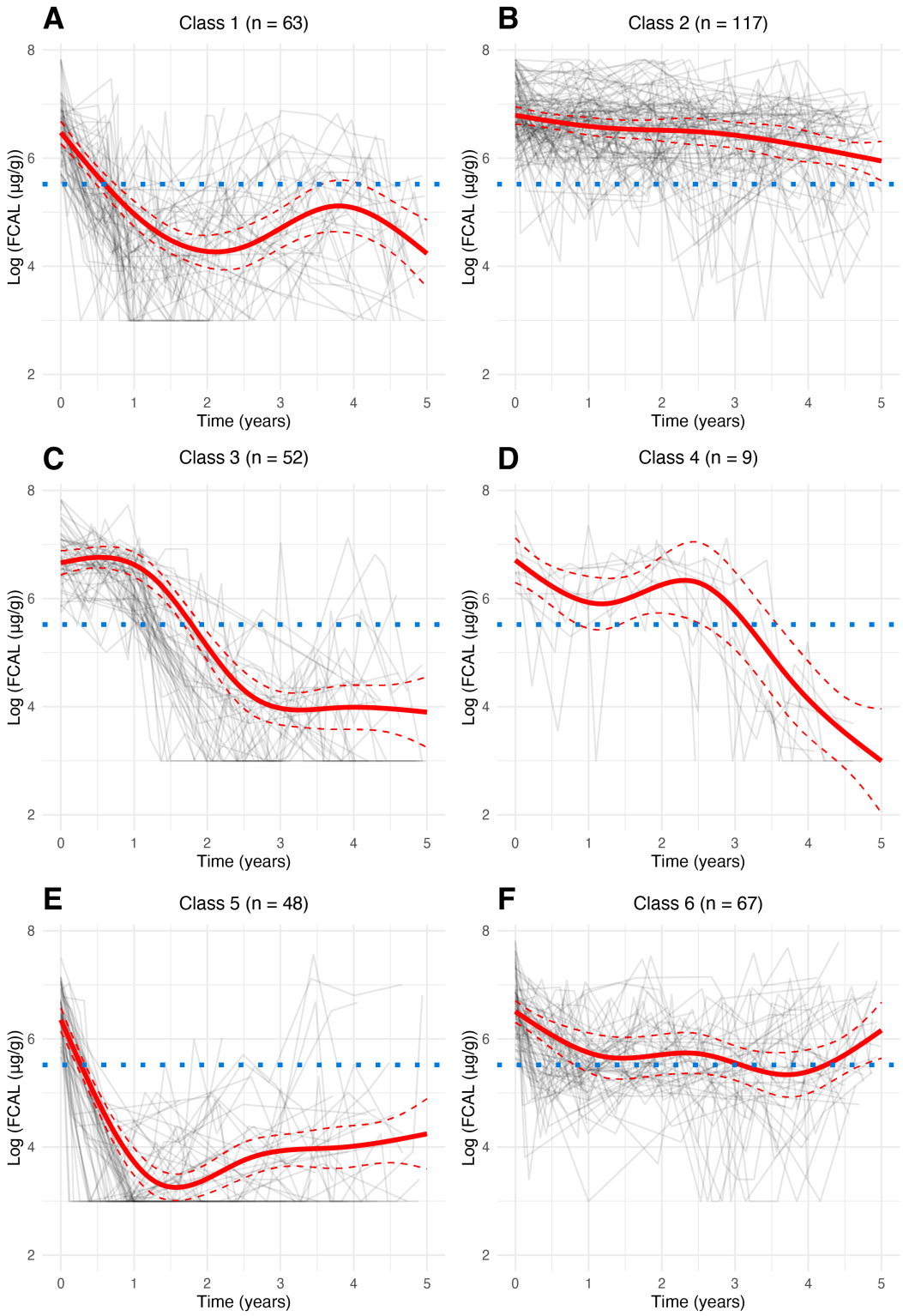

Figure S5: Assuming six latent classes, log-transformed subject-specific five-year faecal calprotectin profiles for the study cohort for **A**, class 1; **B**, class 2; **C**, class 3; **D**, class 4; **E**, class 5; **F**, class 6. The red solid line represents the predicted mean trajectory for each group, whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject. The blue dotted line indicates an FCAL of log(250 $\mu g/g$): the commonly accepted threshold for biochemical remission in Crohn's disease.

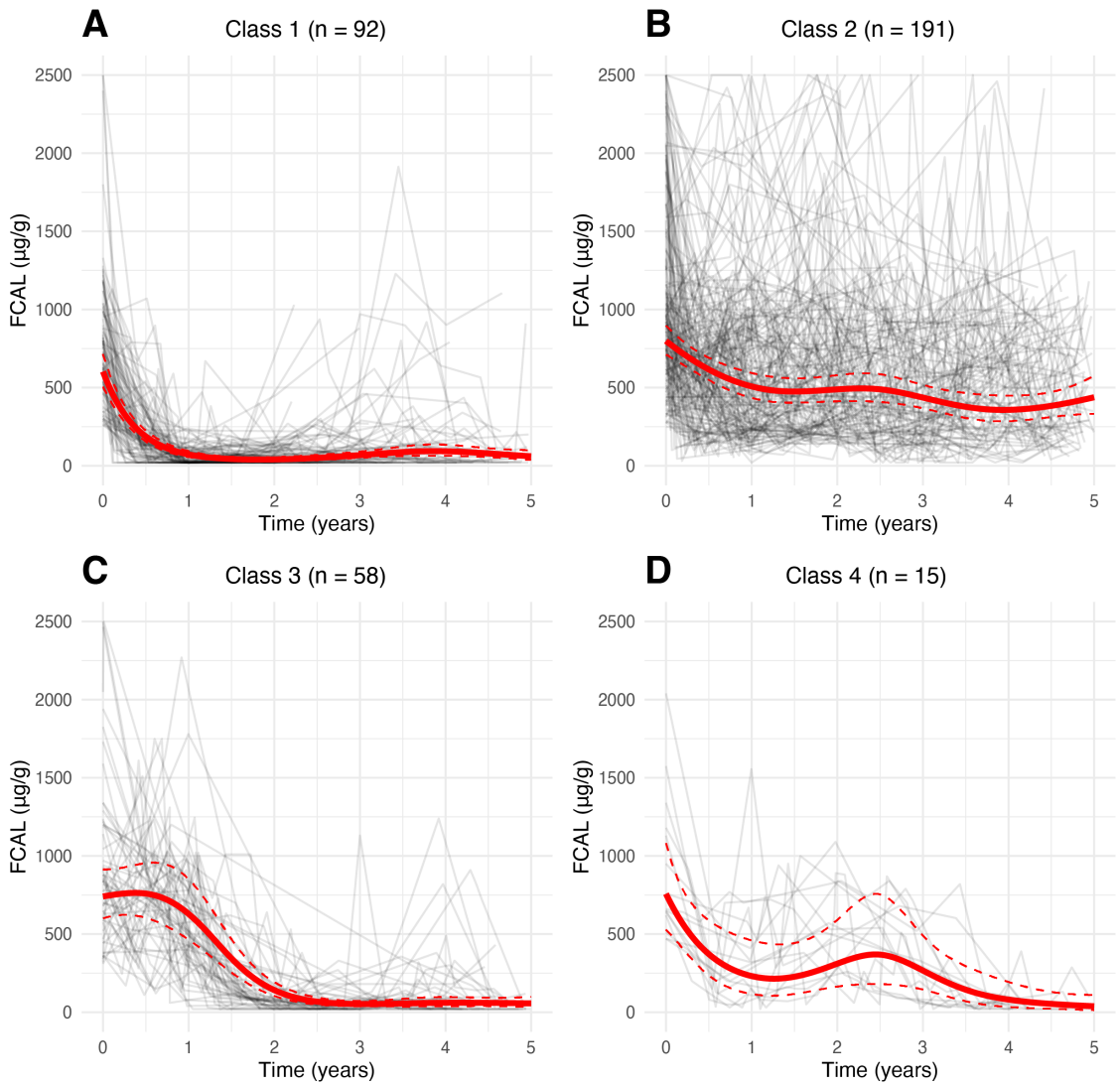

Figure S6: Five-year mean faecal calprotectin (FCAL) trajectories for latent subgroups obtained by fitting latent class mixed models for **A**, class 1; **B**, class 2; **C**, class 3; **D**, class 4. The red solid line represents the predicted mean trajectory for each group, whilst the red dotted lines represent 95% confidence intervals. The grey lines indicate the trajectory of each subject.

Appendix B: Statistical Methods

Formal Definition of the Model

We assume a population of N individuals is heterogeneous and composed of G latent classes: each characterised by a distinct mean profile of FCAL (in logarithmic scale) across time. We assume each subject i has a vector of repeated FCAL measurements of length n_i: allowing the number of measurements to differ across subjects. Random effects specification are used to capture intra-individual correlation in FCAL measurements. We allow each subject i to belong to only one latent class and introduce a discrete random variable c_i which is equal to g if subject i belongs to the latent class g, where $g = 1, \ldots, G$.

The logarithm of the FCAL measurement for the ith subject taken at time t_{ij} is denoted by Y_{ij}. Given that the subject i belongs to class g, the latter is modelled using a latent class mixed model LCMM:

$$Y_{ij}|_{c_i=g} = \mathbf{X}(t_{ij})'\beta_g + \mathbf{X}(t_{ij})'u_{ig} + \epsilon_{ij} \tag{S1}$$

where the vector of regression coefficients β_g capture class-specific fixed effects, u_{ig} denote random effects distributed such that $u_{ig} \sim \mathcal{N}(0, B)$ (the variance-covariance matrix is shared across classes) and ϵ_{ij} indicates an independently distributed Gaussian error term with zero mean and variance σ^2_ϵ.

In (S1), $\mathbf{X}(t) = (1, X_1(t), X_2(t), X_3(t), X_4(t))'$ is a vector of time-dependent covariates used to capture non-linear dependency between log-FCAL values and time (the first element ensures the model includes an intercept term). These are defined using natural cubic splines with three knots (4 cubic polynomials)[1]. The natural cubic splines were calculated as a pre-processing step prior to estimating the model in (S1) using the `ns` function of the `splines` R library[2]. For this purpose, knots were located at the first, second and third quantiles of measurement times across all FCAL measurements.

The probability of $c_i = g$ is given as a class specific probability and is described by a multinomial logistic model:

$$\pi_{ig} = P\left(c_i = g\right) = \frac{e^{\xi_{0g}}}{\sum_{l=1}^{G} e^{\xi_{0l}}} \tag{S2}$$

where ξ_{0g} indicates the intercept for class g in this model. For identifiability, $\xi_{0G} = 0$.

After inferring all model parameters, posterior class-membership probabilities for each subject are given by:

$$\hat{\pi}^Y_{ig} = P(c_i = g|\mathbf{Y}_i, \mathbf{X}(t_{i\cdot}), \hat{\theta}_G) = \frac{\hat{\pi}_{ig}\phi_{ig}(\mathbf{Y}_i|c_i = g, \hat{\theta}_G)}{\sum_{l=1}^{G} \hat{\pi}_{il}\phi_{il}(\mathbf{Y}_i|c_i = l, \hat{\theta}_G)} \tag{S3}$$

where $\mathbf{Y}_i$ denotes a vector of length n_i containing all longitudinal measurements recorded for subject i, $\mathbf{X}(t_{i\cdot})$ is a matrix ($n_i \times 4$) comprised of all the corresponding time-dependent covariates for subject i, $\hat{\theta}_G$ denotes the estimates obtained for all model parameters ($\beta_1, \ldots, \beta_G, B, \sigma^2_\epsilon$) and $\hat{\pi}_{ig}$ corresponds to (S2) evaluated on $\hat{\theta}_G$. Finally, $\phi_{ig}(\mathbf{Y}_i|c_i = g, \hat{\theta}_G)$ denotes a multivariate normal density function with mean $\mathbf{X}(t_{i\cdot})\hat{\beta}_g$ and variance covariance $\mathbf{X}(t_{i\cdot})\hat{B}\mathbf{X}(t_{i\cdot})' + \hat{\sigma}^2_\epsilon I_{n_i}$, where I_{n_i} denotes an identify matrix with dimension n_i.

References

[1] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*: 241-247; Springer Series in Statistics. Springer New York. 2nd ed. 2009

[2] R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2022.