



Universiteit
Leiden
The Netherlands

Ecology meets human health: studies on human gut microbiota in health and disease

Pinto, S.

Citation

Pinto, S. (2025, November 20). *Ecology meets human health: studies on human gut microbiota in health and disease*. Retrieved from <https://hdl.handle.net/1887/4283645>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4283645>

Note: To cite this publication please use the final published version (if applicable).



Part II

**Gut microbiota and
inflammatory bowel disease**

Chapter 4

Heterogeneous associations of gut microbiota with Crohn's disease activity

Susanne Pinto¹, Elisa Benincà², Gianluca Galazzo^{3,4},
Daisy M. A. E. Jonkers^{3,5}, John Penders^{3,4}, Johannes A. Bogaards^{6,7}

- 1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
- 2 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
- 3 School for Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University, Maastricht, the Netherlands
- 4 Department of Medical Microbiology, Infectious Diseases and Infection Prevention, Maastricht UMC, Maastricht, the Netherlands
- 5 Department of Gastroenterology-Hepatology, Maastricht UMC, Maastricht, the Netherlands
- 6 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
- 7 Amsterdam Institute for Infection and Immunity (AI&I), Amsterdam UMC, Amsterdam, the Netherlands

Gut Microbes. 2024 Jan-Dec;16(1):2292239. doi: 10.1080/19490976.2023.2292239.

Heterogeneous associations of gut microbiota with Crohn's disease activity

Abstract

The multifactorial involvement of gut microbiota in Crohn's disease (CD) necessitates robust analysis to uncover potential associations with specific microbes. CD has been linked to certain bacteria, but reported associations vary widely across studies. This inconsistency may result from heterogeneous associations across individual patients, resulting in no apparent or only weak relationships with the means of bacterial abundances. We investigated the relationship between bacterial relative abundances and disease activity in a longitudinal cohort of CD patients ($n = 57$) and healthy controls ($n = 15$). We applied quantile regression, a statistical technique that allows investigation of possible relationships outside the mean response. We found several significant and mostly negative associations with CD, especially in lower quantiles of relative abundance on family or genus level. Associations found by quantile regression deviated from the mean response in relative abundances of Coriobacteriaceae, Pasteurellaceae, Peptostreptococcaceae, Prevotellaceae, and Ruminococcaceae. For the family Streptococcaceae we found a significant elevation in relative abundance for patients experiencing an exacerbation relative to those who remained without self-reported symptoms or measurable inflammation. Our analysis suggests that specific bacterial families are related to CD and exacerbation, but associations vary between patients due to heterogeneity in disease course, medication history, therapy response, gut microbiota composition, and historical contingency. Our study underscores that microbial diversity is reduced in the gut of CD patients, but suggests that the process of diversity loss is rather irregular with respect to specific taxonomic groups. This novel insight may advance our ecological understanding of this complex disease.

Introduction

Crohn's disease (CD) is a chronic inflammatory disorder that can affect any part of the digestive tract, but mostly involves the ileum and colon.³⁵³ The disease is characterized by periods of inflammation (exacerbation) interspersed by periods without symptoms (remission). During exacerbation, the patients are suffering from a range of different symptoms, including diarrhea, abdominal pain, bloody stool, fatigue, and weight loss. Prolonged inflammation can lead to severe complications, such as damage to the gastrointestinal tract and malnutrition.³⁵³ While the exact cause of CD is unknown, an inappropriate immune response against commensal gut bacteria, host genetics, and environmental factors are all thought to be involved in disease pathophysiology.³⁵⁴ The gut microbiota in CD patients is characterized by a reduced diversity and lower long-term stability as compared to healthy individuals.³⁵⁵ Also, shifts in abundance of specific bacterial genera or families have been associated with CD,³⁵⁶ its disease course,¹⁷⁷ and disease activity.¹⁸¹

Several studies have investigated relations between specific microbial groups and CD. *Faecalibacterium prausnitzii* (Ruminococcaceae), *Clostridium leptum* (Clostridiaceae), and

Clostridium coccooides (Clostridiaceae) were found to be negatively associated with CD as well as disease activity.^{181, 356, 357} Conversely, the family Enterobacteriaceae was found to be positively associated with CD and with disease activity.^{171, 358} However, the patterns of association with specific microbes are not always consistent among studies. Within the Bacteroidaceae family conflicting results were found. For example, CD patients showed both lower relative abundances,^{356, 358} as well as higher relative abundances³⁵⁷ in *Bacteroides* (Bacteroidaceae) compared to healthy individuals.

The inconsistency in findings might be partly due to technical artifacts, such as differences between studies in sequencing methods to quantify gut microbiota composition and the compositional nature of data obtained by most next generation sequencing (NGS) techniques. Another explanation is that the heterogeneous responses among patients may derive from multifactorial dependencies, between microbial elements themselves and between gut microbiota and host factors, such as treatment with immunomodulatory drugs, lifestyle, and diet, but also in underlying disease characteristics such as disease location, severity, and epigenetic immune regulation.^{22, 359} This heterogeneity among patients is reflected by a strong variation in disease course, the response to medication, and the need for surgery among subgroups of patients.³⁶⁰ The involvement of specific bacterial groups in CD will likewise depend on multiple factors. Some of these factors can be accounted for when relating CD to gut microbiota composition, although correction relies on adequate model specification which is difficult in multi-factorial systems. Moreover, many factors which may strongly determine the observed relationships between bacterial abundance and CD activity have not been identified or are not routinely measured. One such factor is the order in which specific bacteria have been acquired throughout life. Rapid colonization by maternal and environmental bacteria occurs within days of birth and is unique per person. The temporal development of the microbiota is directed, implying that the growth of certain species precedes the growth of others, leading to the unique microbiomes in adult life. This historical contingency of gut microbiota might also influence how microbes react to future perturbations in that gut community.⁶⁷

The multi-factorial involvement of specific microbial groups with CD necessitates robust analysis to uncover possible associations, as there may be no apparent or only weak relationships with the means of bacterial abundances. Here, we apply quantile regression, an extension of the general linear model that allows for investigation of relationships across different quantiles of the distribution of a response variable.^{361, 362} Quantile regression extends regression of the mean to the analysis of the entire conditional distribution of the response variable.³⁶² Examining quantile regression functions across the entire range of quantiles provides a more complete view of the response variable distribution than achieved by standard regression analysis.³⁶¹ The idea behind this method is that not all individuals are equally responsive to changes in abundance of specific bacterial groups, due to hidden bias and complex dependencies in ecological datasets.³⁶¹ Quantile regression is less sensitive to outliers than conventional regression and is not dependent on homoscedastic errors.³⁶³ In particular, we tested whether associations between relative abundances of specific families with CD can be found with a clinical diagnosis (i.e., remission vs. exacerbation), but also with specific markers (i.e., fecal calprotectin (FC), serum C-reactive protein (CRP), and the Harvey Bradshaw index (HBI)) of disease activity in repeatedly sampled CD patients and healthy controls.¹⁸¹

Methods

Data and procedures

The study population has previously been described in Galazzo et al. (2019).¹⁸¹ A total of 57 CD patients were included in this study. Demographic variables and subject characteristics are provided in Appendix Table 4.1 and medication use between visits is provided in Appendix Table 4.2. The CD patients formed a subset of the Inflammatory Bowel Disease South Limburg Cohort.¹⁶⁹ As a reference group, 15 healthy cohort (HC) subjects, all without any gastrointestinal disease, gastrointestinal symptoms, or comorbidities, were recruited among the controls who participated in the Maastricht Irritable Bowel Syndrome (IBS) Cohort.³⁶⁴ Clinical data, blood, and feces were collected at two time points. The CD group comprised 22 remission-exacerbation (RE) patients with baseline sampling at time of remission and subsequent sampling during an exacerbation, and 35 remission-remission (RR) patients, with two subsequent samples while maintaining remission, i.e., without any flares in between subsequent samples. The median time between baseline and follow up samples was 14 (IQR 11–21), 20 (8–36), and 13 (12–16) weeks for RR patients, RE patients, and HCs, respectively (Appendix Table 4.2). All study subjects gave written informed consent prior to participation. Both studies have been approved by the Medical Ethics Committee of Maastricht University Medical Center and have been registered in the US National Library of Medicine (www.clinicaltrials.gov; NCT02130349 and NCT00775060, respectively).

Fecal samples were collected at home, kept at room temperature, and brought to the hospital within 12 hours after defecation. Part of the fecal sample of the CD patients was sent to the laboratory of Clinical Chemistry for routine analysis of FC. The remaining part was aliquoted and frozen at -80°C for microbiota analysis. Disease activity was defined by FC, serum CRP, and HBI. Patients were included in the study when patients were in remission at baseline, i.e., $\text{FC} < 100 \mu\text{g/g}$ and $\text{CRP} < 5 \text{ mg/L}$ or $\text{FC} < 100 \mu\text{g/g}$, $\text{CRP} < 10 \text{ mg/L}$, and $\text{HBI} \leq 4$. Exacerbation at the second time point was defined by $\text{FC} > 250 \mu\text{g/g}$ or $\text{FC} > 100 \mu\text{g/g}$ with at least a 5-fold increase from baseline (Appendix Figure 4.1). The fecal microbiota composition was assessed by Illumina Miseq sequencing of the V4-region of the 16S rRNA gene. A detailed description of metagenomic DNA isolation, sequencing, and quality control is provided in the supplementary information of Galazzo et al. 2019.¹⁸¹ The 16S rRNA gene sequencing data are released in the European Nucleotide Archive. The accession number is: PRJEB62578 (ERP147674). Information on microbial profiling and the selection of bacterial families for quantile regression analysis can be found in Box 4.1.

Box 4.1 - Data procedures and family selection. Data demultiplexing, length and quality filtering, and clustering of reads into Operational Taxonomic Units (OTUs) at 97% sequence identity was done using the online Integrated Microbial Next Generation Sequencing (IMNGS) platform using default settings except for minimum and maximum length for amplicons, which were set at 100 and 500 bp, respectively.³⁶⁵ After quality filtering, binning, and removing unassigned reads, sequences were clustered in 640 OTUs. Normalization was performed by dividing OTU counts per sample for their total count (sample depth) and then multiplying the obtained relative abundances by the lowest sample depth. OTU sequences assigned to chloroplasts were removed prior to the statistical analyses. Then, the 18 different families used in the main text were selected by removing rare reads (not seen more than three times in at least 20% of the samples). We also performed quantile regression on the remaining families (which were not selected by the base case threshold), these results are placed in the Appendices.

Linear quantile regression mixed models (LQMM)

The quantile regression model takes the form $Q_{Y|X}(\tau) = X\beta_\tau$, where $Q_{Y|X}(\tau)$ denotes the τ^{th} quantile of the response variable Y , which is predicted from a vector X of explanatory variables with quantile specific parameters β_τ . The τ^{th} quantile is the inverse of the cumulative distribution function of Y , i.e., $q_\tau(\tau) = F_Y^{-1}$ or reciprocally $F_Y(q_\tau) = P(Y \leq y) = \tau$; where $\tau \in [0,1]$. It denotes the smallest value where the probability of finding an even smaller value is less than or equal to τ , whereas the probability of finding a larger value is less than or equal to $1 - \tau$.³⁶¹ A parametric distribution is assumed for the deterministic part of the model, but the random error part does not assume any distributional form. Further information on inclusion of covariates and model building strategy is supplied in Box 4.2.

Box 4.2 - Model building strategy. The analysis was divided in three parts. First, we investigated whether the relative abundances of the bacterial families could be explained by the group to which each individual belongs (i.e., healthy control (HC), remission-remission (RR), or remission-exacerbation (RE)). We added the interaction with visit number, to allow for different temporal changes in bacterial relative abundance over time between healthy controls, CD patients who experienced an exacerbation at the second visit, and those who remained in remission. Secondly, we excluded the healthy individuals from the model and investigated whether the relative abundances of the bacteria could be explained by disease activity (i.e., remission vs. exacerbation) in the CD patient group. Thirdly, we additionally investigated if the relative abundances of the bacterial families could be related to a quantitative disease indicator (i.e., HBI, CRP, or FC) other than the clinical definition of disease activity (remission and exacerbation). The models contain two time points per individual. Therefore, we used a random intercept per patient as well as a random effect for the variable 'visit number', because temporal changes in bacterial family's relative abundances may differ within patients, even when accounting for the fixed effect of disease trajectory (e.g., experiencing an exacerbation at the second visit).

Prior to the analyses, relative abundances were multiplied with 1000 and log-transformed with the natural-log function assuming a lower detection limit of 100 reads (which is $\frac{1}{4}^{\text{th}}$ of the lowest measurable value in the data). Prior to variable selection, all models contained the variables sex (male vs. female), smoking (current, ex, or never), and age (centered around mean age of 39.6 years) (Appendix Table 4.1). The models for CD patient cohort data also contained the disease-specific variables disease location (colonic (C), ileal (I), or ileocolonic (IC)), age at diagnosis (younger than 40 years (0) or older than 40 years (1)), surgery (no (0) vs. yes (1)), disease phenotype (non-stricturing/non-penetrating vs. stricturing/penetrating), and current treatment (mesalazines (no (0) vs. yes (1)), thiopurines (no (0) vs. yes (1)), biologicals (no (0) vs. yes (1)), induction (no (0) vs. yes (1)), and proton pump inhibitors (PPI) (no (0) vs. yes (1))) (Appendix Table 4.1). Variable selection was performed by running all possible models and then selecting the model with the lowest Bayesian Information Criterion (BIC) in the 50% quantile. For the model with the disease indicators (HBI, CRP, and FC), variable selection was performed on a model that contained all three indicators. For the sake of comparison, the selected variables were also used in the separate models for HBI, CRP, and FC. FC was divided by 1000 to improve numerical precision in quantile regression. Moreover, HBI, CRP, and FC are measured on different scales, therefore the data was normalized beforehand to make the quantile regression model estimates comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles.

We used the 'lqmm' R package (version 1.5.5)³⁶⁶ of the R statistical analysis software (www.R-project.org) to perform the quantile regression analysis. Analyses were performed separately for each bacterial family. Although genus level might be preferred, this would have resulted in too many models. Therefore, we only looked at certain genus levels, when significant results were found at family level (within the base case selection as described in Box 4.1). To accommodate repeated sampling on the individual level, we employed a linear quantile mixed model (LQMM) framework. A mixed model contains both fixed effects and random effects, and can then account for correlation in repeated measurements from the same individual as these are likely to be more similar than observations from different individuals.³⁶⁷ We estimated the series of quantile regression functions from the 10th to the 90th percent quantile. We used the Benjamini-Hochberg (BH) procedure per quantile to control for the expected proportion of 'false discoveries' across microbial families.²⁶⁶ However, the BH procedure assumes independency in multiple testing, which is likely not the case in the gut microbiota. Therefore, the BH correction might provide too conservative estimates, and we choose to also report the unadjusted results.

Results

Differences in abundance between healthy individuals and CD patients

We found several associations between the relative abundances of bacterial families with CD, and more specific with remission or disease exacerbation (Figure 4.1). The quantiles that were significantly associated with CD are different per bacterial family (Appendix Figure 4.2). For example, patients with baseline sampling at time of remission and subsequent sampling during an exacerbation (RE) displayed significantly distinct distributions in relative abundance in the family Coriobacteriaceae (Figure 4.2A), both at baseline (visit 1) and at the second visit, compared to the healthy control subjects (HC). At baseline, there was a positive association in the higher quantiles and over time (at time of exacerbation) there was a negative association in the lower quantiles. This means that the distribution of Coriobacteriaceae abundance among RE patients is skewed to higher values at baseline, but to lower values at the follow-up visit, as compared to healthy controls (see also Figure 4.1D). However, these effects were no longer significant after BH correction (Appendix Figure 4.3). For Coriobacteriaceae we also found a significant relation in the higher quantiles of patients in the RE group compared to the patients with two subsequent samples while maintaining remission (RR) (Appendix Figure 4.4). Thus, a significant fraction of patients in the RE group had higher Coriobacteriaceae abundance than healthy individuals and RR patients at baseline.

The family Erysipelotrichiaceae (Figure 4.2B) displayed negative associations in relative abundance over almost all quantiles (except the most upper quantiles) for both patient groups compared to the healthy controls. We still found significant differences after BH correction, but these were only present in the lowest quantiles (Appendix Figure 4.3). Looking at genus level, the genera *Holdemanella* and *Turicibacter* both displayed a similar pattern of significant results (Appendix Figures 4.5 to 4.7). We did not find a significant difference among the patients in the RE and RR group (Appendix Figure 4.4). This implies that the relative abundance of Erysipelotrichiaceae is severely skewed to lower values in CD patients (see also Figure 4.1G). The same kind of relationship was also found for Ruminococcaceae (Figure 4.2D). Similar to Erysipelotrichiaceae, the associations were only found at baseline, suggesting that this is a characteristic of CD, but not related to disease activity.

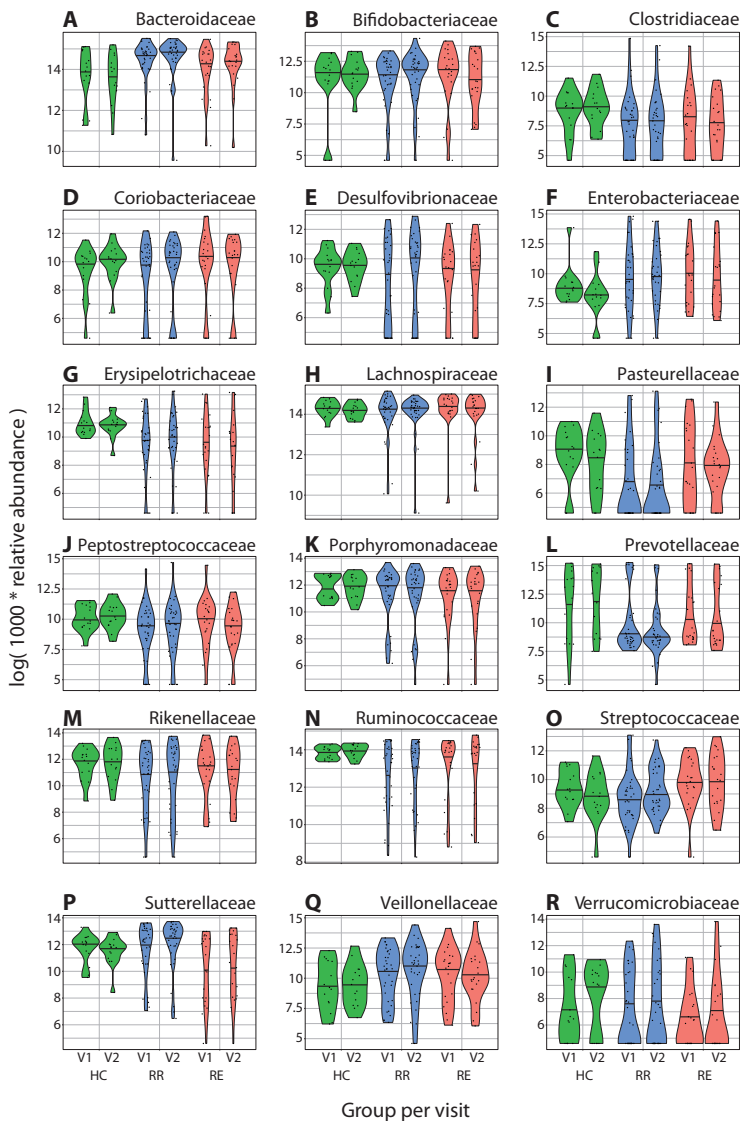


Figure 4.1 - Violin plots of transformed relative abundances of base case bacterial families by group and time point. In green the healthy controls (HC), in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50% quantile is shown with a black horizontal line. Genera are given in Appendix Figure 4.5 and families outside the base case criterion in Appendix Figure 4.8.

As another example, the relative abundances of Sutterellaceae among patients in the RE group were significantly skewed to lower values compared to both the healthy controls and the RR group patients (Figures 4.1P, 4.2F, and Appendix Figure 4.4). We also found a significant negative relation between the abundance of the family Pasteurellaceae in the RR group at baseline compared to the healthy controls (Figure 4.2C). For the family Streptococcaceae, we did not find many significant associations at baseline (except for one quantile), but the association for the RR x visit variable was significant for almost all quantiles (Figure 4.2E). This means that patients from the RR group experienced stronger increases in relative abundance of Streptococcaceae over time as compared to the healthy controls.

We also found a significant difference between the RR and RE patient groups for the family Streptococcaceae, with the RE patients having elevated abundances across the entire quantile range (Appendix Figure 4.4, see also Figure 4.1 - panel O). In our data, Pasteurellaceae and Streptococcaceae both only consisted of one classified genus. When refining the analyses for these classified genera, we did not find significant results within *Mannheimia* (Pasteurellaceae) or *Streptococcus* (Streptococcaceae) (Appendix Figures 4.5 to 4.7).

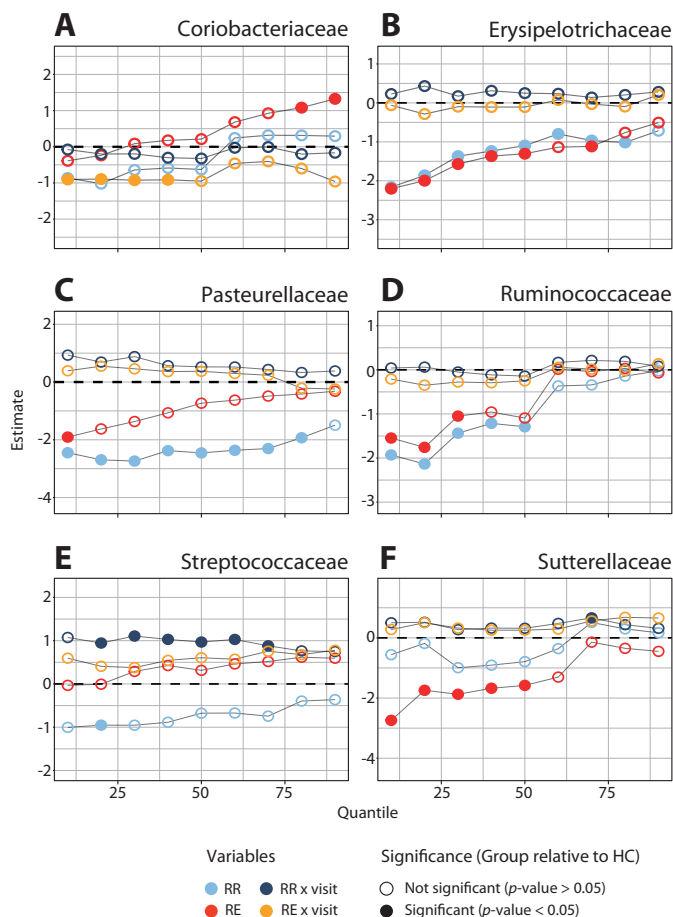


Figure 4.2 - Examples of quantile regression profile plots for some of the base case families. A) Coriobacteriaceae, B) Erysipelotrichaceae, C) Pasteurellaceae, D) Ruminococcaceae, E) Streptococcaceae, and F) Sutterellaceae. The various dots represent sample estimates (y-axis) of differences in relative abundance compared to healthy controls across the 10th to the 90th percent quantile (x-axis). Differences at baseline (visit 1) are visualized in light blue (RR) and red (RE), while the interaction with visit number (in dark blue and orange) displays the difference in changes over time. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance at that particular quantile. Significant variables (p -value < 0.05) are indicated with a closed circle. The other base case families are in Appendix Figure 4.2.

On top of the examples given in Figure 4.2, we also found that the relative abundances of Clostridiaceae, Desulfovibrionaceae, Peptostreptococcaceae, Prevotellaceae, and Rikenellaceae in CD patients were different from the relative abundances in the microbiota of healthy controls (Figure 4.3, Appendix Figure 4.2). These results, except for the families Prevotellaceae and Rikenellaceae, remained significant after BH correction (Appendix Figure 4.3). Most significant relations were negative and were found in the lower quantiles (Figure 4.3), meaning that CD patients more often displayed negatively than positively skewed abundance distributions (see also Figure 4.1). Besides, only a few associations between bacterial family abundance and covariates were found, with sex being the only significant covariate (males having higher abundances than females) (Figure 4.3). We also identified some significant associations in the families which fall under the sensitivity analyses of the families outside the base case selection criterion (Appendix Figure 4.8 to 4.10). We found that the relative abundances of Victivallaceae and Clostridiales fam. i.s. XI were different from the healthy controls for both the patients that stayed in remission and the patients that experienced an exacerbation. We also found significant results for the family Enterococcaceae for the patients that stayed in remission compared to the healthy controls and the families Actinomycetaceae and Lactobacillaceae for the patients that experienced an exacerbation at the second visit (Appendix Figure 4.9). However, these results disappeared after applying BH correction for multiple testing (Appendix Figure 4.10).

We compared our results from the LQMM models with the results obtained from an ordinary linear mixed-effects model (with similar variables as used in the LQMM models) using the 'lme' function from the 'nlme' R package (version 3.1).³⁶⁸ Example code of the LQMM and LME models can be found on the GitHub repository (susanepinto/Quantile-Regression-CD). Most associations found by quantile regression could also be found with ordinary regression, as the mean response in the linear mixed-effects model provides somewhat of an average response over all quantiles. Nevertheless, some differences were also noticeable (Appendix Figure 4.11). For example, the family Coriobacteriaceae has a positive estimate in the higher quantiles for patients in the RE group relative to the healthy control group, which is not visible in the mean response (Appendix Figure 4.11D). Likewise, patients in the RR group displayed significant reductions in abundance in the lower to middle quantiles of Prevotellaceae and Streptococcaceae abundance, that were apparent in a reduced mean response, but without statistical significance. Conversely, the reduced mean responses regarding Ruminococcaceae in both RR and RE patients hide the fact that reductions only apply to lower and middle quantiles of abundance (Figure 4.1N and Appendix Figure 4.11N). Comparable findings were obtained for Pasteurellaceae and Peptostreptococcaceae. In some instances, linear mixed-effects regression yielded imprecise (cf. Lachnospiraceae) or biased (cf. group $\times$ visit in Prevotellaceae) estimates as compared to quantile regression (Appendix Figure 4.11).

Gut microbiota changes in relation to Crohn's disease activity

The relation between bacterial family abundance and disease activity (exacerbation) was mainly negative across the quantile range, indicating reduced abundance among RE patients as compared to RR patients at the second visit. However, this association was only statistically significant for upper quantiles of Coriobacteriaceae after adjustment for covariates (Figure 4.4). Instead, significant associations were revealed with several clinical variables (e.g., phenotype, surgery, proton pump inhibitors (PPI), and biologicals), suggesting that differences between RR and RE patients might have been confounded by disease-specific variables (Figure 4.4).

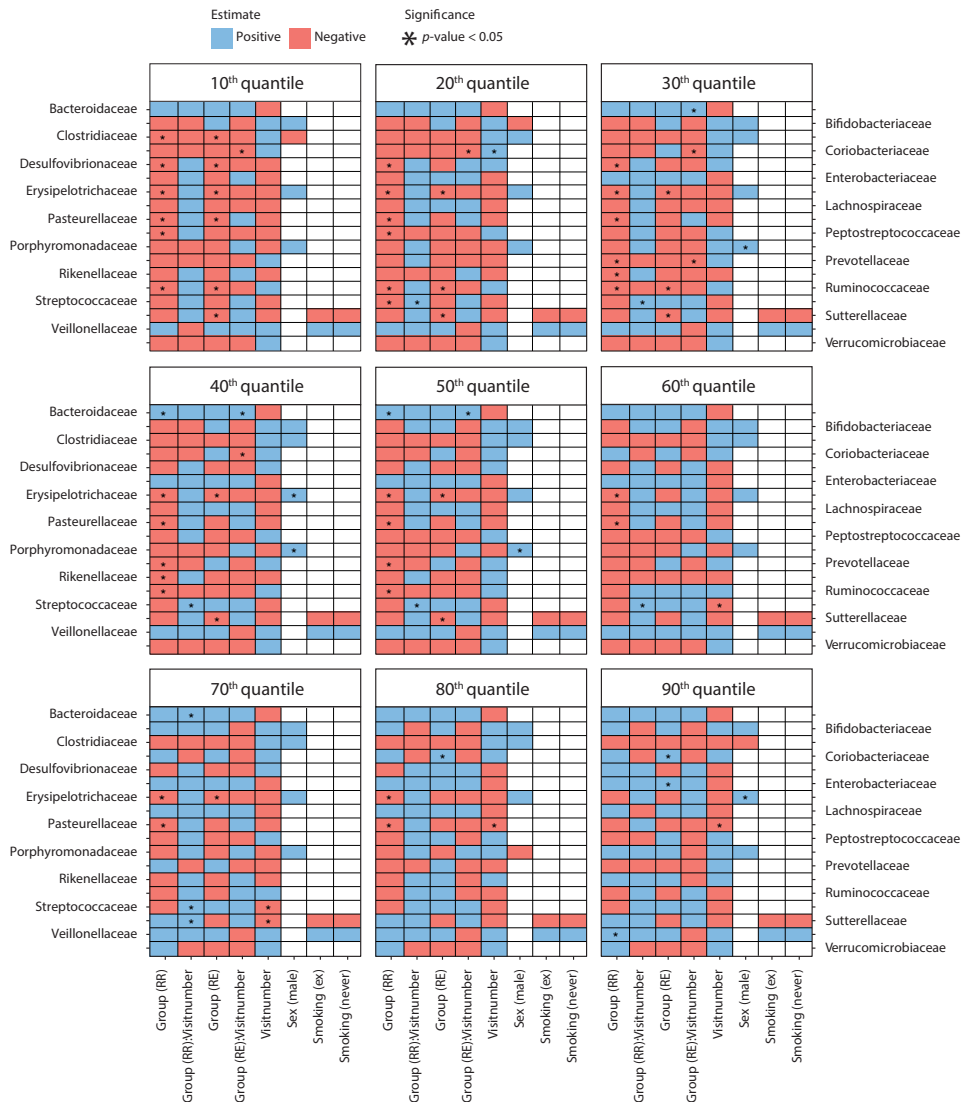


Figure 4.3 - Heatmap of quantile regression estimates across quantiles of relative abundance for base case families and common variables.

The model included all groups, i.e., healthy control (HC), remission-remission (RR), and remission-exacerbation (RE), with healthy controls as reference group. The red boxes indicate negative regression estimates, the blue boxes indicate positive regression estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05) are indicated with an asterisk (*), results adjusted with the BH procedure are given in Appendix Figure 4.3. Other families outside the base case criterion are given in Appendix Figure 4.9 and Appendix Figure 4.10.

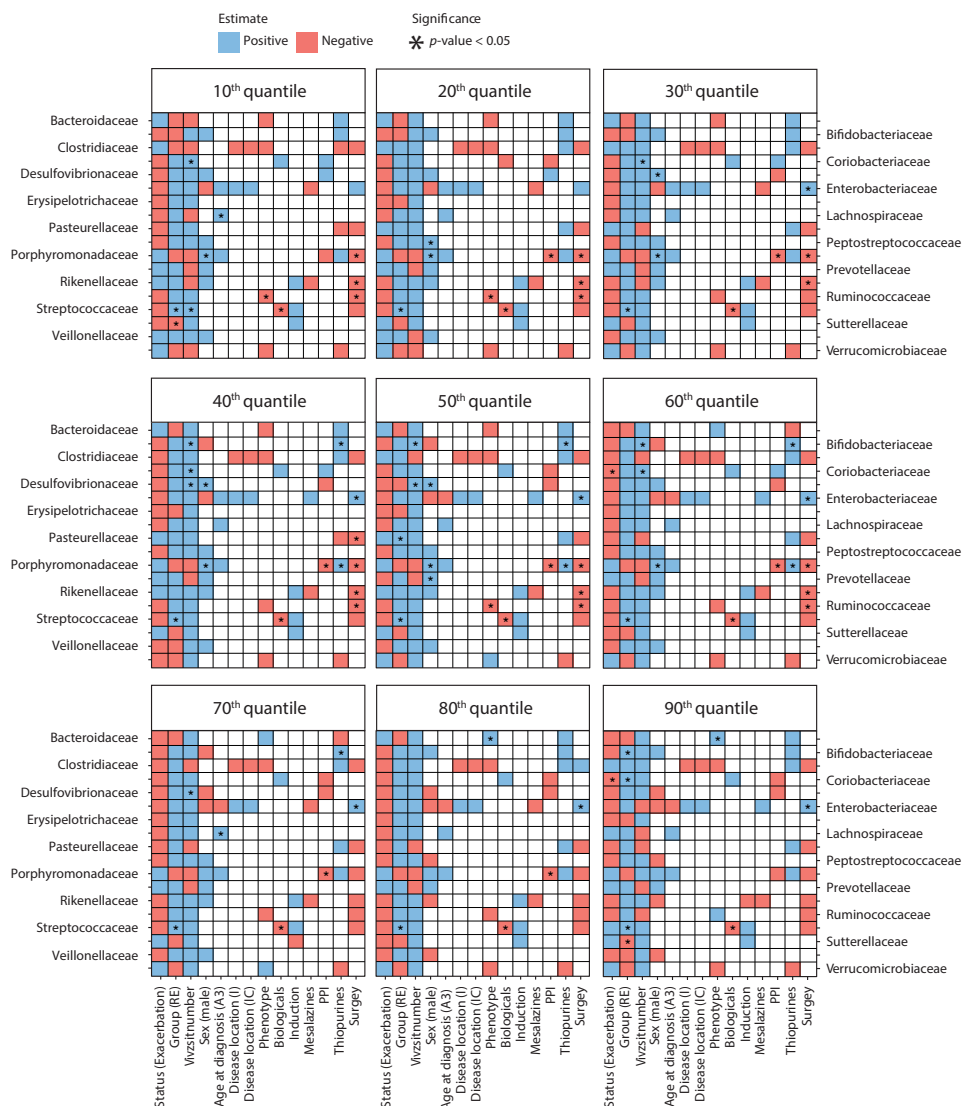


Figure 4.4 - Heatmap of quantile regression estimates across quantiles of relative abundance for base case families and clinical variables in CD patients only. The red boxes indicate negative regression estimates, the blue boxes indicate positive regression estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$) are indicated with an asterisk (*), results adjusted with the BH procedure are given in Appendix Figure 4.12.

Of note, many associations with disease activity also disappeared after BH correction for multiple testing (Appendix Figure 4.12), but the finding that RE patients had elevated Streptococcaceae abundances across the entire quantile range (irrespective of visit number) remained significant; likewise, treatment with biologicals remained significantly associated with lower Streptococcaceae abundance (Figure 4.4 and Appendix Figure 4.12). Further results on genera and other families can be found in Appendix Figures 4.13 to 4.16.

Bacterial relative abundances in relation to different disease activity indicators

When comparing regression on clinically defined exacerbation with different indicators of disease activity, we found that especially FC levels gave distinct results compared to the other indicators (Figure 4.5 and Appendix Figure 4.17). For almost all bacterial families, we did not observe a signal (estimates around zero) for clinical status (remission or exacerbation), CRP, and HBI after correction for clinical variables. In contrast, the normalized estimates of FC were much stronger (Appendix Figure 4.17), and significantly negative across the entire quantile range for Porphyromonadaceae and Verrucomicrobiaceae (Figure 4.5). The results of Porphyromonadaceae still hold after BH correction (Appendix Figure 4.18D). Further results on genera and other families can be found in Appendix Figures 4.19 to 4.22.

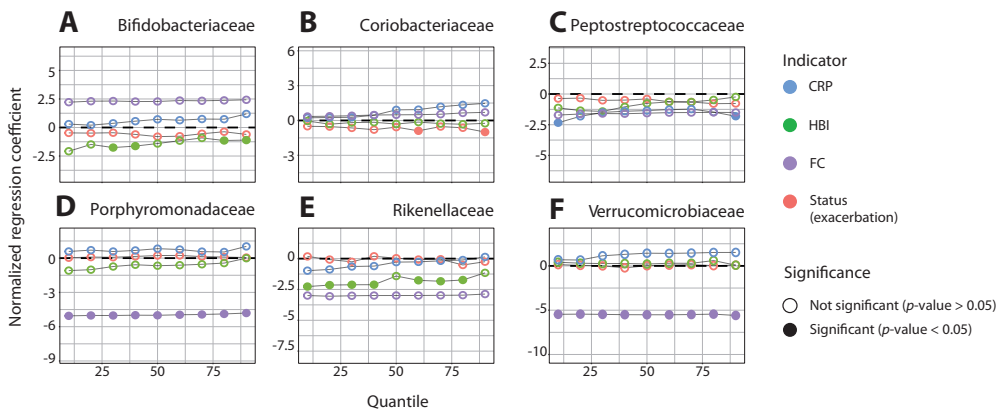


Figure 4.5 - Quantile regression profile plots for different disease activity indicators and clinical variables in CD patients only. For the families:

A) Bifidobacteriaceae, B) Coriobacteriaceae, C) Peptostreptococcaceae, D) Porphyromonadaceae, E) Rikenellaceae, and F) Verrucomicrobiaceae. Only families with significant results are given, the other families are in Appendix Figure 4.17. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance for that particular quantile. The regression estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data was normalized beforehand to make the regression coefficients comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p -value < 0.05) are indicated with a closed circle, results adjusted with the BH procedure are given in Appendix Figure 4.18.

Discussion

In this study, we investigated the possible associations between the relative abundance of specific bacterial families with CD and disease activity. We relied on quantile regression to uncover relationships that are not restricted to the mean response across CD patients. We found mainly negative associations with CD at family or genus level, especially in lower quantiles of relative bacterial abundance. These results are consistent with the frequently cited reduced microbial diversity in the gut of CD patients compared to healthy controls, but they also highlight that reductions for specific microbes are usually limited to a minority of patients. Thus, while CD coincides with a loss of microbial diversity in the gut, the process of diversity loss seems rather irregular with respect to specific taxonomic groups.

Associations outside of the mean response can be a result of heterogeneity among CD patients and may arise from the complex interactions among members of the microbiota. Microbes alter the environment through metabolic by-products, creating new ecological niches that promote diversification. However, some metabolites may adversely affect the growth of other microorganisms.⁸ Whether systemic changes, such as those induced by CD, lead to niche reduction or expansion for a particular microbe probably depends as much on the characteristics of that particular microbe as on the microbial ecosystem of the individual host. If specific bacterial groups respond to disease or disease activity in some of the patients but not in others, their associations with CD or disease activity are more likely to be found in upper or lower quantiles than in the mean or median response across CD patients.

Interestingly, almost all significant associations we found were negative and applicable to the lower quantiles of bacterial abundance. While positive associations in upper quantiles have been attributed to unmeasured factors that limit the potential response to a positive stimulus,³⁶¹ this opposite pattern is reminiscent of an ecosystem response to stress: the ability to maintain healthy bacterial abundances is gradually lost once the system gets close to a tipping point.⁹⁷ The loss of some species in the microbial network can still be compensated for by other species with similar ecosystem functions (functional redundancy), but the loss of too many may lead to a loss of resilience and critical transition to an alternative stable state.^{104, 369} Although the existence of tipping points in the onset or exacerbation of CD has not been demonstrated, a large-scale study by Lahti et al. (2014) showed distinct bimodal abundance patterns of certain bacterial species (i.e., tipping elements) among healthy human hosts.¹⁰² These species were present in either a high or low abundance state, supporting the idea of alternative stable states in the human gut microbiota. Taken together, concepts of ecosystem resilience and critical transitions in the gut microbiota may explain why some individuals respond strongly to systemic changes, as induced by CD, while others display a more robust microbiota composition.²²

Associations between the relative abundance of bacterial families and CD across the entire quantile range can also be identified with methods that focus on the mean response, such as ordinary linear regression. Families that exhibit such uniform responses could be considered to represent keystone bacterial groups, as their response to CD or disease activity is less dependent on other microbes or host factors as compared to families that are only responsive in some patients. This feature of robustness would be preferred for clinical diagnostics, prioritization for treatment, or monitoring of disease course, because guidelines can then be developed and used for all CD patients. However, it is also important to understand the less generic differences in the microbiota of CD patients because less robust associations may

also shed light on etiology and progression of CD and may provide leads for personalized treatment strategies. This is especially important considering the heterogeneous disease course, therapy response, and potentially contributing factors to microbiota perturbations. Moreover, the results of our analysis might help to reconcile inconsistencies in previously published findings as regards involvement of specific bacterial families in CD.

Our results confirm previously identified associations of CD with the families Erysipelotrichaceae, Peptostreptococcaceae, Prevotellaceae, Rikenellaceae, Ruminococcaceae (e.g., *Faecalibacterium prausnitzii*), and Veillonellaceae.^{171, 181, 358} In addition, we also identified previously unreported associations between CD and the families Coriobacteriaceae, Desulfovibrionaceae, Streptococcaceae, and Sutterellaceae. Mixed results (both negative and positive associations) have been reported for the family Bacteroidaceae (e.g., *Bacteroides fragilis*).^{171, 357, 358} Our results suggest that these mixed results can be explained by a change in association (from positive to negative) across the quantile range. Likewise, we found a negative association with the family Pasteurellaceae when the patients were compared to healthy individuals, especially with regard to patients who remained in remission. This is in contrast with previous results showing a positive association between relative abundances of Pasteurellaceae and CD.¹⁷¹

Previous studies did not make use of quantile regression to identify possible associations between the microbiota and factors related to inflammation. Most studies compared samples based on their means or medians (by Student's *t*-test, Mann-Whitney U test, or the analysis of variance), without taking into account the confounding effect of covariates, such as medication use or the age of the patient.^{173, 356-358} Other studies used methods that can take covariates into account, such as generalized linear regression models, but these still only consider the mean count or relative abundance and do not consider distinct associations across patients.^{171, 173, 175, 357} Lastly, supervised classifiers (e.g., Random Forest) and clustering algorithms (e.g., agglomerative hierarchical clustering) are used to predict the presence or activity of disease by the pattern in relative abundance of many families at once.¹⁸¹ While these methods are not constrained by the strict assumptions of regression models, they have difficulties in dealing with repeated measurements and covariates. In addition, these methods require many patients, which are often not available in longitudinal clinical cohorts.

A practical advantage of quantile regression is its usefulness in situations when assumptions of other methods are violated. For example, quantile regression does not require homoscedastic and normally distributed data. On the contrary, the method enables detection and description of changes in the conditional distribution of the response variable when there is heteroscedasticity, skewness, or kurtosis in the data.³⁶¹ Another limitation of quantile regression is that it is hard to use for the purpose of prediction. Nevertheless, quantile regression is powerful when heterogeneous response distributions should be expected, e.g., if many interdependencies and potentially limiting factors play a role. If those co-factors are differently distributed among patients and not included in the model, they lead to (hidden) bias in conventional regression but can be dealt with in quantile regression.³⁷⁰

We found several significant associations of bacterial abundance with the presence of CD. However, associations with disease activity were less evident in our data. Although we found some differences between the two groups of CD patients, we only found significantly elevated abundances for Streptococcaceae (at baseline) and Coriobacteriaceae (during active disease) in patients experiencing an exacerbation relative to patients remaining in remission, and the latter effect disappeared after correction for multiple testing.

Multiple explanations are possible for the lack of significant associations between exacerbation and remission. Firstly, the microbiota of CD patients might not be responsive to exacerbation as compared to remission. Multiple studies underline our finding of no clear significant differences between remission and active disease.^{177, 181} In other words, the observed differences in bacterial abundances are disease-related community differences that even persist in the absence of active inflammation, and therefore this pattern is not significantly reflected between the disease states.^{371, 372}

However, most studies, including ours, also likely lacked the statistical power to find such potential differences. A second possibility is that potential associations are confounded with other factors that are likely to play a role in shaping the microbiota, such as disease severity,¹⁷⁵ disease duration,¹⁷⁷ disease location,¹⁷³ treatments,³⁷³ and host characteristics (such as smoking).¹⁷² Treatments such as PPIs have been shown to change gut microbiota and individuals undergoing surgery will have been given antibiotics to prevent infection.³⁷³ Lastly, we might not have looked at the right taxonomic level, while the resolution of taxonomic profiling could impact the accuracy and specificity of our findings. Most differences are possibly present only at the species or strain levels, or may even require metabolic or functional analysis.^{21, 22, 172} Moreover, a change in relative abundance at taxonomic level might not reflect a change in ecosystem functioning, as expansion in certain species can compensate for the loss of another (functionally similar) species.²² Nonetheless, we did find stronger associations with fecal calprotectin than with the clinical definition of CD activity. This suggests that a quantitative measure of inflammation carries information about the microbial involvement in disease activity. As the level of fecal calprotectin is only a proxy of inflammation in the gut, the associations might become even clearer when specific immunological markers, or even hormones, would be used.

It is important to acknowledge several limitations that may impact the generalizability and interpretation of our findings. Firstly, our study was conducted within a relatively small cohort of CD patients ($n = 57$) and healthy controls ($n = 15$). While this cohort size allowed us to perform longitudinal analyses, it may limit the generalizability of our results to broader CD populations. Also, quantile regression is not insensitive to outliers, especially in the highest and lowest quantiles when there is not much data left for estimation, potentially affecting the robustness of our statistical analyses. With a larger dataset, one might consider a finer quantile division to obtain a smoother quantile regression profile, while with a smaller dataset, a coarser division would be more appropriate. Furthermore, as with any observational study, causation cannot be inferred from our results, and further mechanistic investigations are needed to elucidate whether there is a role of the highlighted bacterial families in CD pathogenesis. Lastly, the dynamic nature of the gut microbiota and potential temporal variations were not extensively explored in this study, which might have limited our ability to capture the full spectrum of microbial changes associated with CD over time. In light of these limitations, caution should be exercised when interpreting our results, and future research hopefully addresses these constraints through larger, more diverse cohorts and a finer taxonomic resolution to provide a more comprehensive understanding of the gut microbiota's role in CD.

In this study, we showed that associations between CD with relative bacterial abundances can be different for subsets of individuals. Our findings revealed significant negative associations with CD for several bacterial families, such as Pasteurellaceae, Peptostreptococcaceae, Prevotellaceae, and Ruminococcaceae, highlighting their potential roles in CD pathogenesis.

Furthermore, the significant differences in the relative abundance of Sutterellaceae and Streptococcaceae among CD patients who experienced exacerbations, relative to those who maintained remission, had not been seen before and underscore the dynamic nature of microbial associations in relation to disease activity. The subtle variations observed in the family Coriobacteriaceae, which could not be seen in the mean response, further emphasize the complexity of these relationships.

Importantly, our study underscores the heterogeneity of CD and its impact on gut microbiota, suggesting that associations may only become evident when considering patients' diverse disease courses, medication histories, therapy responses, and gut microbiota compositions. Associations with specific bacterial families may only be detectable in a minority of patients, hence they cannot generally be considered to identify CD or disease activity. The novelty of our study lies in its rigorous approach to exploring associations in subsets of patients, acknowledging the heterogeneity between them. In such situations, quantile regression is a useful tool for distilling potential relationships that may remain unidentified by commonly used methods. We recommend its use in even larger cohorts, to gain a better understanding of CD in relation to the gut microbiota.

Appendices of Chapter 4

Appendix Table 4.1 - Clinical and demographic information of healthy controls and CD patients (RR and RE).

Note that two samples per individual ($n = 72$ individuals) were collected ($n = 144$ samples).

	HC ($n = 30$)	RR ($n = 70$)	RE ($n = 44$)	Overall ($n = 144$)
Sex				
Female	14 (46.7%)	50 (71.4%)	24 (54.5%)	88 (61.1%)
Male	16 (53.3%)	20 (28.6%)	20 (45.5%)	56 (38.9%)
Smoking				
Ex	4 (13.3%)	34 (48.6%)	20 (45.5%)	58 (40.3%)
Never	26 (86.7%)	20 (28.6%)	20 (45.5%)	66 (45.8%)
Current	0 (0%)	16 (22.9%)	4 (9.1%)	20 (13.9%)
Age				
Mean (SD)	26.7 (5.93)	42.6 (12.8)	43.6 (16.8)	39.6 (14.7)
Median (min, max)	25.0 (20.0, 45.0)	43.0 (17.0, 67.0)	42.5 (19.0, 68.0)	38.5 (17.0, 68.0)
Disease location^a				
C	NA	16 (22.9%)	14 (31.8%)	30 (20.8%)
I	NA	24 (34.3%)	14 (31.8%)	38 (26.4%)
IC	NA	30 (42.9%)	16 (36.4%)	46 (31.9%)
Age at diagnosis^b				
A2	NA	62 (88.6%)	28 (63.6%)	90 (62.5%)
A3	NA	8 (11.4%)	16 (36.4%)	24 (16.7%)

Table continued on next page

Surgery				
0	30 (100%)	54 (77.1%)	36 (81.8%)	120 (83.3%)
1	0 (0%)	16 (22.9%)	8 (18.2%)	24 (16.7%)
Phenotype				
0	NA	52 (74.3%)	24 (54.5%)	76 (52.8%)
1	NA	18 (25.7%)	20 (45.5%)	38 (26.4%)
Mesalazines				
0	30 (100%)	60 (85.7%)	35 (79.5%)	125 (86.8%)
1	0 (0%)	10 (14.3%)	9 (20.5%)	19 (13.2%)
Thiopurines				
0	30 (100%)	46 (65.7%)	28 (63.6%)	104 (72.2%)
1	0 (0%)	24 (34.3%)	16 (36.4%)	40 (27.8%)
Biologicals^c				
0	30 (100%)	32 (45.7%)	15 (34.1%)	77 (53.5%)
1	0 (0%)	38 (54.3%)	29 (65.9%)	67 (46.5%)
Induction				
0	30 (100%)	57 (81.4%)	36 (81.8%)	123 (85.4%)
1	0 (0%)	13 (18.6%)	8 (18.2%)	21 (14.6%)
PPI^d				
0	30 (100%)	56 (80.0%)	28 (63.6%)	114 (79.2%)
1	0 (0%)	14 (20.0%)	16 (36.4%)	30 (20.8%)
HBI^e				
Mean (SD)	NA	2.41 (2.95)	3.00 (3.39)	2.08 (2.97)
Median (min, max)	NA	1.00 (0, 11.0)	2.00 (0, 13.0)	1.00 (0, 13.0)
Missing	NA	0 (0%)	2 (4.5%)	2 (1.4%)
CRP^f				
Mean (SD)	NA	2.57 (2.05)	3.76 (3.28)	2.31 (2.61)
Median (min, max)	NA	2.00 (0.9, 11.0)	2.80 (0.9, 13.0)	1.40 (0, 13.0)
Missing	NA	5 (7.1%)	8 (18.2%)	13 (9.0%)
FC^g				
Mean (SD)	NA	28.6 (20.9)	290 (742)	102 (426)
Median (min, max)	NA	14.0 (14.0, 98.0)	110 (14.0, 4900)	15.0 (0, 4900)

^a Disease location: colonic (C), ileal (I), or ileocolonic (IC)

^b Age at diagnosis: younger than 40 years (A2) or older than 40 years (A3)

^c All biological treatments concerned anti-TNF therapy

^d PPI: proton pump inhibitors

^e HBI: Harvey Bradshaw index (Appendix Figure 4.1)

^f CRP: serum C-reactive protein (Appendix Figure 4.1)

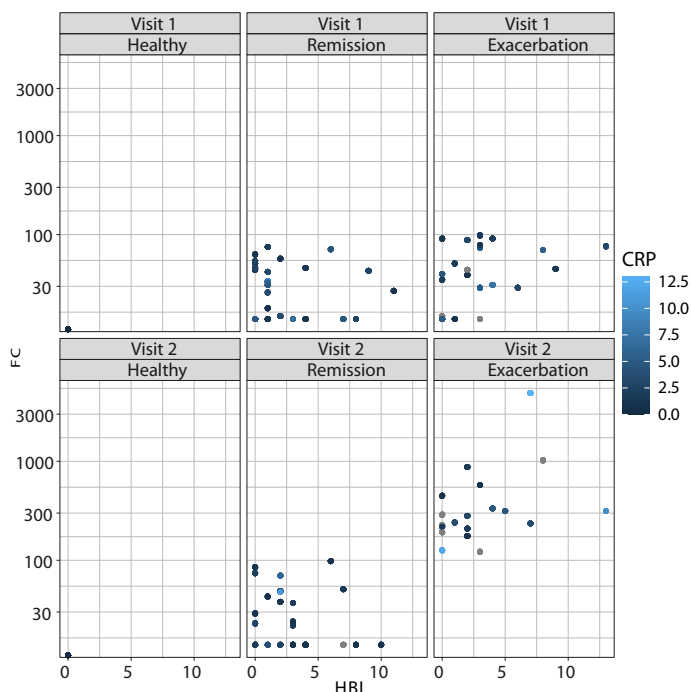
^g FC: fecal calprotectin (Appendix Figure 4.1)

Appendix Table 4.2 - Medication use and time between sampling moments for remission and exacerbation samples.

	RR (n = 35)		RE (n = 22)	
	Remission	Remission	Remission	Exacerbation
Medication ^a				
Mesalazine	5 (14.3%)	5 (14.3%)	4 (18.2%)	5 (22.7%)
Thiopurines	11 (31.4%)	11 (31.4%)	9 (40.9%)	7 (31.8%)
Biologicals	18 (51.4%)	19 (54.3%)	13 (59.1%)	15 (68.2%)
Corticosteroids	1 (2.9%)	0 (0%)	1 (4.5%)	1 (4.5%)
PPI	7 (20%)	7 (20%)	8 (36.4%)	8 (36.4%)
Antibiotics ^b	1 (2.9%)	0 (0%)	1 (4.5%)	0 (0%)
Time between sampling moments (week, median, IQR)	14 (11–21)		20 (11–21)	

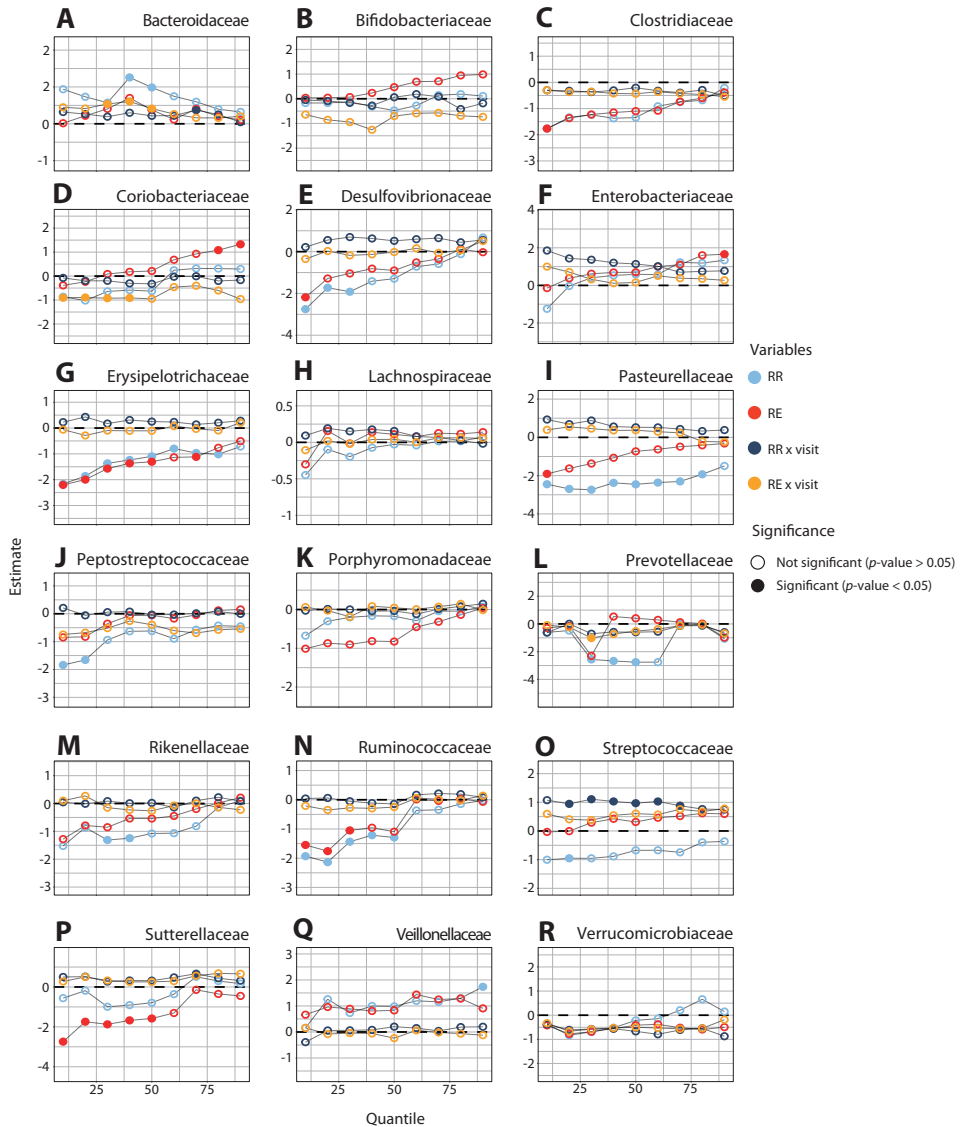
^a Six RR and five RE patients had a medication change between consecutive samples during the study period. In the RR group, mesalazine was stopped by one patient, prednisone by one patient and biologicals by two patients, while one patient started mesalazine and one patient started with biologicals. In the RE group, two patients started with biologicals, two patients stopped with thiopurines, and one patient started with mesalazine.

^b Ciprofloxacin and cotrimoxazole were used two and one month prior to sample collection, respectively.



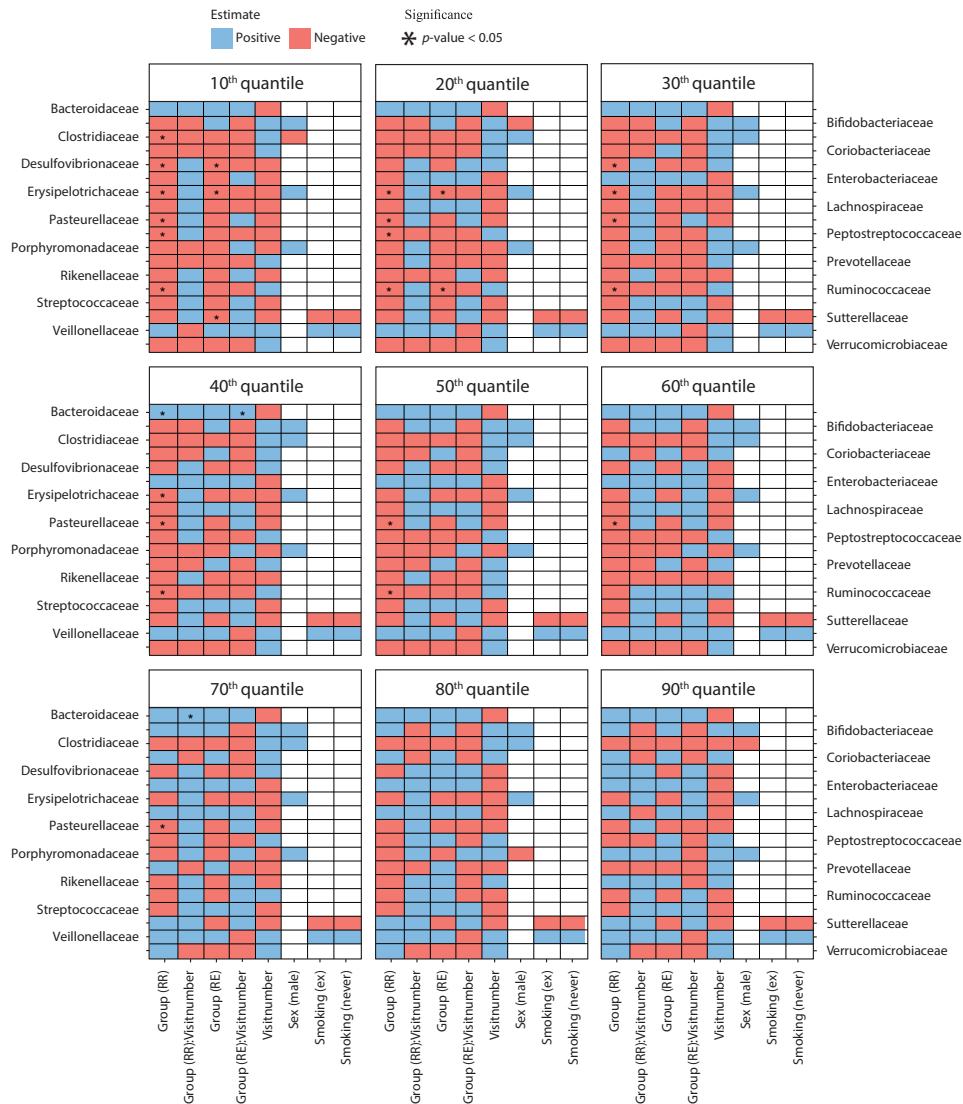
Appendix Figure 4.1 - Disease indicators (FC, serum CRP, and HBI).

Remission at baseline was defined by FC < 100 µg/g and CRP < 5 mg/L or FC < 100 µg/g, CRP < 10 mg/L, and HBI ≤ 4. Disease activity at the second visit was defined by FC, serum CRP, and HBI, i.e., FC > 250 µg/g or FC > 100 µg/g with at least a 5-fold increase from baseline.

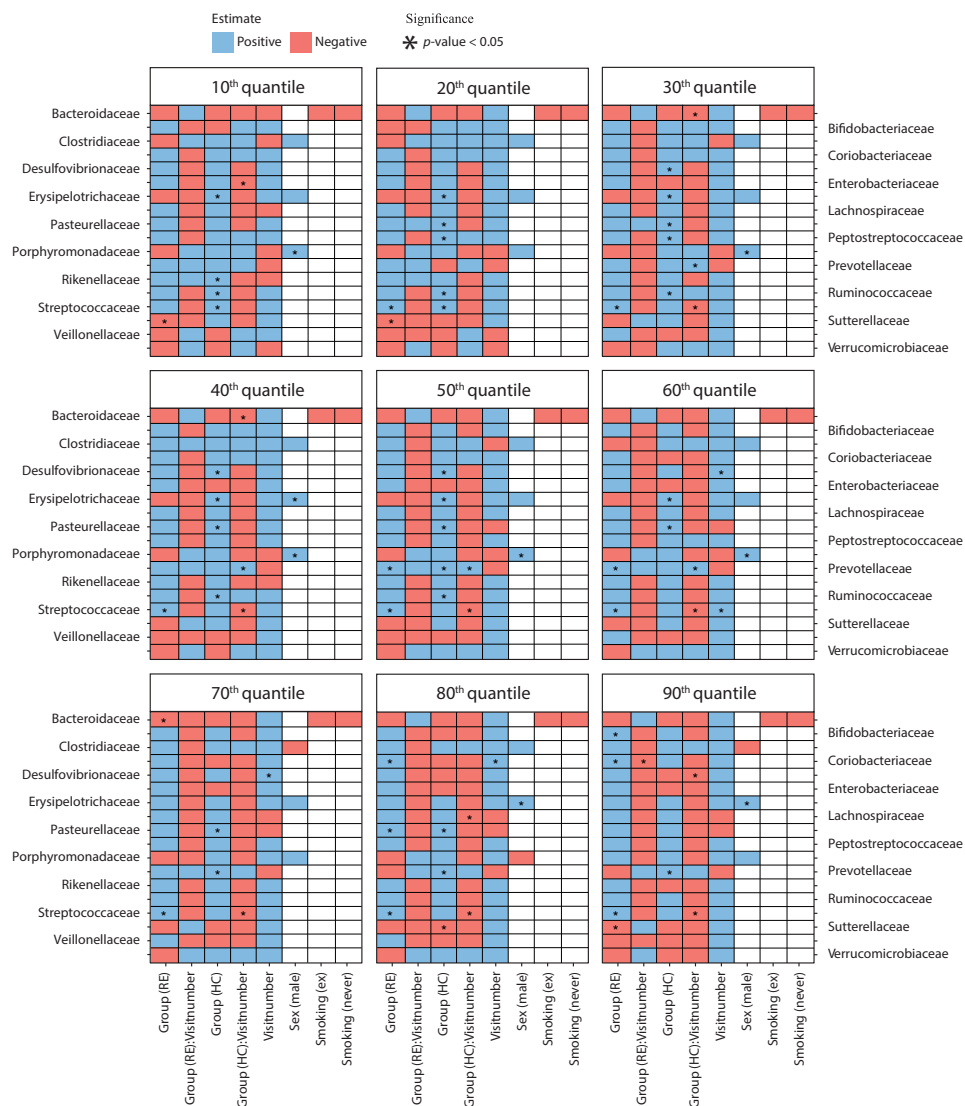


Appendix Figure 4.2 - Quantile regression profile plots for all base case families. The various dots represent sample estimates (y-axis) of differences in relative abundance compared to healthy controls across the 10th to the 90th quantile (x-axis). Differences at baseline (visit 1) are visualized in light blue (RR) and red (RE), while the interaction with visit number (in dark blue and orange) displays the difference in changes over time. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance at that particular quantile. Significant variables ($p\text{-value} < 0.05$) are indicated with a closed circle.

Sensitivity analyses: Differences in abundance between healthy individuals and CD patients

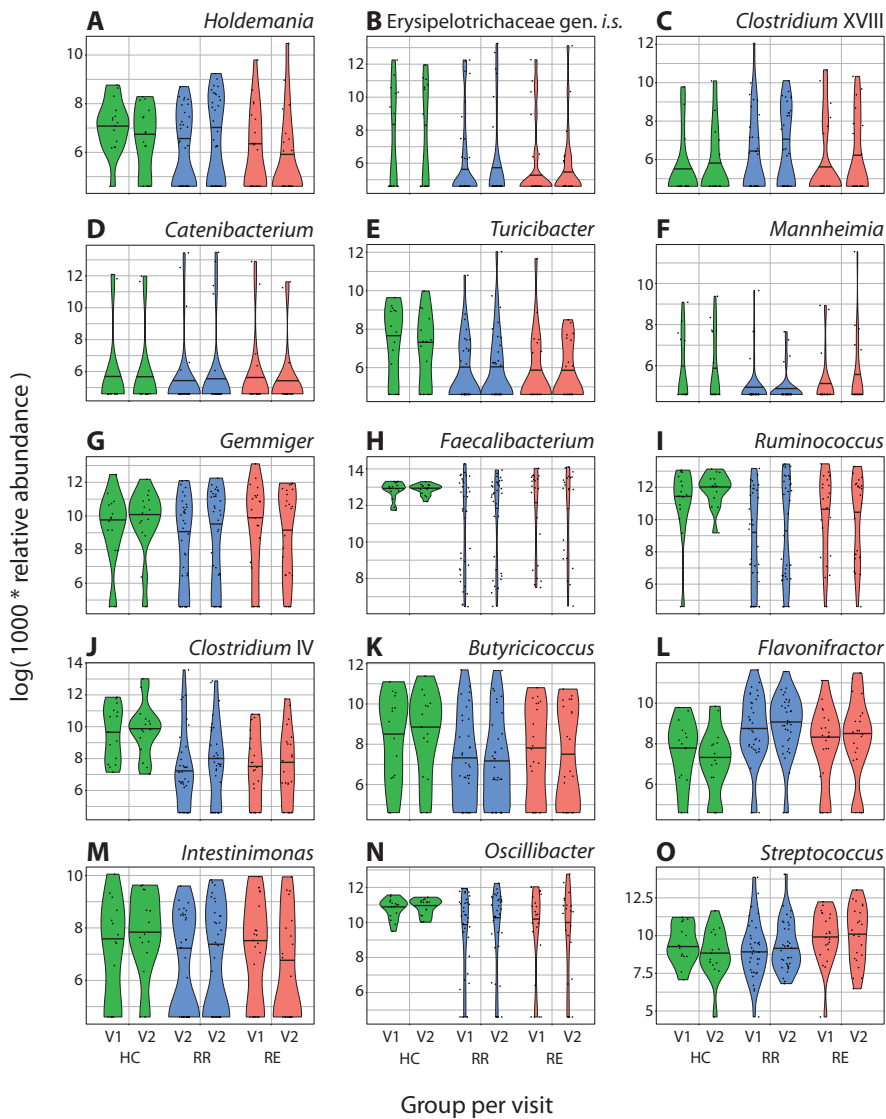


Appendix Figure 4.3 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and common variables, with p -values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05 after BH adjustment) are indicated with an asterisk (*).

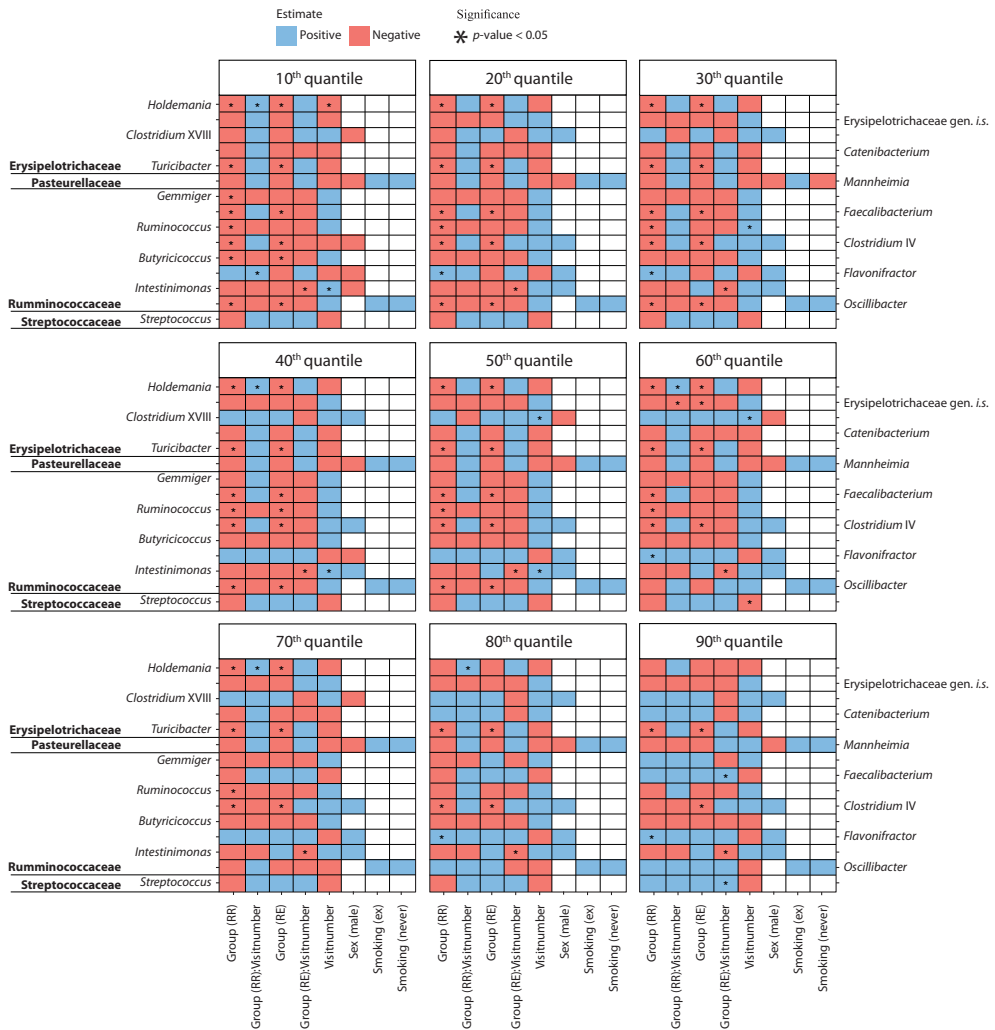


Appendix Figure 4.4 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and common variables, using the RR group as reference instead of the HC group.

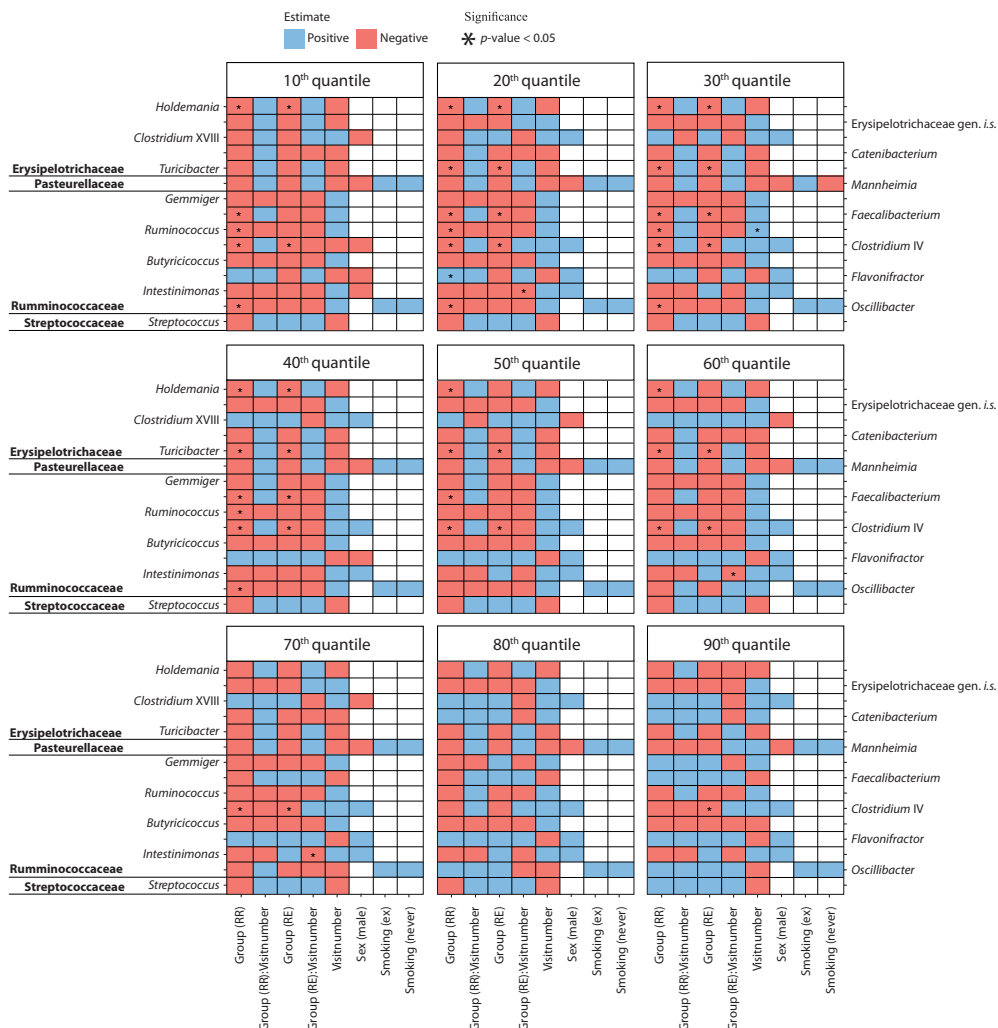
The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05) are indicated with an asterisk (*).



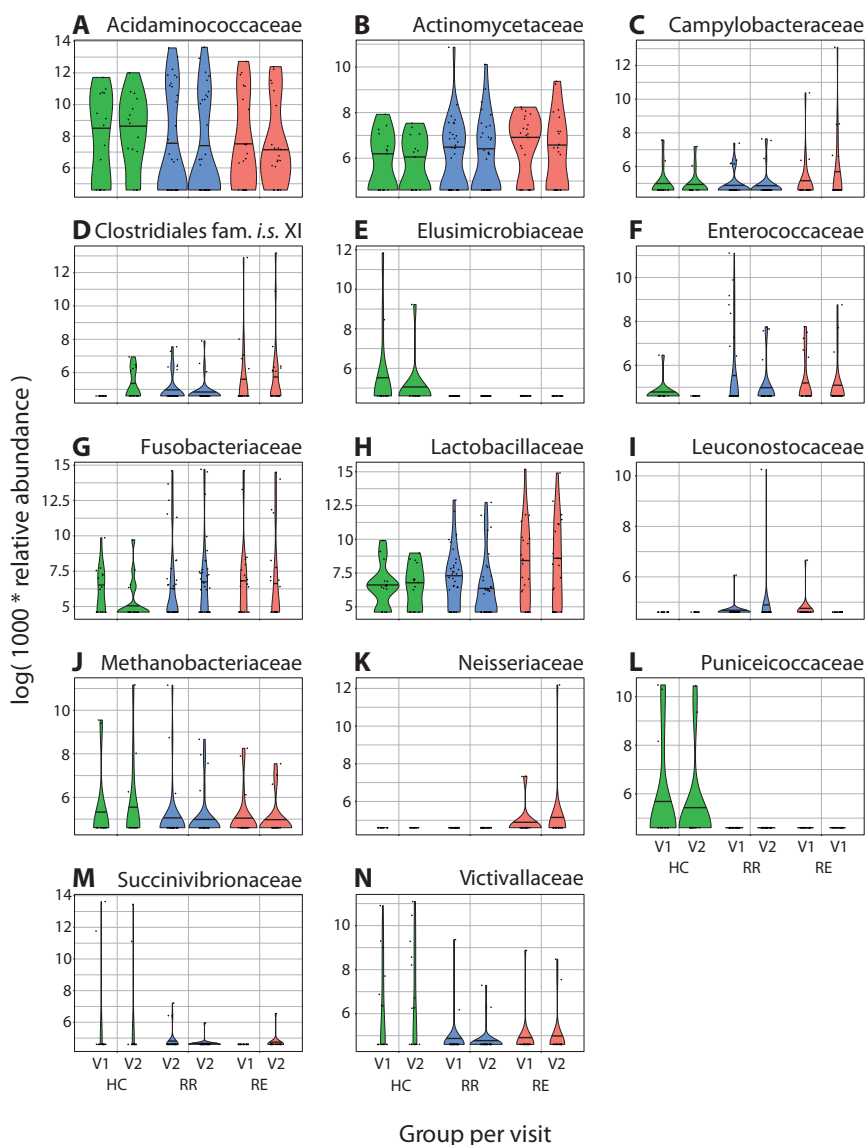
Appendix Figure 4.5 - Violin plots of transformed relative abundances for selected genera by group and time point. In green the healthy controls, in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50th quantile is shown with a black horizontal line.



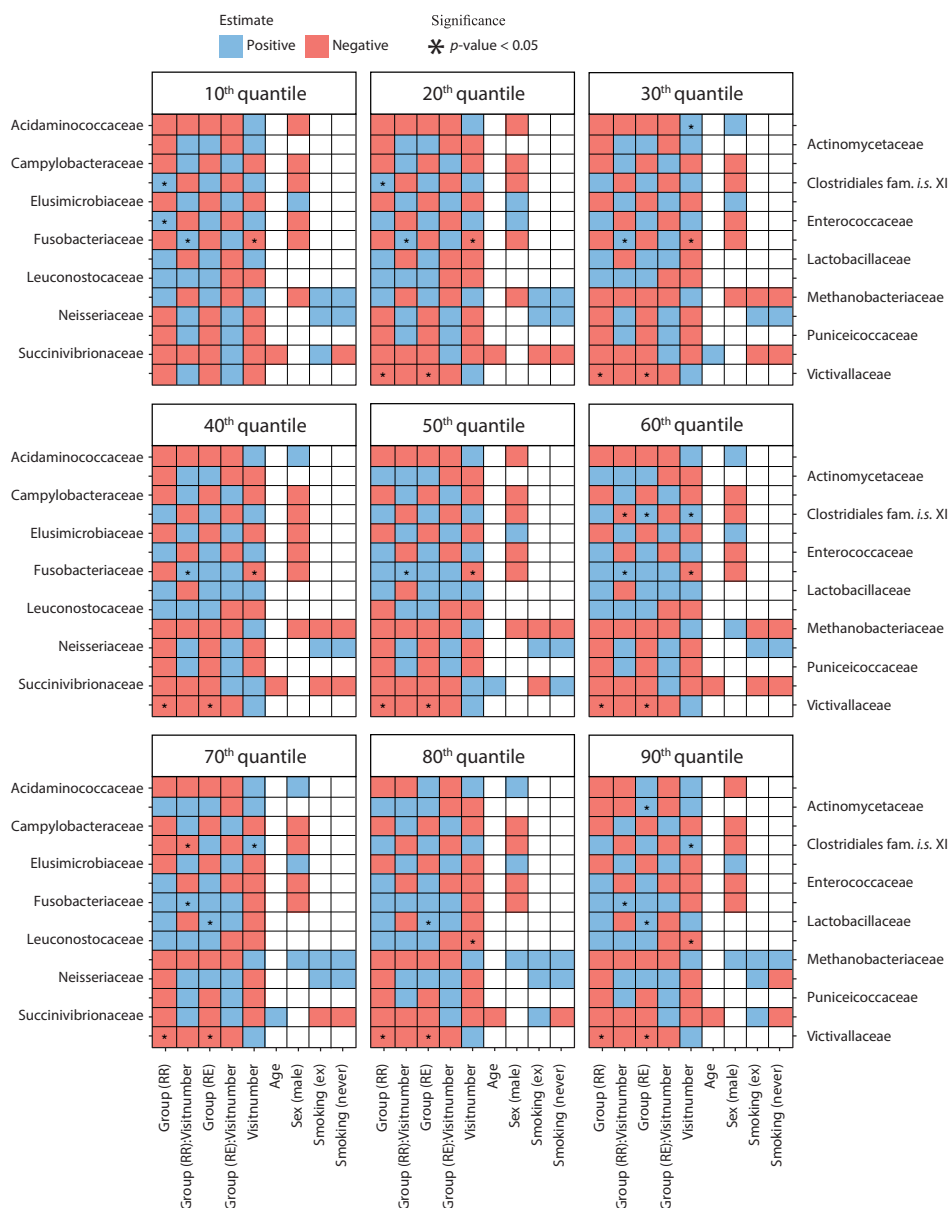
Appendix Figure 4.6 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and common variables. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05) are indicated with an asterisk (*).



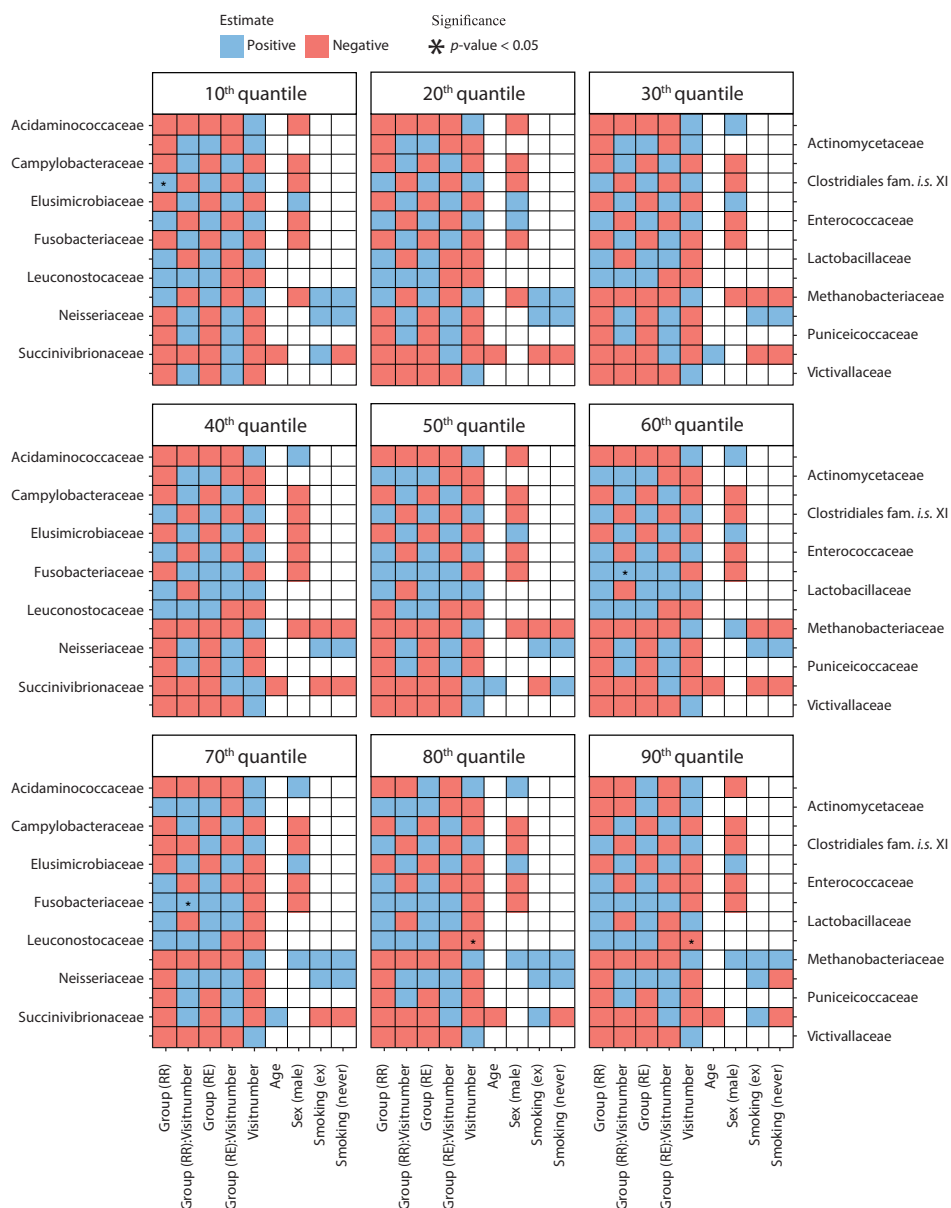
Appendix Figure 4.7 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and common variables, with p -values adjusted using the BH procedure. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05 after BH adjustment) are indicated with an asterisk (*).



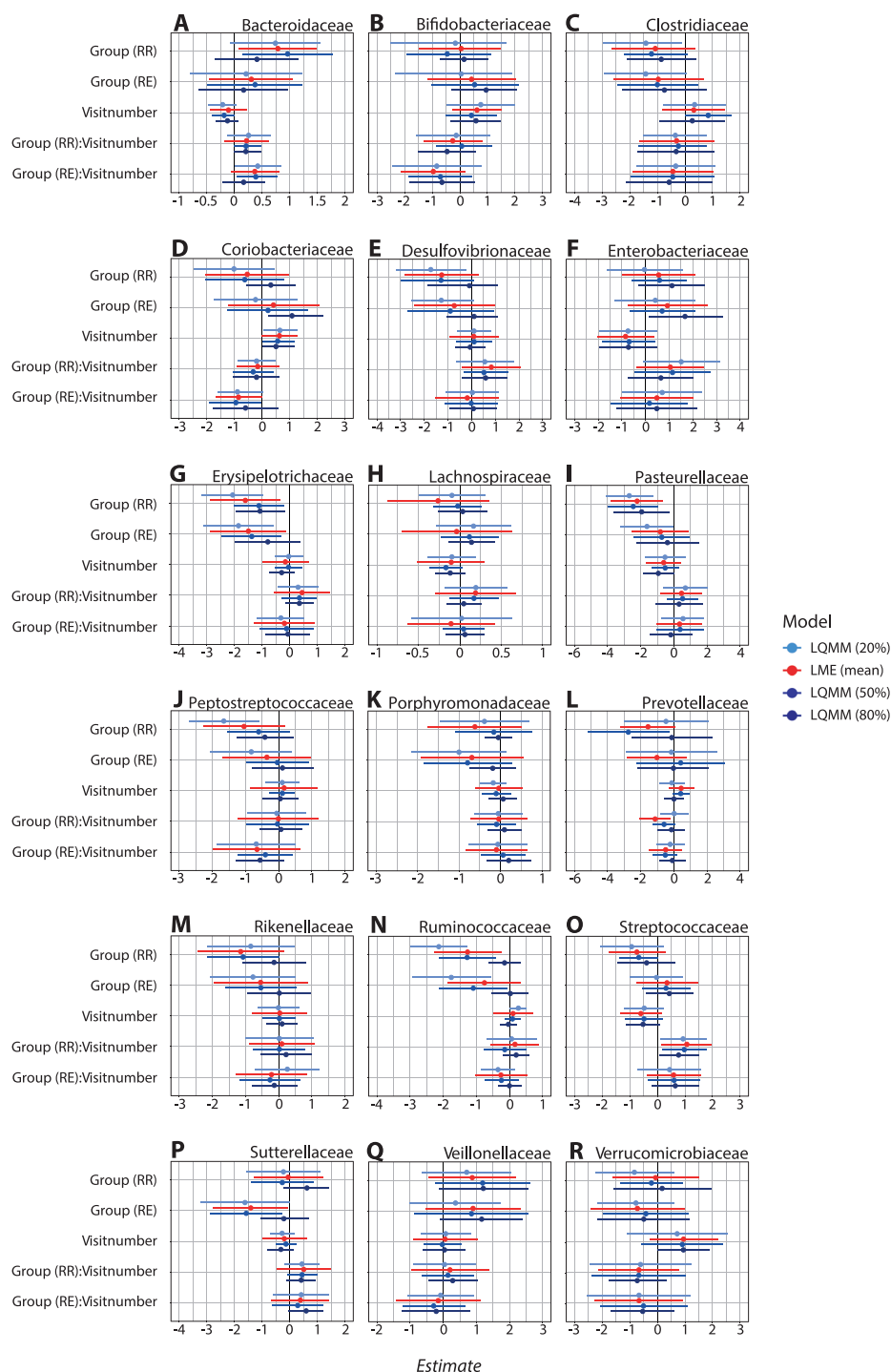
Appendix Figure 4.8 - Violin plots of transformed relative abundances for families outside the base case selection criterion by group and time point. In green the healthy controls, in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50th quantile is shown with a black horizontal line.



Appendix Figure 4.9 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and common variables. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$) are indicated with an asterisk (*).

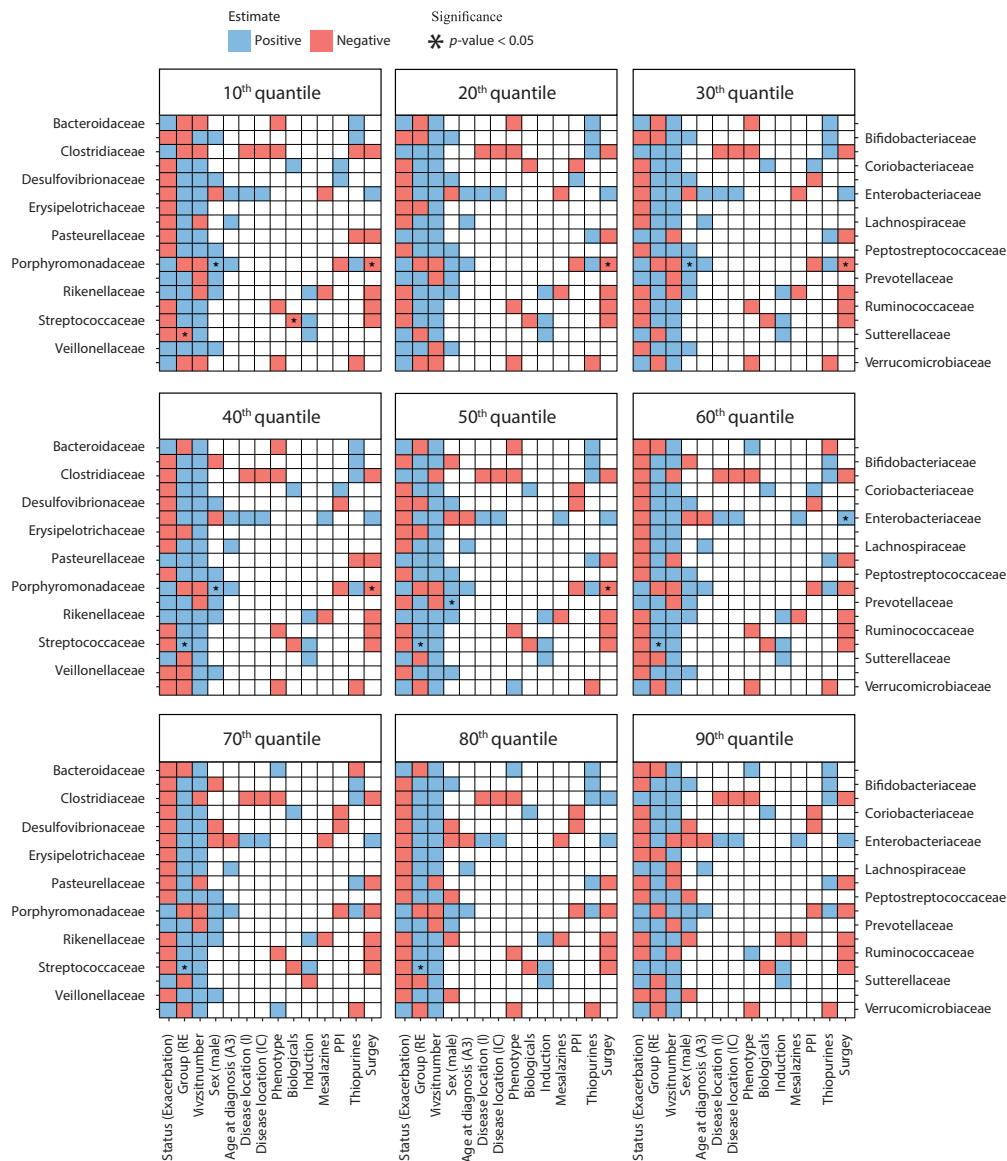


Appendix Figure 4.10 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and common variables, with p -values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$ after BH adjustment) are indicated with an asterisk (*).

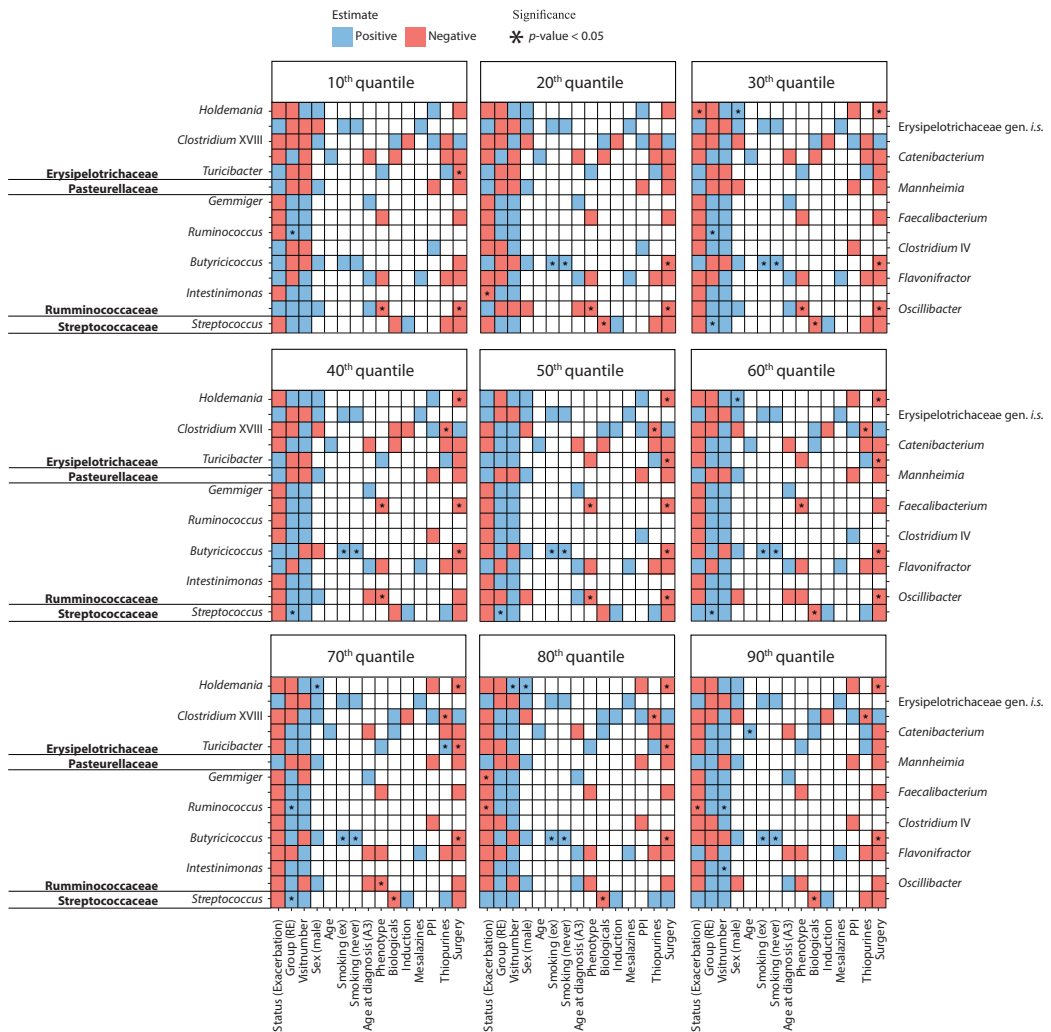


Appendix Figure 4.11 - Comparison of the LQMM analysis results (20%, 50%, and 80% quantiles) with linear mixed-effects model results. The point estimates, 95% confidence intervals, and a reference line at 0 (in black) are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different (p -value < 0.05) from 0.

Sensitivity analyses: Gut microbiota changes in relation to Crohn's disease activity

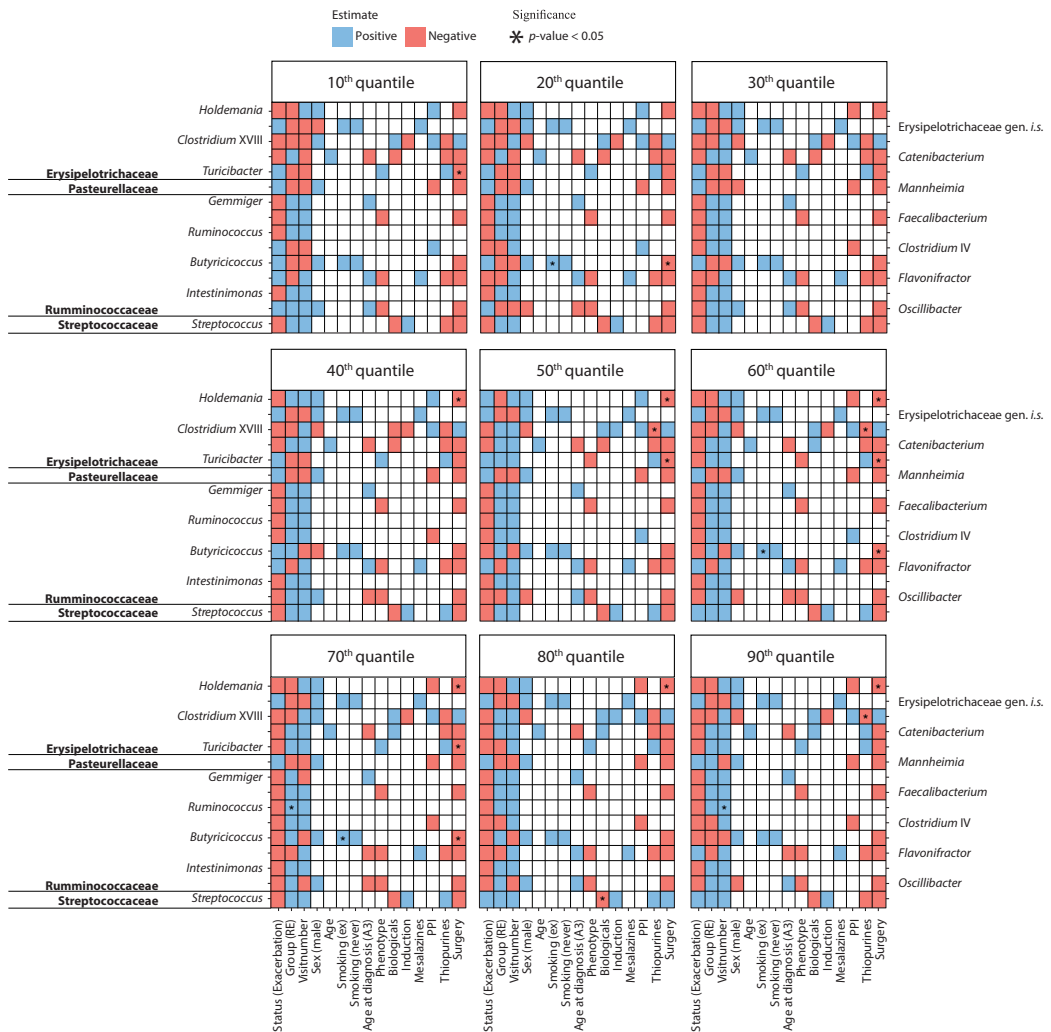


Appendix Figure 4.12 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and clinical variables in CD patients only, with p -values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05 after BH adjustment) are indicated with an asterisk (*).

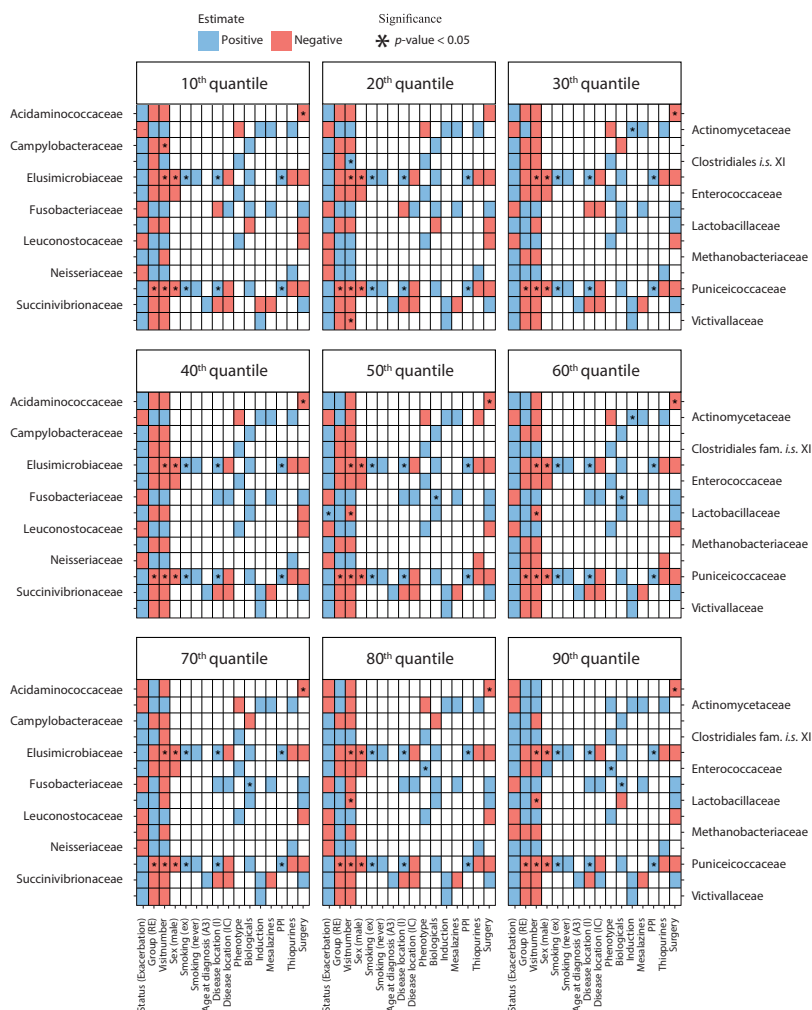


Appendix Figure 4.13 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and clinical variables in CD patients only.

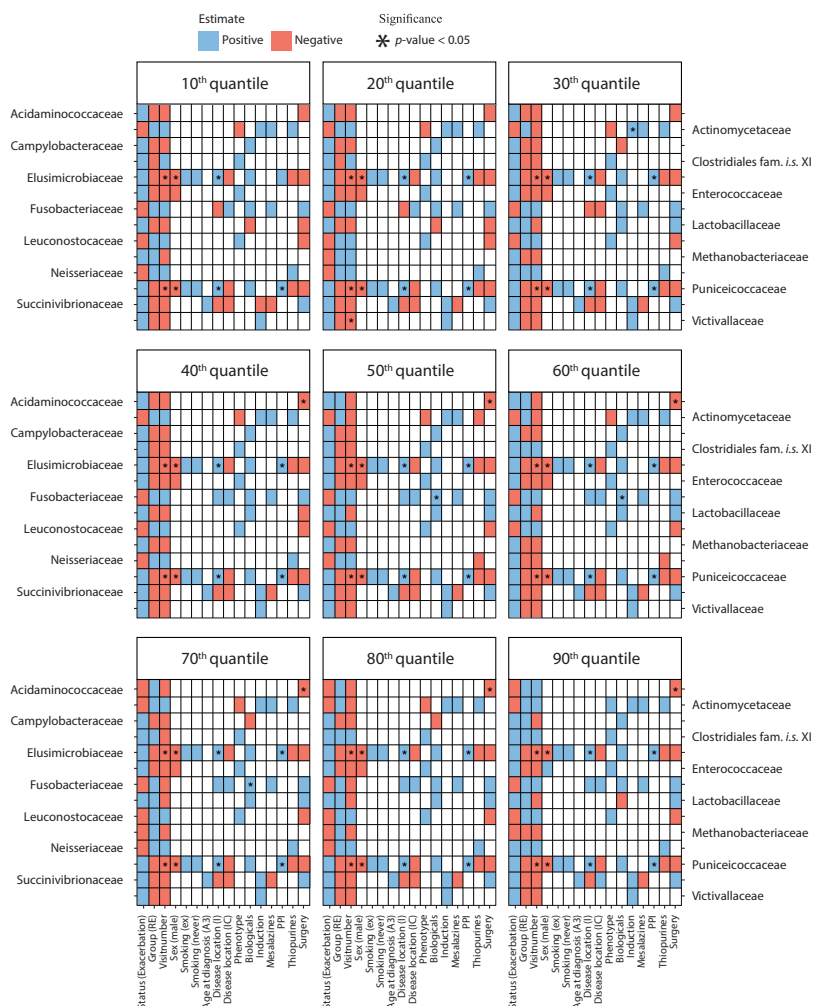
The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$) are indicated with an asterisk (*).



Appendix Figure 4.14 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and clinical variables in CD patients only, with p -values adjusted using the BH procedure. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$ after BH adjustment) are indicated with an asterisk ("*").

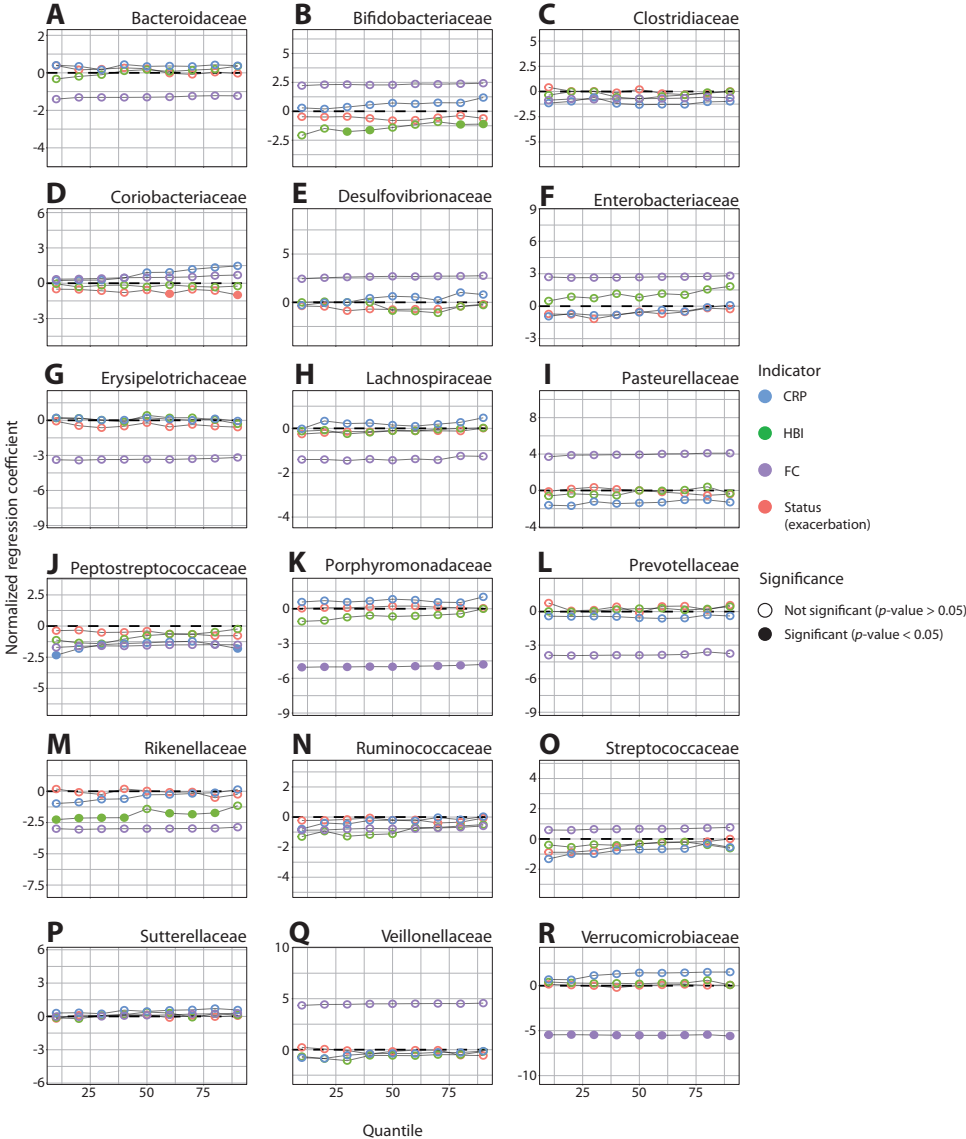


Appendix Figure 4.15 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and clinical variables in CD patients only. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables ($p\text{-value} < 0.05$) are indicated with an asterisk (*).

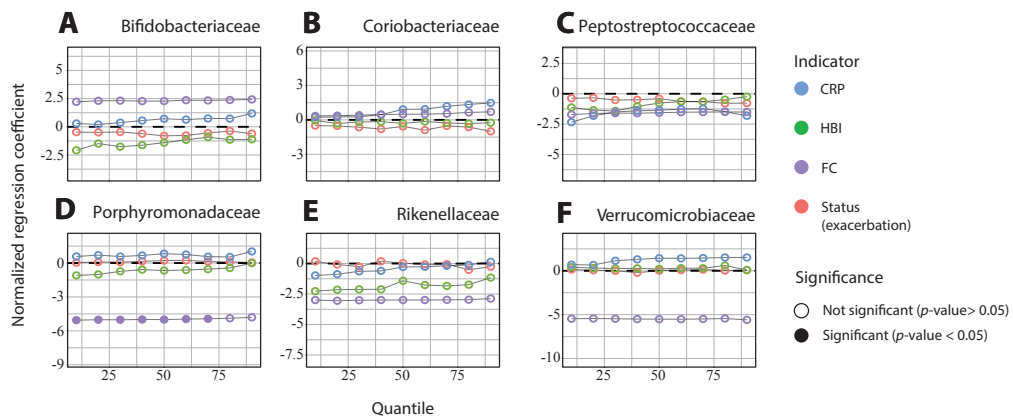


Appendix Figure 4.16 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and clinical variables in CD patients only, with p -values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p -value < 0.05 after BH adjustment) are indicated with an asterisk (*).

Sensitivity analyses: relative abundances of bacterial families in relation to different disease activity indicators

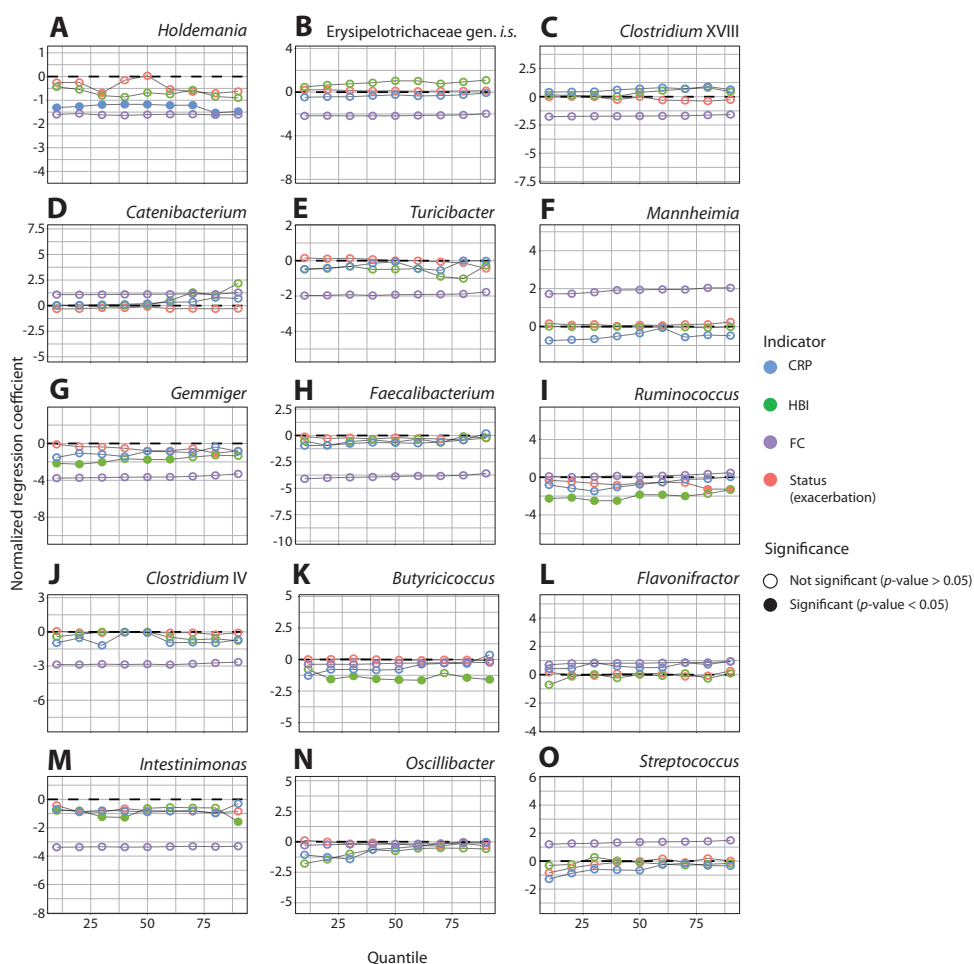


Appendix Figure 4.17 - Quantile regression profile plots for different disease activity indicators and clinical variables for all base case families in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p -value < 0.05) are indicated with a closed circle.

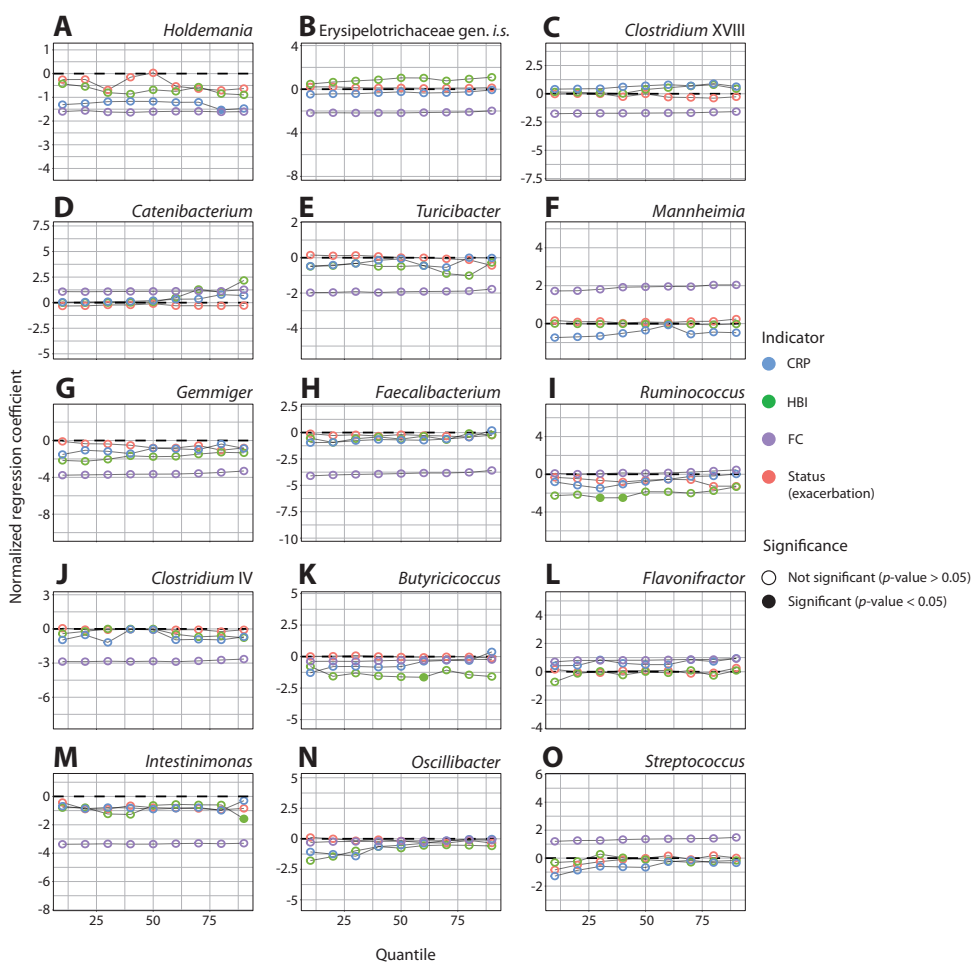


Appendix Figure 4.18 - Quantile regression profile plots for different disease activity indicators and clinical variables for base case families in CD patients only, with p-values adjusted using the BH procedure.

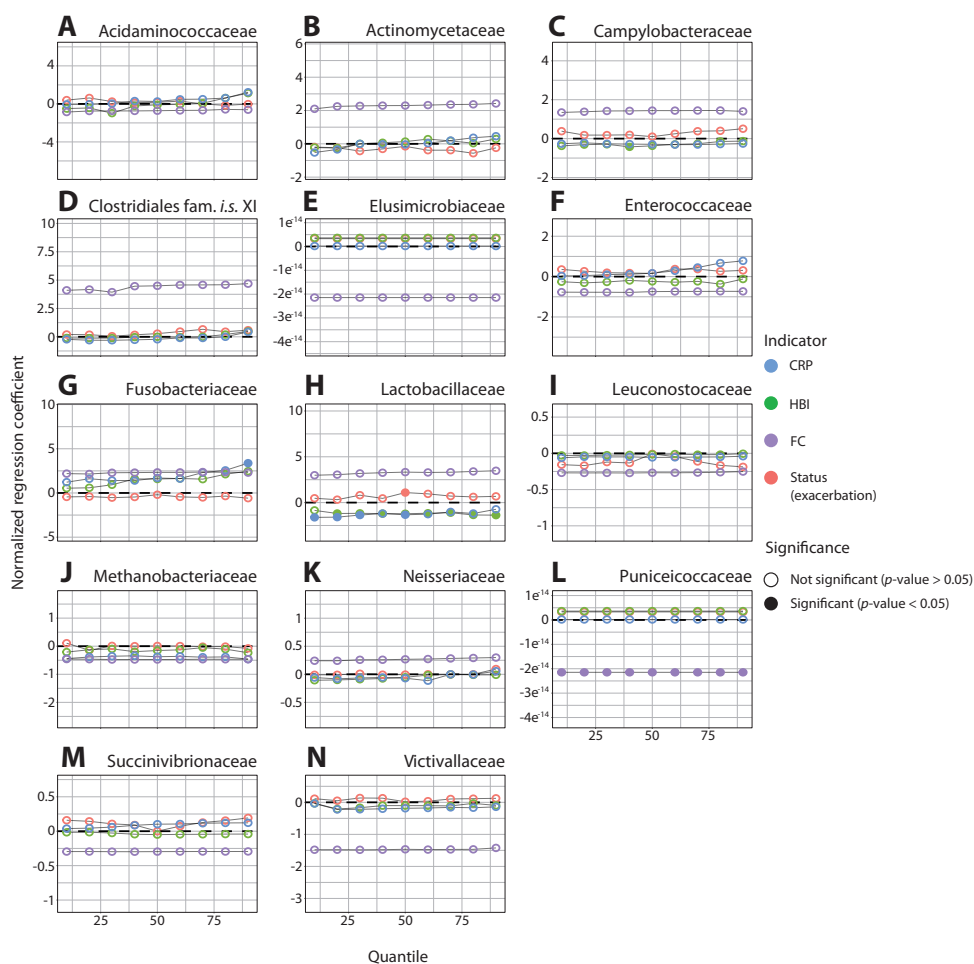
The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables ($p\text{-value} < 0.05$ after BH adjustment) are indicated with a closed circle.



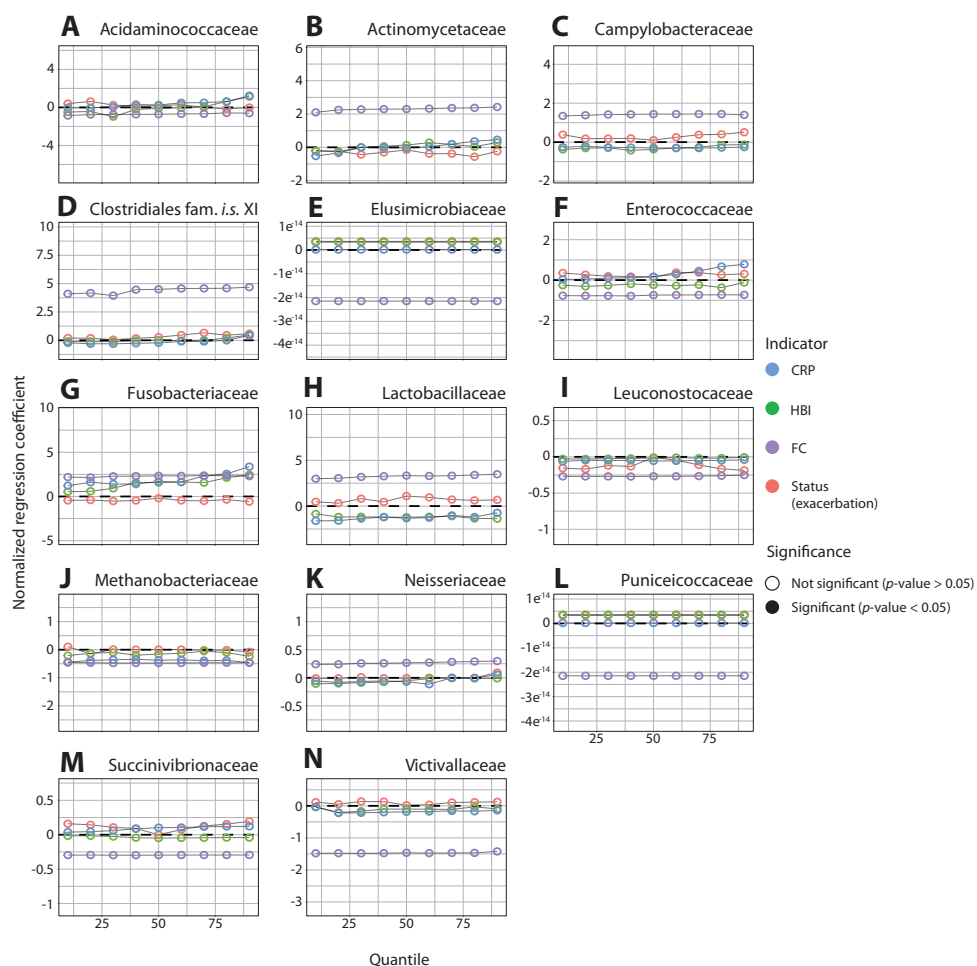
Appendix Figure 4.19 - Quantile regression profile plots for different disease activity indicators and clinical variables for selected genera in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p -value < 0.05) are indicated with a closed circle.



Appendix Figure 4.20 - Quantile regression profile plots for different disease activity indicators and clinical variables for selected genera in CD patients only, with p-values adjusted using the BH procedure. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p -value < 0.05 after BH adjustment) are indicated with a closed circle.



Appendix Figure 4.21 - Quantile regression profile plots for different disease activity indicators and clinical variables for families outside the base case selection criterion in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables ($p\text{-value} < 0.05$) are indicated with a closed circle.



Appendix Figure 4.22 - Quantile regression profile plots for different disease activity indicators and clinical variables for families outside the base case selection criterion in CD patients only, with p-values adjusted using the BH procedure.

The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p -value < 0.05 after BH adjustment) are indicated with a closed circle.