medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Original research

Phenotyping and prediction of paediatric abdominal pain: A machine learning approach

Kazuya Takahashi[1,2], Michalina Lubiatowska[1], Huma Shehwana[3], James K. Ruffle[1,4], John A Williams[5,6,7], Animesh Acharjee[5,6,7], Shuji Terai[2], Georgios V Gkoutos[5,6,7], Humayoon Satti[3], and Qasim Aziz[1]

1. Centre for Neuroscience and Trauma, Wingate Institute of Neurogastroenterology, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

2. Division of Gastroenterology and Hepatology, Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan.

3. Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Pakistan.

4. Queen Square Institute of Neurology, University College London, UK

5. College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK.

6. Health Data Research UK, Midlands Site, Birmingham, UK.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

7. Centre for Health Data Science, Birmingham, UK

H.Satti and Q.Aziz are joint senior authors.

Corresponding Author: Qasim Aziz

Centre for Neuroscience and Trauma, Wingate Institute of Neurogastroenterology, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

Tel: +44 2078822655

E-mail: q.aziz@qmul.ac.uk

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Due to patient heterogeneity, the exact mechanisms of paediatric abdominal pain (AP) remain unknown. We sought to resolve this by identifying paediatric AP phenotypes and developing predictive models to determine their associated factors. In 13,790 children from a large birth cohort, the frequencies of paediatric and maternal demographics and comorbidities were catalogued from general practitioner records. Unsupervised machine learning clustering was used to identify phenotypes of paediatric AP with shared characteristics. Predictive paediatric AP models were constructed using paediatric and maternal demographics and comorbidities.

1,274 children experienced AP (9.2 %) (average age: 8.4 ± 1.1 years old, male/female: 615/659), who clustered into 3 distinct phenotypes: phenotype 1 with an allergic predisposition (n = 137), phenotype 2 with maternal comorbidities (n = 676), and phenotype 3 with minimal other comorbidities (n = 340). As the number of allergic diseases or maternal comorbidities increased, so did the frequency of AP, with 17.6% of children with ≥ 3 allergic diseases and 25.6% of children with ≥ 3 maternal comorbidities. Furthermore, in high-risk children who met both ≥ 3 allergic diseases and ≥ 3 maternal comorbidities, 30.8% had AP. Predictive models demonstrated modest fidelity in predicting paediatric AP (AUC 0.66), showing that a child's

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ethnicity and paediatric/maternal comorbidities were strongly predictive factors. Our findings reveal distinct phenotypes and associated factors of paediatric AP, suggesting targets for future research to elucidate the exact mechanisms of paediatric AP related to allergic diseases, ethnicity, and maternal comorbidities.

Keywords: abdominal pain, machine learning, artificial intelligence, Uniform Manifold Approximation and Projection,

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1. Introduction

Abdominal pain (AP) is one of the most common symptoms among children and adolescents, with prevalence rates across the USA and Europe ranging from 0.3% to 19.0% [1–3]. In primary care, children presenting with AP are diagnosed with functional AP or non-medically explained in 80% of the cases, while organic causes are considerably less frequent [4,5]. According to the internationally recognized Rome Criteria, functional AP is now described as disorders of gut-brain interaction (DGBI) [6]. Though rarely life-threatening, AP is often refractory to treatment and associated with psychiatric comorbidities, such as anxiety and depressive disorders [7–9], significantly affecting health-related quality of life [10]. Therefore, early recognition and intervention are warranted.

Organic diseases, as well as early life events ranging from psychological abuse, allergy status, psychological comorbidity, and parental factors, are proposed risk factors for paediatric AP [5,11–16]. Moreover, these diverse risk factors can influence each other, contributing to its complexity and heterogeneity. A more robust stratification with large patient cohorts is essential to understanding the aetiology of paediatric AP, plausibly disclosing novel insights into its underlying mechanisms.

A robust approach for identifying subgroups of patients with shared characteristics is

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

data-driven clustering by unsupervised machine learning (ML) [17–20]. Any yielded subgroups may share an underlying mechanism associated with AP. Furthermore, supervised ML is considered a powerful tool for clinical outcome prediction [21,22], and it could aid clinicians in assessing the risk of AP development in early childhood [23]. We hypothesised that an ML algorithm could phenotype children with AP according to common routinely available characteristics and thus help unravel the complex underlying disease mechanisms. Additionally, we hypothesised that ML algorithms could predict the development of paediatric AP and help identify the important factors associated with it.

In this study, we systematically evaluated the risk factors of paediatric AP in a large birth cohort. Using unsupervised ML, we delineated phenotypes of paediatric AP using paediatric and maternal clinical data. ML models tasked to predict the development of paediatric AP moreover reveal underlying factors linked to its frequency, catalytic for future hypothesis-generating research.

2. Methods

2.1. Participants

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Between 2007 and 2011, 12,453 pregnant women (recruited at 26 - 28 weeks) and 13,858 children were registered in the Born in Bradford (BiB) cohort. The detailed demographics of the whole cohort are described elsewhere [24]. For this study, we included children who had linked general practitioner (GP) records and whose comorbidities could be identified using the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT). Additionally, the clinical records of their mothers were included in the analysis. The details of the study population are illustrated in **Figure 1**.

2.2. Definition of AP and frequency survey of paediatric and maternal comorbidities

AP was defined as the presence of one of the following diagnoses in SNOMED-CT, regardless of its cause or whether it was acute or chronic: AP, abdominal wall pain, or generalized AP. The frequencies of comorbidities, such as gastrointestinal (GI), psychological, and allergic diseases (asthma, eczema, urticaria, or hay fever), as well as diseases causing somatic pain, were also retrieved using SNOMED-CT, since these are associated with both organic and functional AP [11,13–15,25–27]. Additionally, the comorbidities of the children's mothers were extracted to identify any associations between the maternal comorbidities and paediatric AP. The diseases of the children

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and their mothers investigated in this study and their SNOMED-CT codes are listed in **Supplemental digital content 1**. Fathers were not included in our analysis pipeline due to a significant proportion of missing values (exceeding 74.7%).

2.3. Study 1. Identifying factors associated with paediatric AP

In Study 1, we aimed to determine the frequency of AP, elucidate the clinical characteristics of children with AP, and identify its associated factors. First, we statistically compared the demographics (gender, ethnicity, and mode of delivery) and frequency of comorbidities between children with and without AP. Then, we identified factors associated with paediatric AP from paediatric and maternal demographics and comorbidities. In Study 1, all comorbidities listed in the GP records were used to assess the association between AP and comorbidities, regardless of whether they occurred before or after the onset of AP.

2.4. Study 2. Phenotyping paediatric AP using unsupervised ML clustering

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

To elucidate the mechanisms of paediatric AP, we deemed patient stratification crucial. Study 2 focused on children experiencing AP, utilising unsupervised ML clustering to delineate deep phenotypes. In Python (version 3.7.12) [28], Uniform Manifold Approximation and Projection (UMAP), a non-linear dimension reduction technique, was used to embed all variables for subsequent clustering into a three-dimensional latent space [29]. Scatterplots were generated based on these latents, and phenotypic AP clusters were identified using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), a density-based hierarchical method [30]. Subsequently, demographics and comorbidities of children and their mothers were compared across the identified phenotypes to elucidate their distinct characteristics. Additionally, we examined the frequency of AP and the quantitative burden of significant comorbidities associated with AP identified in the comparison between the phenotypes.

2.5. Study 3. Using supervised ML to predict the development of paediatric AP

To facilitate early intervention through the prediction of AP onset, we developed extreme gradient-boosted tree (XGB) classifiers, a highly effective and well-validated

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

supervised ML algorithm extensively detailed elsewhere [31]. Given the rarity of children with AP compared to those without, data imbalance could potentially reduce prediction model performance. In line with common research practice, we initially randomly selected an equal number of children without AP from the control group to match those with AP (**Fig. 1**).

For predicting AP, model inputs included demographic and pre-existing comorbidities diagnosed before the onset of AP. Rare comorbidities with a frequency of less than 1% were excluded to prevent overfitting. The variables used for predictive models and their frequencies in children with or without AP are summarised in **Supplemental digital content 2**.

The entire dataset was randomly partitioned into 70% for model training and 30% for out-of-sample testing. Using the training data, we assessed the incremental value of variable categories by constructing several models: Model 1 included all variables, Model 2 used only children's variables, and Model 3 used only mothers' variables. Additionally, based on insights from the clustering analysis, we developed Model 4, focusing on counts of allergic diseases and maternal comorbidities. Further details on the methodology for constructing predictive models are provided in **Supplemental**

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

digital content 3. We assessed the performance of the predictive models by creating receiver operating characteristic (ROC) curves and computing the area under the curve (AUC).

To evaluate how each variable contributed to the models, we used Shapley Additive exPlanations (SHAP) values. SHAP values provide a unified measure of the importance of variables by assigning each variable an importance value for a particular prediction. This method helps in understanding the contribution of each variable to the model's output by considering the impact of each variable on the prediction, averaged over all possible combinations of variables. By using SHAP values, we can gain insights into the relationship between the variables and the predicted outcome, thereby enhancing the interpretability of the model's predictions [32,33].

2.6. Statistical analysis

Continuous data were expressed as mean ± standard deviation (SD). Categorical data were expressed as numbers plus percentages. The Student’s t-test and Chi-squared test were used for numerical data and categorical data, respectively. Univariate and multivariate logistic regression analyses were utilised to calculate odds ratios (OR) and

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

95% confidence intervals (CI) for each variable, determining variables associated with paediatric AP. Variables significant in the univariate analysis were included in the multivariate analysis. A p-value < 0.05 was considered statistically significant. EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan) was used for statistical analysis [34].

3. Results

3.1. Study 1. Identifying factors associated with paediatric AP

A total of 13,790 children were included in the analysis, of whom 1,274 (9.2%) experienced AP (male/female = 615/659) (**Table 1**). The average age at diagnosis of AP was 5.6 ± 2.7 years. Compared to children without AP, those with AP were more likely to be female (51.7% vs. 48.1%, P = 0.01). Additionally, more children of Pakistani origin were in the AP group (58.6% vs. 45.8%, P < 0.01), while fewer were of white British origin (22.3% vs. 36.4%, P < 0.01).

Children with AP had higher rates of comorbidities, including allergic diseases (47.2% vs. 37.3%, P < 0.01), constipation (4.5% vs. 1.1%, P < 0.01), and gastro-oesophageal

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

reflux disease (GORD) (5.2% vs. 2.7%, $P < 0.01$). Moreover, mothers of children with AP reported higher incidences of AP (49.5% vs. 31.8%, $P < 0.01$), allergic diseases (47.8% vs. 39.4%, $P < 0.01$), arthritis (8.6% vs. 4.0%, $P < 0.01$), depressive and/or bipolar disorder (24.3% vs. 21.0%, $P = 0.01$), GORD (13.4% vs. 7.0%, $P < 0.01$), irritable bowel syndrome (IBS) (9.9% vs. 7.0%, $P < 0.01$), and migraine (22.4% vs. 14.7%, $P < 0.01$).

Univariate and multivariate logistic regression analysis identified several significant variables associated with paediatric AP, summarized in **Table 2** (non–significant results are summarized in **Supplemental digital content 4**). In multivariate analysis, significant associations with paediatric AP included female gender (OR, 1.17; 95% CI, 1.04–1.33; $P = 0.01$), Pakistani origin (OR, 1.57; 95% CI, 1.38–1.78; $P < 0.01$), allergic diseases (OR, 1.20; 95% CI, 1.06–1.36; $P < 0.01$), migraine (OR, 2.92; 95% CI, 1.61–5.30; $P < 0.01$), and Ehlers-Danlos syndrome (EDS) (OR, 4.09; 95% CI, 1.17–14.30; $P = 0.03$), as well as various GI diseases. Regarding maternal comorbidities, allergic diseases (OR, 1.16; 95% CI, 1.03–1.32; $P = 0.02$), GORD (OR, 1.47; 95% CI, 1.21–1.77; $P < 0.01$) and migraine (OR, 1.23; 95% CI, 1.05–1.43; $P = 0.01$), were significantly associated with paediatric AP, suggesting the adverse effect of maternal comorbidities on paediatric AP.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3.2. Study 2. Phenotyping children using unsupervised ML clustering

In Study 1, multiple factors were significantly associated with paediatric AP. However, due to the heterogeneity of the AP group, it was unclear which factors were relevant to specific cases. To address this, we employed unsupervised ML clustering to phenotype the children with AP, aiming to identify specific subgroups of AP by analysing stratified patient groups. The unsupervised model classified children with AP into three distinct phenotypes: 137 children in Phenotype 1 (10.8%), 677 children in Phenotype 2 (53.1%), and 340 children in Phenotype 3 (26.7%) (**Fig. 2**). The remaining 120 children (9.4%) exhibited miscellaneous characteristics and were sparsely distributed on the plot rendering them unclassifiable by the model; these children were excluded from downstream analyses.

Subsequently, to investigate the characteristics of the three phenotypes identified by ML clustering, we compared the clinical characteristics among the three phenotypes (**Table 3**). Age at diagnosis of AP, gender, route of birth, and ethnicity showed almost no significant differences among the three phenotypes. The frequency of GI disorders was low in all phenotypes and showed few significant differences between the

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

phenotypes. A summary of the other clinical characteristics of each phenotype is described below.

Phenotype 1 (AP with allergic predisposition)

Most children in Phenotype 1 had allergic diseases (99.3%), and all mothers had allergic diseases (100%), which were significantly higher than in Phenotypes 2 and 3 ($P < 0.01$). On the other hand, mother's AP was not observed in this phenotype. In short, this phenotype was characterized as 'AP with allergic predisposition', suggesting relevance of allergic mechanisms in AP development in children.

Phenotype 2 (AP with mother's comorbidities)

In this phenotype, the frequency of allergic diseases is relatively high at 68.7%, but other comorbidities in the children were uncommon. In contrast, maternal comorbidities showed the highest frequencies in this phenotype: AP (70.0%), allergic diseases (40.0%), depressive and/or bipolar disorder (37.2%), GORD (18.9%), and migraine (34.0%). We termed Phenotype 2 as 'AP with mothers' comorbidities,' suggesting that the mothers' illnesses may play a role in the development of paediatric AP rather than comorbidities in the children themselves.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Phenotype 3 (AP with the least comorbidities)

The frequency of mother's AP (46.2%) was the second highest in Phenotype 3. However, the frequencies of other comorbidities, including children's allergic diseases and other maternal comorbidities were uncommon. This phenotype was termed 'AP with the least comorbidities', suggesting mechanisms of AP onset that could not be explained by paediatric or maternal comorbidities investigated in this study.

Impact of allergic diseases and maternal comorbidities on paediatric AP

Since ML phenotyping suggested that there were groups of children whose AP was associated with paediatric allergies and maternal comorbidities, we investigated the effect of allergic diseases and maternal comorbidity burdens on the frequency of paediatric AP. As the number of paediatric allergic diseases increased, the frequency of AP in children increased commensurately (**Fig. 3A**). Specifically, 17.6% of children with ≥ 3 allergic diseases experienced AP, which was significantly more frequent than those with 0–2 allergic diseases ($P < 0.01$). Similarly, it was found that as the number of maternal comorbidities increased, the frequency of AP in children also increased (**Fig. 3B**). In cases where mothers had ≥ 3 comorbidities, the frequency of paediatric AP was 25.6%, which was significantly higher than in cases with 0–2 comorbidities (*P*

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

< 0.01). Furthermore, when comparing the high-risk children (those who met the criteria of having both ≥ 3 allergic diseases and ≥ 3 maternal comorbidities) to the others (control), the high-risk group had a significantly higher frequency of AP (30.8 vs. 9.1%, $P < 0.01$) (**Fig. 3C**). When visualising and comparing the frequency of children with ≥ 3 allergic diseases and/or ≥ 3 maternal comorbidities across each phenotype using a Venn diagram, it was found that the frequency of having ≥ 3 maternal comorbidities was significantly higher in phenotype 2 than in the other phenotypes ($P < 0.01$) (**Fig. 3D-3F**).

3.3. Study 3. Development of ML predictive models for paediatric AP

Finally, we constructed predictive models for the development of paediatric AP. Model 1, which used all the paediatric and maternal clinical and demographic data (detailed in **Supplementary digital content 2**) achieved an AUC value of 0.63 (95% CI, 0.59–0.67) on the test dataset (**Fig. 4A**). Model 2, which included only the paediatric clinical data, showed almost the same performance as Model 1, with an AUC of 0.64 (95% CI, 0.60–0.68) on the test dataset. Both models demonstrated modest ability to distinguish between children with and without AP. According to SHAP feature importance,

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

variables such as ethnicity and gender were more influential in Models 1 and 2 compared to paediatric or maternal comorbidities (**Figs. 4B–4E**). Specifically, the model predicted that White British children were less likely to develop AP, while Pakistani children and females were more likely to exhibit AP. Meanwhile, the performance of Model 3, which utilised only maternal clinical data, was weaker, with an AUC of 0.54 (95% CI, 0.50–0.58) on the test dataset. In Models 1 and 3, maternal AP was the most significant contributing factor among the maternal comorbidities, indicating that children with maternal AP were more likely to develop AP (**Figs. 4F–4G**).

Based on the results of the analysis in Study 2, we further developed Model 4, which focused on the number of allergic diseases and maternal comorbidities (**Figs 4H–4I**). Despite being a simpler model using only variables such as ethnicity, gender, the number of allergic conditions, and the number of maternal comorbidities, Model 4 achieved an AUC of 0.66 (95% CI, 0.62–0.70), superiorly performant to the other models.

4. Discussion

4.1. Summary of findings

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The frequency of paediatric AP in our study cohort of 13,790 children from the BiB birth cohort was 9.2 %. The univariate and multivariate logistic regression analysis demonstrated that some paediatric and maternal factors, such as paediatric allergic diseases and maternal AP, in addition to paediatric GI disorders, were associated with paediatric AP. ML-based clustering successfully identified three paediatric AP phenotypes, implying subgroups within children with AP. The phenotype characteristics suggest that allergic diseases and maternal comorbidities are relevant in developing paediatric AP across two of those phenotypes. The performance of the predictive models for paediatric AP was moderate when using the information from GP records. The greatest determinants of paediatric AP were ethnicities, maternal comorbidities (especially maternal AP), paediatric gender, and paediatric allergic predisposition.

Associated factors for paediatric AP

Our study showed that maternal comorbidities, such as GORD and migraine, were significant associated factors for paediatric AP. This aligns with previous research, showing that parental factors are related to functional AP in children [11,14,15]. One of those studies has shown that children of mothers with IBS had more GI symptoms,

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

which was explained by the effect of parents' solicitous behavioural response to children's symptoms [15]. Moreover, IBS and migraine are more common among mothers of children with functional AP [35]. Thus, maternal comorbidities and paediatric AP appear associated.

Our study also revealed that paediatric and maternal allergic predispositions were important factors for paediatric AP. Some studies have reported that allergic diseases form significant risk factors for functional GI disorders such as IBS in children [12,26,36]. Furthermore, a recent study revealed that injection of food antigens into the rectosigmoid colon of IBS patients induces local oedema and mast cell activation, proposing IBS may be a food-induced disorder mediated by mast cell activation localized to the intestine [37]. This study implies that GI neuro-immune reactions are one of the possible causes of AP in children with allergic predispositions.

4.2. Cluster analysis identifies 3 distinct phenotypes

The ML clustering revealed 3 distinct phenotypes of paediatric AP. In all phenotypes, the frequency of paediatric GI diseases was very low, suggesting that paediatric AP could not be simply explained by GI diseases. Since Phenotype 1 was characterized by allergic diseases, the aforementioned allergic mechanism could account for AP in

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Phenotype 1 [37]. Furthermore, in Phenotype 2, in which maternal comorbidities were associated, maternal illness behaviour or their solicitous response to their children are potential mechanisms of paediatric AP [15,38]. Our results reveal areas for future research and suggest tailored therapeutic approaches based on the identified phenotypes and risk factors that may enhance the management of paediatric AP.

In Phenotype 3, maternal abdominal pain was relatively frequent (46.2%). Therefore, some of the AP in children included in Phenotype 3 may be explained by maternal comorbidities. However, since the frequency of maternal comorbidities is much lower compared to Phenotype 2, it is unlikely that all cases can be attributed to this factor. It is possible that children in Phenotype 3 are characterized more by other variables, and the information available from GP records may have been insufficient to deeply phenotype these children. Furthermore, 9.4% of children were classified as unclassified in this study. With more detailed data, it might be possible to identify these unclassified children as a distinct phenotype with specific characteristics.

4.3. Machine prediction of paediatric AP

The predictive models' performance was moderate. The combination of the child's own factors and the mother's comorbidities tended to yield good performance. This result

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

implies that the development of paediatric AP is potentially a result of an interacting combination of the children's innate factors and those of their surrounding family or genetics.

In the models that utilised paediatric clinical characteristics (Models 1, 2, and 4), ethnicity emerged as an important contributor, consistent with its significance in the logistic regression model. The BiB cohort is unique in that children in this cohort are mainly comprised of White British (29.3%) and Pakistani ethnicities (39.5%) [24]. Socioeconomic status was significantly different among ethnicities in the BiB cohort [39], and lower socioeconomic status is a risk factor for pain conditions in children, including AP [40]. Therefore, the variations in the cohort's socioeconomic status might be one reason why ethnicity was an important factor both in the logistic regression analysis and in the predictive modelling.

The model focusing on the number of allergic diseases and the number of maternal comorbidities demonstrated performance comparable to the other models. As discussed in the previous section, this suggests that these factors are particularly important in the development of paediatric AP. Furthermore, it is interesting to note that while maternal AP was not significant in the multivariable logistic regression, it was the most

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

important factor among maternal comorbidities in the predictive model construction, indicative of a complex nonlinear interaction simpler models could not disclose.

4.4. Limitations

Our study is not without limitation. Firstly, the diagnosis of diseases was based on GP records; therefore, the criteria used for each diagnosis and the exact causes of AP in each case are unknown and beyond our control. However, most children with AP are diagnosed with functional AP pain at the primary care level [5], and the frequencies of organic diseases related to AP were rare in this cohort. Hence, functional AP would be the most likely cause of AP. Second, although we evaluated the performance of our predictive models in an out-of-sample test partition, we did not entirely validate them using a different cohort of children, which would maximise generalisability further. Finally, although the study relied on GP records, by incorporating more detailed data related to AP, such as socioeconomic status, lifestyle, functionality of the autonomic nervous system, and the gut microbiota composition, more detailed phenotyping and more accurate predictions of paediatric AP may be possible [41–43].

4.5. Conclusions

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Our study identified paediatric and maternal comorbidities as significant associated factors for paediatric AP. Using data-driven clustering techniques, we uncovered three distinct phenotypes of paediatric AP. These findings provide a foundation for future research aimed at elucidating the precise mechanisms underlying paediatric AP. The predictive models developed in this study highlight the potential for early identification and intervention. Further studies incorporating more detailed socio-economic, lifestyle, and biological data could refine phenotyping and these models, leading to more accurate predictions and paving the way for optimised, patient-personalised treatment strategies.

Funding

This study was supported by JA Niigata Kouseiren Grant (Niigata University School of Medicine). JKR was supported by the Medical Research Council.

References

1. Chitkara DK, Rawat DJ, Talley NJ. The epidemiology of childhood recurrent

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

abdominal pain in western countries: A systematic review. Am J Gastroenterol. 2005;100: 1868–1875.

2. Korterink JJ, Diederen K, Benninga MA, Tabbers MM. Epidemiology of pediatric functional abdominal pain disorders: A meta-analysis. PLoS One. 2015;10: 1–17.

3. Devanarayana NM, de Silva DGH, de Silva HJ. Recurrent abdominal pain syndrome in a cohort of Sri Lankan children and adolescents. J Trop Pediatr. 2008;54: 178–183.

4. Vila M, Kramer T, Obiols JE, Garralda ME. Abdominal pain in British young people: Associations, impairment and health care use. Journal of Psychosomatic Research. 2012;73: 437–442.

5. Spee LA, Van Den Hurk AP, Van Leeuwen Y, Benninga MA, Bierma-Zeinstra SM, Passchier J, et al. Childhood abdominal pain in primary care: Design and patient selection of the HONEUR abdominal pain cohort. BMC Fam Pract. 2010;11: 6–8.

6. Sperber AD, Bangdiwala SI, Drossman DA, Ghoshal UC, Simren M, Tack J, et al. Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study. Gastroenterology. 2021;160: 99–

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

114.e3.

7. Dufton LM, Dunn MJ, Compas BE. Anxiety and somatic complaints in children with recurrent abdominal pain and anxiety disorders. J Pediatr Psychol. 2009;34: 176–186.

8. Garralda ME. Unexplained physical complaints. Pediatr Clin North Am. 2011;58: 803–813.

9. Chu A, Torres L, Kao G, Gilbert C, Monico EC, Chumpitazi BP. Multidisciplinary Care for Refractory Pediatric Functional Abdominal Pain Decreases Emergency and Inpatient Utilization. J Pediatr Gastroenterol Nutr. 2022;74: 248–252

10. Warschburger P, Hänig J, Friedt M, Posovszky C, Schier M, Calvano C. Health-related quality of life in children with abdominal pain due to functional or organic gastrointestinal disorders. J Pediatr Psychol. 2014;39: 45–54.

11. Korterink J, Devanarayana NM, Rajindrajith S, Vlieger A, Benninga MA. Childhood functional abdominal pain: Mechanisms and management. Nat Rev Gastroenterol Hepatol. 2015;12: 159–171.

12. Sjölund J, Kull I, Bergström A, Järås J, Ludvigsson JF, Törnblom H, et al.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Allergy-related diseases in childhood and risk for abdominal pain-related functional gastrointestinal disorders at 16 years—a birth cohort study. BMC Med. 2021;19: 1–11.

13. McOmber M, Rafati D, Cain K, Devaraj S, Weidler EM, Heitkemper M, et al. Increased Gut Permeability in First-degree Relatives of Children with Irritable Bowel Syndrome or Functional Abdominal Pain. Clin Gastroenterol Hepatol. 2020;18: 375-384.e1.

14. Zeevenhooven J, Rutten JMTM, van Dijk M, Peeters B, Benninga MA. Parental Factors in Pediatric Functional Abdominal Pain Disorders: A Cross-sectional Cohort Study. J Pediatr Gastroenterol Nutr. 2019;68: e20–e26.

15. Levy RL, Whitehead WE, Walker LS, Von Korff M, Feld AD, Garner M, et al. Increased somatic complaints and health-care utilization in children: Effects of parent IBS status and parent response to gastrointestinal symptoms. Am J Gastroenterol. 2004;99: 2442–2451.

16. Nimnuan C, Hotopf M, Wessely S. Medically unexplained symptoms: an epidemiological study in seven specialities. J Psychosom Res. 2001;51: 361–367.

17. Ruffle JK, Farmer AD, Aziz Q. Artificial Intelligence-Assisted Gastroenterology

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

— Promises and Pitfalls. Am J Gastroenterol. 2019;114: 422–428.

18. Yang Y, Sun H, Zhang Y, Zhang T, Gong J, Wei Y, et al. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. Cell Rep. 2021;36: 109442.

19. Ko S, Pareek A, Jo C, Han H-S, Lee MC, Krych AJ, et al. Automated Risk Stratification of Hip Osteoarthritis Development in Patients With Femoroacetabular Impingement Using an Unsupervised Clustering Algorithm: A Study From the Rochester Epidemiology Project. Orthop J Sports Med. 2021;9: 23259671211050613.

20. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37: 38–47.

21. Takahashi K, Sato H, Shimamura Y, Abe H, Shiwaku H, Shiota J, et al. Achalasia phenotypes and prediction of per-oral endoscopic myotomy outcomes using machine learning. Dig Endosc. 2024;36: 789–800.

22. Ruffle JK, Patel A, Giampietro V, Howard MA, Sanger GJ, Andrews PLR, et al. Functional brain networks and neuroanatomy underpinning nausea severity can

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

predict nausea susceptibility using machine learning. J Physiol. 2019;597: 1517–1529.

23. Bravo-Merodio L, Acharjee A, Russ D, Bisht V, Williams JA, Tsaprouni LG, et al. Translational biomarkers in the era of precision medicine. Adv Clin Chem. 2021;102: 191–232.

24. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort profile: The born in bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42: 978–991.

25. Spee LAA, Lisman-Van Leeuwen Y, Benninga MA, Bierma-Zeinstra SMA, Berger MY. Prevalence, characteristics, and management of childhood functional abdominal pain in general practice. Scand J Prim Health Care. 2013;31: 197–202.

26. Olén O, Neuman A, Koopmann B, Ludvigsson JF, Ballardini N, Westman M, et al. Allergy-related diseases and recurrent abdominal pain during childhood - A birth cohort study. Aliment Pharmacol Ther. 2014;40: 1349–1358.

27. Fang ZY, Zhang HT, Lu C, Lu QM, Yu CH, Wang HY. Association between Allergic Diseases and Irritable Bowel Syndrome: A Retrospective Study. Int Arch Allergy Immunol. 2018;177: 153–159.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12: 2825–2830.

29. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML]. 2018. Available: http://arxiv.org/abs/1802.03426

30. Melvin RL, Xiao J, Godwin RC, Berenhaut KS, Salsbury FR. Visualizing correlated motion with HDBSCAN clustering. Protein Sci. 2018;27: 62–75.

31. Cappon G, Facchinetti A, Sparacino G, Georgiou P, Herrero P. Classification of postprandial glycemic status with application to insulin dosing in type 1 diabetes—an in silico proof-of-concept. Sensors . 2019;19: 1–11.

32. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. arXiv [cs.AI]. 2017. Available: http://arxiv.org/abs/1705.07874

33. Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. arXiv [cs.LG]. 2018. Available: http://arxiv.org/abs/1802.03888

34. Kanda Y. Investigation of the freely available easy-to-use software “EZR” for

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medical statistics. Bone Marrow Transplant. 2013;48: 452–458.

35. Campo JV, Bridge J, Lucas A, Savorelli S, Walker L, Di Lorenzo C, et al. Physical and emotional health of mothers of youth with functional abdominal pain. Arch Pediatr Adolesc Med. 2007;161: 131–137.

36. Jones MP, Walker MM, Ford AC, Talley NJ. The overlap of atopy and functional gastrointestinal disorders among 23 471 patients in primary care. Aliment Pharmacol Ther. 2014;40: 382–391.

37. Aguilera-Lizarraga J, Florens MV, Viola MF, Jain P, Decraecker L, Appeltans I, et al. Local immune response to food antigens drives meal-induced abdominal pain. Nature. 2021;590: 151–156.

38. Levy RL. Exploring the intergenerational transmission of illness behavior: From observations to experimental intervention. Annals of Behavioral Medicine. 2011;42: 174–182.

39. Petherick ES, Pearce N, Sunyer J, Wright J. Ethnic and socio-economic differences in the prevalence of wheeze, severe wheeze, asthma, eczema and medication usage at 4 years of age: Findings from the Born in Bradford birth cohort. Respir Med. 2016;119: 122–129.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

40. Grøholt E-K, Stigum H, Nordhagen R, Köhler L. Recurrent pain in children, socio-economic factors and accumulation in families. Eur J Epidemiol. 2003;18: 965–975.

41. Zhou Q, Verne GN. Epigenetic modulation of visceral nociception. Neurogastroenterol Motil. 2022;34: e14443.

42. Gibler RC, Jastrowski Mano KE. Systematic Review of Autonomic Nervous System Functioning in Pediatric Chronic Pain. Clin J Pain. 2021;37: 281–294.

43. Rajindrajith S, Zeevenhooven J, Devanarayana NM, Perera BJC, Benninga MA. Functional abdominal pain disorders in children. Expert Rev Gastroenterol Hepatol. 2018;12: 369–390.

Figure legends

Figure 1. Study population

Among the 13,858 children from the Born in Bradford cohort, 13,790 children with general practitioner (GP) records were included in Study 1, where we identified factors associated with paediatric abdominal pain (AP). Among these children, 1,274 children

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

with AP were included in Study 2, where we performed phenotyping of paediatric AP using unsupervised machine learning (ML) clustering. Then, for Study 3, 1,274 children with AP and 1,274 children randomly selected from 12,516 children without AP were included to create supervised ML models for paediatric AP.

Figure 2. Phenotyping of children with abdominal pain

Scattering plots of based on 3 new variables created by unsupervised uniform manifold approximation and projection (UMAP). Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) found 3 distinct phenotypes.

Figure 3. The frequency of paediatric abdominal pain depending on the number of allergic diseases and the number of maternal comorbidities

A-B. The bar graphs illustrate the relationship between the frequency of paediatric abdominal pain (AP) and the number of allergic diseases in children, as well as the number of comorbidities in their mothers. Error bars in the figure represent the 95% confidence intervals (CI) for the frequency of paediatric AP. **C** The high-risk group is defined as children with both ≥ 3 allergic diseases and ≥ 3 maternal comorbidities while the remaining children are regarded as the control group. The frequency of AP is compared between the high-risk and control groups. **D-F.** The Venn diagrams illustrate

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the overlap between children with ≥ 3 allergic diseases and mothers with ≥ 3 comorbidities in each phenotype. The blue, green, and red circles represent all children in each phenotype, children with ≥ 3 allergic diseases, and children whose mothers have ≥ 3 comorbidities, respectively. The white, light grey, and dark grey areas of the circles indicate patients with one, two, and three overlaps, respectively. The number of children is displayed in each circle.

Figure 4. Results of supervised machine learning predictive models for paediatric abdominal pain

A. Receiver operating characteristic (ROC) curves of Models 1-4. **B, D, F, and H.** The bar graphs show the mean Shapley Additive exPlanations (SHAP) value of each feature. Variables written in red represent maternal comorbidities, while those written in black represent variables related to the children themselves. **C, E, G, and I**. The beeswarm plots show the SHAP feature importance and the direction of the effect of each variable on Models 1-4. The horizontal axis represents the SHAP value. Since all the variables used in the predictive models were binary, red and blue plots indicate positive or negative for each variable, respectively. For example, in the case of 'White

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

British', blue plots tend to be distributed in the positive SHAP value and red plots in the negative SHAP value. This indicates that White British children are less likely to have abdominal pain compared to those who are not White British.

Table 1. Background characteristics of the whole study cohort (13,790 children) and their frequency of diseases

Background	With abdominal pain (n = 1274)	Without abdominal pain (n = 12516)	P value
Average age at the extraction ± SD (years)	8.4 ± 1.1	8.3 ± 1.1	<0.01
Average age at the diagnosis of abdominal pain ± SD (years)	5.6 ± 2.7	N/A	N/A
Gender, female, n (%)	659 (51.7)	6015 (48.1 %)	0.01
Route of birth, (vaginal)*, n (%)	943 (75.3)	9415 (77.0 %)	0.17
Ethnicity**, n (%)			
Pakistani	703 (58.6)	4750 (45.8)	<0.01
White British	267 (22.3)	3771 (36.4)	<0.01
Other ethnicities	229 (19.1)	1847 (17.8)	0.27
Frequency of diseases, n (%)			
Allergic diseases	601 (47.2)	4665 (37.3)	<0.01
Asthma	166 (13.0)	1217 (9.7)	<0.01

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Hay fever	131 (10.3)	720 (5.8)	<0.01
Urticaria	94 (7.4)	601 (4.8)	<0.01
Eczema	410 (32.2)	3418 (27.3)	<0.01
Appendicitis	27 (2.1)	24 (0.2)	<0.01
Arthritis	1 (0.1)	6 (0.0)	0.49
Celiac disease	9 (0.7)	24 (0.2)	<0.01
Constipation	57 (4.5)	142 (1.1)	<0.01
FD	4 (0.3)	3 (0.0)	<0.01
GORD	66 (5.2)	342 (2.7)	<0.01
IBD, colitis	2 (0.2)	4 (0.0)	0.1
IBS	8 (0.6)	9 (0.1)	<0.01
Migraine	16 (1.3)	53 (0.4)	<0.01
EDS, JHS	4 (0.3)	9 (0.1)	<0.01
Autism	4 (0.3)	36 (0.3)	0.87
Intellectual disability	2 (0.2)	16 (0.1)	0.78

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Mother’s abdominal pain	631 (49.5)	3985 (31.8)	<0.01
Mother’s allergic disease	609 (47.8)	4931 (39.4)	<0.01
Mother’s appendicitis	13 (1.0)	94 (0.8)	0.31
Mother’s arthritis	109 (8.6)	495 (4.0)	<0.01
Mother’s Celiac disease	14 (1.1)	73 (0.6)	0.04
Mother’s chronic fatigue syndrome	3 (0.2)	25 (0.2)	0.74
Mother’s chronic muscle pain	30 (2.4)	167 (1.3)	0.01
Mother’s constipation	10 (0.8)	128 (1.0)	0.55
Mother’s depressive disorder, bipolar disorder	309 (24.3)	2632 (21.0)	0.01
Mother’s FD	33 (2.6)	162 (1.3)	<0.01
Mother’s GORD	171 (13.4)	876 (7.0)	<0.01
Mother’s IBD	10 (0.8)	84 (0.7)	0.59
Mother’s IBS	126 (9.9)	877 (7.0)	<0.01
Mother’s migraine	285 (22.4)	1845 (14.7)	<0.01
Mother’s EDS, JHS	5 (0.4)	47 (0.3)	0.93

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Mother's obsessive-compulsive disorder	10 (0.8)	61 (0.5)	0.15
Mother's schizophrenia	2 (0.2)	6 (0.0)	0.17
Mother's intellectual disability	1 (0.1)	44 (0.4)	0.104

* 311 values are missing in route of birth. ** 2223 values are missing in ethnicity.

SD, standard deviation; FD, functional dyspepsia; GORD, gastro-oesophageal reflux disease; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; EDS, Ehlers-Danlos syndrome; JHS, joint hypermobility syndrome

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Significant results of univariate logistic regression analysis and subsequent multivariate analysis for the diagnosis of paediatric abdominal pain

	Univariate		Multivariate	
	OR (95% CI)	P value	OR (95% CI)	P value
Gender, female	1.16 (1.03-1.30)	0.01	1.17 (1.04-1.33)	0.01
Pakistani (vs. other ethnicities)	1.68 (1.48-1.89)	<0.01	1.57 (1.38-1.78)	<0.01
Allergic diseases	1.50 (1.34-1.69)	<0.01	1.20 (1.06-1.36)	<0.01
Appendicitis	11.30 (6.48-19.60)	<0.01	9.17 (5.07-16.60)	<0.01
Celiac disease	3.70 (1.72-7.98)	<0.01	3.22 (1.46-7.10)	<0.01
Constipation	4.08 (2.98-5.58)	<0.01	3.11 (2.23-4.34)	<0.01
FD	13.10 (2.94-58.80)	<0.01	12.30 (2.06-73.40)	0.01
GORD	1.94 (1.48-2.55)	<0.01	1.59 (1.19-2.12)	<0.01
IBS	8.78 (3.38-22.80)	<0.01	8.55 (3.05-24.00)	<0.01
Migraine	2.99 (1.70-5.25)	<0.01	2.92 (1.61-5.30)	<0.01

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

EDS, JHS	4.38 (1.35-14.20)	0.01	4.09 (1.17-14.30)	0.03
Mother's abdominal pain	2.10 (1.87-2.36)	<0.01	1.68 (1.48-1.91)	0.13
Mother's allergic disease	1.41 (1.25-1.58)	<0.01	1.16 (1.03-1.32)	0.02
Mother's arthritis	2.09 (1.39-3.16)	<0.01	1.41 (0.91-2.20)	0.13
Mother's Celiac disease	1.89 (1.07-3.37)	0.03	1.41 (0.76-2.62)	0.28
Mother's chronic muscle pain	1.78 (1.20-2.64)	<0.01	1.07 (0.69-1.65)	0.76
Mother's depressive disorder, bipolar disorder	1.20 (1.05-1.38)	0.01	1.02 (0.88-1.19)	0.76
Mother's FD (%)	2.03 (1.39-2.96)	<0.01	1.08 (0.71-1.65)	0.72
Mother's GORD	2.06 (1.73-2.45)	<0.01	1.47 (1.21-1.77)	<0.01
Mother's IBS	1.46 (1.20-1.77)	<0.01	1.06 (0.85-1.32)	0.6
Mother's migraine	1.67 (1.45-1.92)	<0.01	1.23 (1.05-1.43)	0.01

FD, functional dyspepsia; GORD, gastro-oesophageal reflux disease; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; EDS, Ehlers-Danlos syndrome; JHS, joint hypermobility syndrome

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 3. Comparison of clinical characteristics among 3 pain phenotypes

Phenotype	Phenotype 1 (n = 137)	Phenotype 2 (n = 677)	Phenotype 3 (n = 340)	P value (Phenotype 1 vs. 2)	P value (Phenotype 1 vs. 3)	P value (Phenotype 2 vs. 3)
Background						
Average age at the extraction ± SD (years)	8.5±1.1	8.7±1.1	8.6 ±1.2	0.15	0.52	0.[illegible]
Average age at the diagnosis of abdominal pain ± SD (years)	5.6±2.8	5.6±2.7	5.6±2.7	0.95	0.99	0.[illegible]
Gender, female, n (%)	67 (48.9)	334 (49.3)	172 (50.6)	0.93	0.74	0.[illegible]
Route of birth, (vaginal)*, n (%)	111 (81.0)	477 (70.5)	260 (76.5)	0.01	0.28	0.[illegible]
Ethnicity**, n (%)						
White British	31 (22.6)	146 (22.8)	62 (20.1)	0.97	0.54	0.[illegible]
Pakistani	74 (54.0)	382 (59.6)	183 (59.2)	0.23	0.30	0.9[illegible]
Others	32 (23.4)	149 (22.0)	95 (27.9)	0.73	0.31	0.0[illegible]
Frequency of comorbidities, n (%)						

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Allergic disease	136 (99.3)	465 (68.7)	0 (0.0)	<0.01	<0.01	<0.01
Appendicitis	3 (2.2)	17 (2.5)	5 (1.5)	0.83	0.58	0.26
Arthritis	0 (0.0)	1 (0.1)	0 (0.0)	0.65	0.00	0.46
Celiac disease	3 (2.2)	1 (0.1)	2 (0.6)	0.05	0.44	0.78
Constipation	6 (4.4)	38 (5.6)	10 (2.9)	0.56	0.43	0.[illegible]
FD	0 (0.0)	3 (0.4)	0 (0.0)	0.44	0.00	0.[illegible]
GORD	11 (8.0)	40 (5.9)	13 (3.8)	0.35	0.06	0.16
IBD	1 (0.7)	0 (0.0)	1 (0.3)	0.03	0.51	0.[illegible]
IBS	1 (0.7)	2 (0.3)	5 (1.5)	0.44	0.51	0.[illegible]
Migraine	1 (0.7)	12 (1.8)	3 (0.9)	0.38	0.87	0.[illegible]
EDS, JHS	1 (0.7)	3 (0.4)	0 (0.0)	0.66	0.12	0.[illegible]
Autism	0 (0.0)	0 (0.0)	2 (0.6)	0.00	0.37	0.06
Intellectual disability	0 (0.0)	0 (0.0)	0 (0.0)	NA	NA	NA
Mother's abdominal pain, n (%)	0 (0.0)	474 (70.0)	157 (46.2)	<0.01	<0.01	<0.01
Mother's allergic disease, n (%)	137 (100.0)	271 (40.0)	81 (23.8)	<0.01	<0.01	<0.01

medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

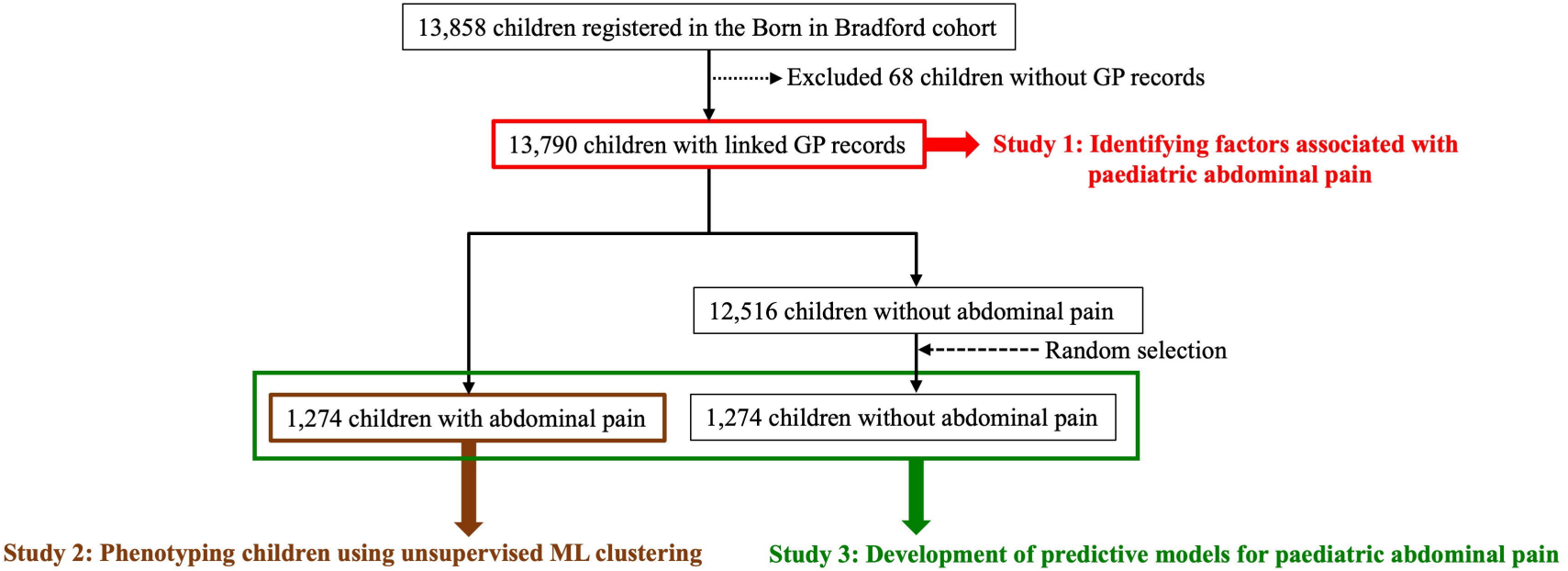
Mother’s appendicitis	2 (1.5)	7 (1.0)	4 (1.2)	0.66	0.80	0.84
Mother’s arthritis, n (%)	10 (7.3)	77 (11.4)	14 (4.1)	0.53	0.50	<0.01
Mother’s Celiac disease	0 (0.0)	9 (1.3)	4 (1.2)	0.18	0.20	0.84
Mother’s chronic fatigue syndrome	1 (0.7)	2 (0.3)	0 (0.0)	0.44	0.12	0.34
Mother’s chronic muscle pain	1 (0.7)	24 (3.5)	4 (1.2)	0.08	0.67	0.09
Mother’s constipation	0 (0.0)	8 (1.2)	2 (0.6)	0.56	0.37	0.3
Mother’s depressive disorder, bipolar disorder	28 (20.4)	252 (37.2)	0 (0.0)	<0.01	<0.01	<0.01
Mother’s FD	1 (0.7)	27 (4.0)	4 (1.2)	0.21	1.00	0.3
Mother’s GORD	14 (10.2)	128 (18.9)	14 (4.1)	0.04	0.05	<0.0
Mother’s IBD	0 (0.0)	5 (0.7)	5 (1.5)	0.31	0.15	0.36
Mother’s IBS	12 (8.8)	86 (12.7)	24 (7.1)	0.75	1.00	0.3
Mother’s migraine, n (%)	15 (10.9)	230 (34.0)	20 (5.9)	<0.01	0.24	<0.01
Mother’s EDS, JHS	0 (0.0)	3 (0.4)	1 (0.3)	0.44	0.53	0.74
Mother’s obsessive-compulsive disorder	2 (1.5)	5 (0.7)	3 (0.9)	0.40	0.58	0.84
Mother’s intellectual disability	0 (0.0)	0 (0.0)	0 (0.0)	NA	NA	NA

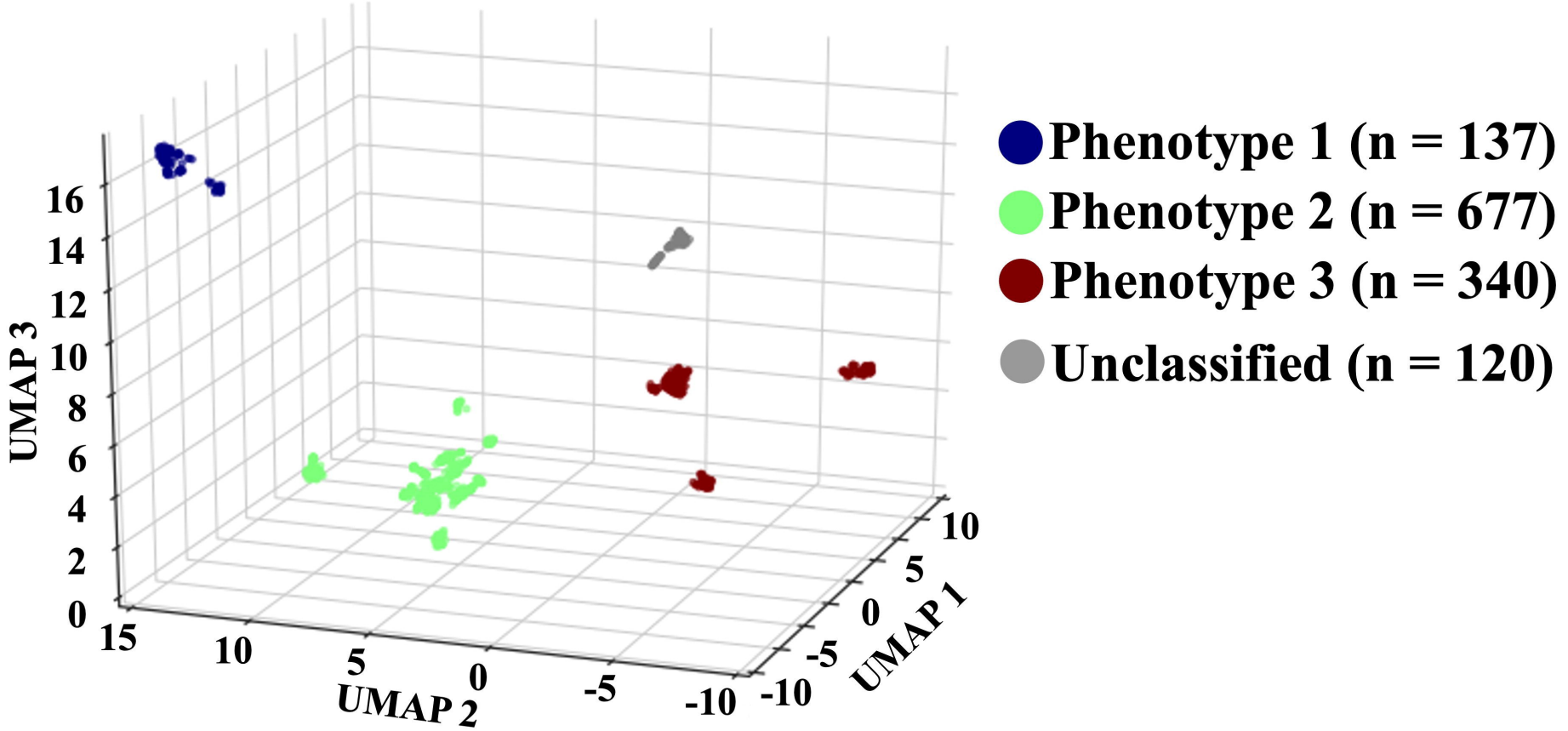
medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

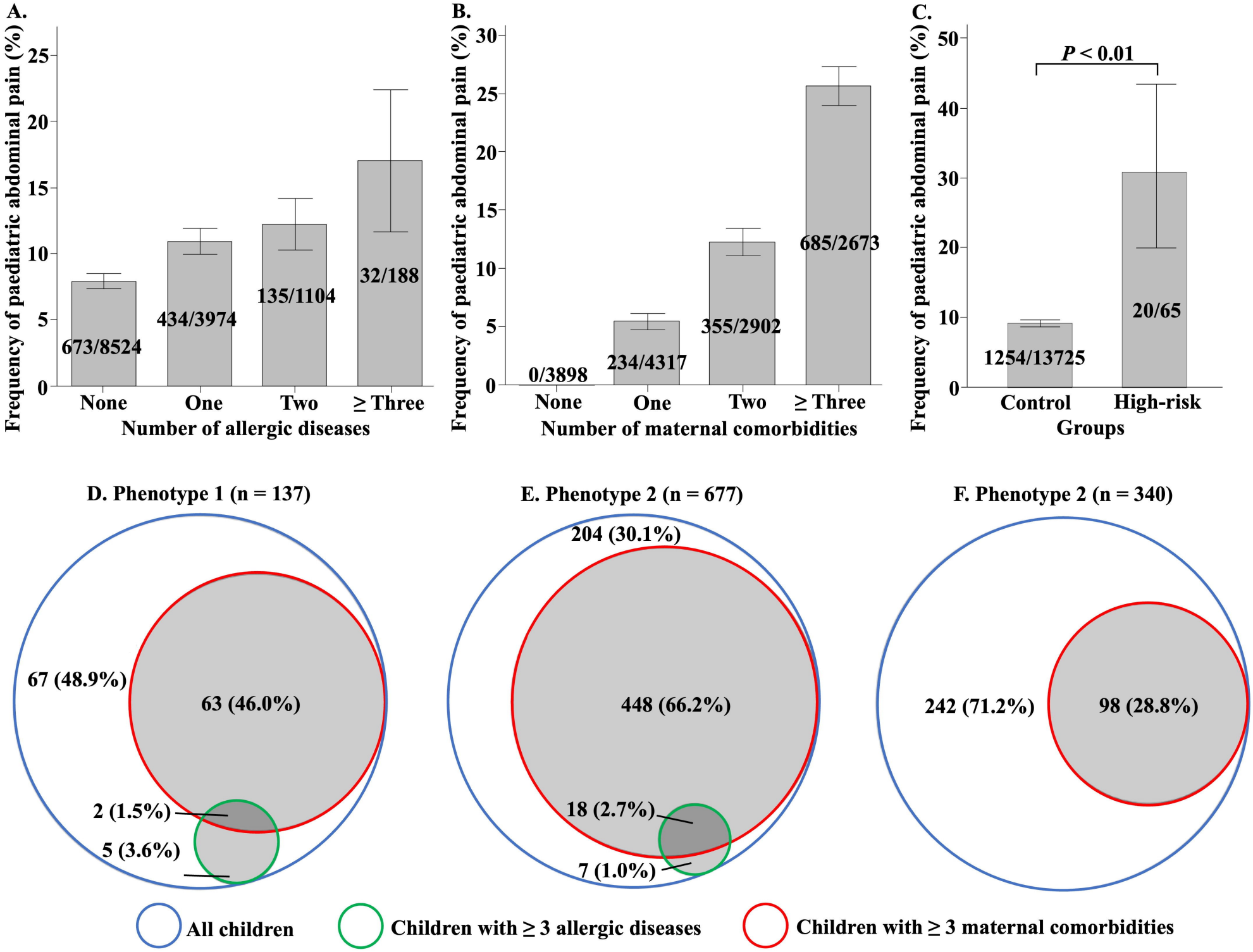
Mother's schizophrenia	1 (0.7)	0 (0.0)	1 (0.3)	0.03	0.51	0.16

FD, functional dyspepsia; GORD, gastro-oesophageal reflux disease; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; EDS, Ehlers-Danlos syndrome; JHS, joint hypermobility syndrome

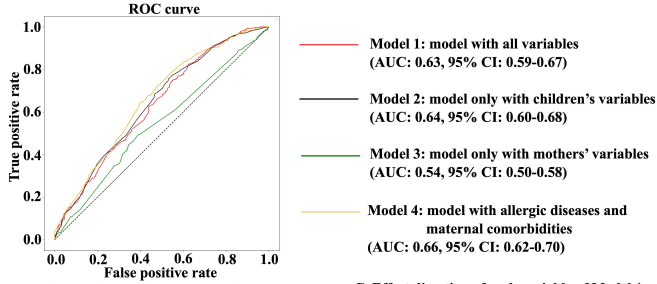
medRxiv preprint doi: https://doi.org/10.1101/2023.04.26.23289185; this version posted September 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


13,858 children registered in the Born in Bradford cohort
Excluded 68 children without GP records
13,790 children with linked GP records
Study 1: Identifying factors associated with paediatric abdominal pain
12,516 children without abdominal pain
Random selection
1,274 children with abdominal pain
1,274 children without abdominal pain
Study 2: Phenotyping children using unsupervised ML clustering
Study 3: Development of predictive models for paediatric abdominal pain

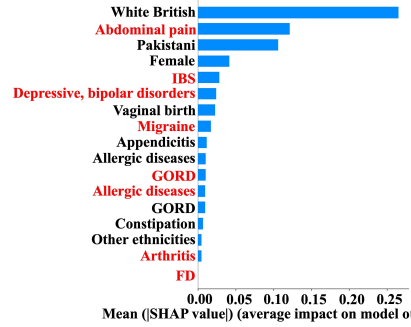

Phenotype 1 (n = 137)
Phenotype 2 (n = 677)
Phenotype 3 (n = 340)
Unclassified (n = 120)
UMAP 3
16
14
12
10
8
6
4
2
0
UMAP 2
15
10
5
0
-5
-10
UMAP 1
10
5
0
-5
-10


A.
Frequency of paediatric abdominal pain (%)
0
5
10
15
20
25
673/8524
434/3974
135/1104
32/188
None
One
Two
≥ Three
Number of allergic diseases
B.
Frequency of paediatric abdominal pain (%)
0
5
10
15
20
25
30
0/3898
234/4317
355/2902
685/2673
None
One
Two
≥ Three
Number of maternal comorbidities
C.
Frequency of paediatric abdominal pain (%)
0
10
20
30
40
50
P < 0.01
1254/13725
20/65
Control
High-risk
Groups
D. Phenotype 1 (n = 137)
67 (48.9%)
63 (46.0%)
2 (1.5%)
5 (3.6%)
E. Phenotype 2 (n = 677)
204 (30.1%)
448 (66.2%)
18 (2.7%)
7 (1.0%)
F. Phenotype 2 (n = 340)
242 (71.2%)
98 (28.8%)
All children
Children with ≥ 3 allergic diseases
Children with ≥ 3 maternal comorbidities

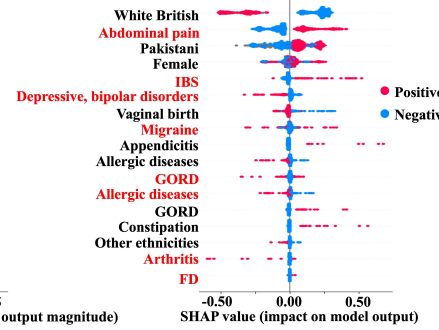
A. Receiver operating characteristics (ROC) curve and areas under the curve (AUC)



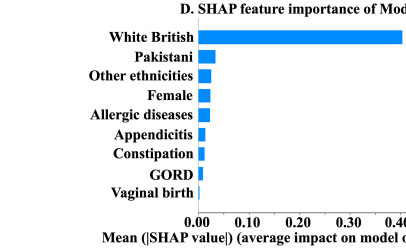
B. SHAP feature importance of Model 1



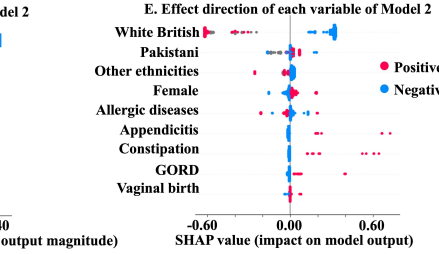
C. Effect direction of each variable of Model 1



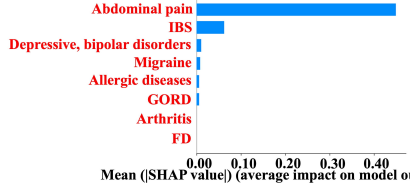
D. SHAP feature importance of Model 2



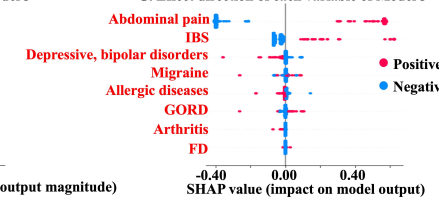
E. Effect direction of each variable of Model 2



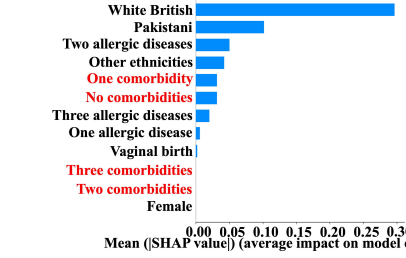
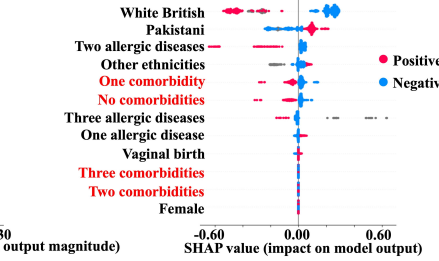
F. SHAP feature importance of Model 3



G. Effect direction of each variable of Model 3



H. SHAP feature importance of Model 4



I. Effect direction of each variable of Model 4