**Connectivity patterns between multiple allergen specific IgE antibodies and their association with severe asthma**

G. Roberts DM\*1-3, S. Fontanella PhD\*4, A. Selby DM1, R. Howard PhD5, Sarah Filippi PhD6, G. Hedlin PhD7, B. Nordlund PhD7, P. Howarth DM1,2, S. Hashimoto PhD8,9, P. Brinkman MSc8, L.J. Fleming MD4,10, C. Murray MD11, A. Bush MD14,10, U. Frey PhD12, F. Singer PhD13, Ann-Marie Malby Schoos MD PhD14, W. van Aalderen PhD9, R. Djukanovic DM1,2, K.F. Chung DSc10, P.J. Sterk PhD8, A. Custovic MD PhD FAAAI4 on behalf of the U-BIOPRED Consortium#.

\* Both authors contributed equally

# Other consortium study team members listed in online supplement

1. Clinical and Experimental Sciences and Human Development in Health Academic Unit, University of Southampton Faculty of Medicine, Southampton, UK

2. NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK

3. David Hide Asthma and Allergy Research Centre, St Mary’s Hospital, Newport, Isle of Wight, UK

4. National Heart and Lung Institute, Imperial College London, London, UK

5. Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK

6. Department of Mathematics, Imperial College London

7. Department of Women’s and Children’s Health and the Centre for Allergy Research, Karolinska Institutet at Karolinska University Hospital, 171 76 Stockholm, Sweden

8. Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

9. Department of Paediatric Respiratory Medicine and Allergy, Emma Children’s Hospital, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

10. Department of Respiratory Paediatrics, Royal Brompton Hospital, London, UK

11. Division of Infection, Immunity and Respiratory Medicine, The University of Manchester, Manchester Academic Health Science Centre, University Hospital of South Manchester NHS Foundation Trust, Manchester, UK

12. University Children’s Hospital Basel UKBB, University of Basel, Spitalstrasse, 4031 Basel, Switzerland

13. Division of Respiratory Medicine, Department of Pediatrics, University Children`s Hospital Bern, University of Bern, 3010 Bern, Switzerland

14. COPSAC; Copenhagen Prospective Studies on Asthma I Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Denmark

**Corresponding author:** Professor Graham Roberts, Paediatric Allergy and Respiratory Medicine (Mailpoint 805), Southampton University Hospital NHS Foundation Trust, Tremona Road, Southampton SO16 6YD, UK; Tel: 02381206160, Email: g.c.roberts@soton.ac.uk

**Number of figures:** 7

**Number of tables:** 1

**Word count:** 3814

**Abstract word count:** 250

**Author Contributions:** Study concept and design, GR, SFo, AS, RH, SFi, RD, KFC, PJS, AC. Acquisition of data, GR, AS, GH, PH, SH, LJF, CM, FS, HB, WvA, RD, KFC, PJS. Analysis and interpretation of data, GR, SFo, AS, RH, SFi, GH, PH, SH, PB, LJF, CM, AB, UF, FS, HB, WvA, RD, KFC, PJS, AC. Drafting of the manuscript, GR, SFo, AS, RH, AC. Critical revision of the manuscript for important intellectual content, GR, SFo, AS, RH, SFi, GH, PH, SH, PB, LJF, CM, AB, UF, FS, AMS, WvA, RD, KFC, PJS, AC. Statistical analysis, SFo, GR,AS, RH, SFi. Obtained funding, RD, KFC, PJS.

**Supported by:** U-BIOPRED is supported through an Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115010, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution (www.imi.europa.eu). We would also like to acknowledge help from the IMI funded eTRIKS project (EU Grant Code No.115446).

**Disclosure of potential conflicts of interest:** GR’s University has received honoraria from Thermo Fisher for a sponsored symposium; AMS has received a travel grant from Thermo Fisher in the last 3 years; and Custovic reports personal fees from Thermo Fisher, during the conduct of the study; personal fees from Novartis, personal fees from Regeneron, personal fees from Philips, personal fees from Sanofi, outside the submitted work. The other authors declare that they have no relevant conflicts of interest.

**Abstract**

**Background:** Allergic sensitization is associated with severe asthma, but assessment of sensitization is not recommended by most guidelines.

**Objective:** We hypothesized that patterns of IgE responses to multiple allergenic proteins differ between sensitized participants with mild/moderate and severe asthma.

**Methods:** IgE to 112 allergenic molecules (components, c-sIgE) was measured using multiplex array among 509 adults, 140 school-age and 131 pre-school children with asthma/wheeze from U-BIOPRED cohort, of whom 595 had severe disease. We applied clustering methods to identify and co-occurrence patterns of components (component clusters) and patterns of sensitization among participants (sensitization clusters). Network analysis techniques explored the connectivity structure of c-sIgE, and differential network analysis looked for differences in **c-sIgE interactions between severe and mild/moderate asthma**.

**Results:** Four sensitization clusters were identified, but with no difference between disease severity groups. Similarly, component clusters were not associated with asthma severity. None of the c-sIgE were identified as associates of severe asthma. The key difference between school-children and adults with mild/moderate compared to those with severe asthma was in the network of connections between c-sIgE. Participants with severe asthma had higher connectivity among components, but these connections were weaker. The mild/moderate network had fewer connections, but the connections were stronger. Connectivity between components with no structural homology tended to co-occur among participants with severe asthma. Results were independent from the different sample sizes of mild/moderate and severe groups.

**Conclusions:** The patterns of interactions between IgE to multiple allergenic proteins are predictors of asthma severity amongst school-children and adults with allergic asthma.

**Key words (MeSH terms):** asthma, allergic sensitization, cluster, network analysis

**Clinical trial registered:** ClinicalTrials.gov (NCT01976767, NCT 01982162)

**Key messages**

* Severe asthma is characterized by higher connectivity among allergen components but these connections are weak whereas mild/moderate asthma is characterized by fewer but stronger connections.
* It may be possible to develop algorithms to apply to allergen component sensitization data to identify clinically important allergic sensitization patterns that predict asthma severity.

**Capsule summary**

There is more co-sensitisation (higher connectivity) among allergen components in severe asthma although these connections were weaker than those in the mild asthma. The mild network had less co-sensitsations but these were stronger.

**List of abbreviations:**

U-BIOPRED Unbiased BIOmarkers for the PREDiction of respiratory diseases outcomes.

IgE Immunoglobulin E

c-sIgE Component-specific IgE

CRD Component-resolved diagnostics

ISAC ImmunoCAP Immuno-Solid phase Allergy Chip

FEV1 Forced expiratory volume in one second

FeNO Fraction of exhaled nitric oxide

HDM House dust mite

SPT Skin prick test

MCMC Metropolis-coupled Markov chain Monte Carlo

ISU ISAC Standardized Units

BMM Bernoulli Mixture Model-

**JDINAC Joint density-based nonparametric differential interaction network analysis and classification**

pLR Penalized logistic regression

**Introduction**

Severe asthma affects around one in 20 children, adolescents and adults with asthma and in these patients the healthcare resources utilization and mortality are disproportionately higher1. Both the diagnosis and treatment of severe asthma are time-consuming and require specialist expertise2. Despite modern pharmacotherapy, including biological agents, there is a significant unmet need in this group of patients3,4.

Many severe asthmatics have co-existing allergic conditions such as eczema, allergic rhinitis and food allergy55,678. Severe asthma is associated with allergic sensitization5-8, and the US National Institutes of Health expert group recommends a multi-allergen screening as the core biomarker for asthma9. However, the assessment of sensitization is not recommended by most clinical guidelines, and its use varies considerably between different countries, specialties and departments. This is in part because confirmation of sensitization using skin prick tests (SPT) and specific serum IgE to whole allergen extracts does not prove that patient’s symptoms are caused by allergy10 and is some patients may be a chance finding11. Furthermore, allergic sensitization may be an agglomerate of several subtypes12,13, some clinically relevant and others not associated with symptomatic disease14, which differ in their association with asthma presence and severity15,16. However, identifying clinically relevant sensitization in asthma is challenging14, and this limits the value of currently available tests17.

Some limitations of the tests for evaluating sensitization could potentially be overcome by component-resolved diagnostics (CRD)18, a method that detects IgE specific to individual allergenic molecules (components; c-sIgE), rather than whole extracts19. CRD arrays for multiple components generate complex and rich data sets, which create conditions to application of novel techniques such as machine learning to interrogate the data20-24. These methodologies have identified clusters of c-sIgEs to multiple allergenic proteins associated with high risk of asthma in cross-sectional20,21 and longitudinal studies22,23. A recent study investigating the connectivity structure among component-specific IgE responses measured by a CRD multiplex array has demonstrated that the key associate of childhood asthma is not the IgE response to any individual molecule but, rather, complex interaction patterns between c-sIgE24. Of note, the predictive value for asthma of a differential network of pairwise interactions between 18 c-sIgE was high, with the area under the curve of 0.9424.

In this study we postulate that disambiguation of sensitization using CRD may predict whether sensitized participants with asthma have mild or severe disease, and that this may help the assessment of both childhood and adult severe asthma. Specifically, we hypothesized that connections between IgE responses to multiple allergenic proteins differ among sensitized participants with mild/moderate when compared to those with severe asthma. We also aimed to determine whether these associations differ between children and adults. To address our hypotheses, we measured c-sIgE to 112 allergenic molecules using a multiplex CRD array in serum samples from participants in the Unbiased Biomarkers for the Prediction of Respiratory Disease Outcomes (U-BIOPRED) project, which aims to improve our understanding of severe asthma25,26. We used machine learning techniques **to explore how different** c-sIgEs **interact between each other, and to identify sensitization profiles among children** and adults with asthma. Using network analyses, we explored the connectivity structure of c-sIgE in mild/moderate and severe asthma**.**

**Methods**

**Study design, setting, and participants**

U-BIOPRED is a European multicenter, observational study. The adult cohort was recruited from 16 centers in 11 countries25 and the pediatric cohort from seven centers in five countries26. The study was approved by the ethics committee for each participating institution and is registered on ClinicalTrials.gov (NCT01976767, NCT01982162). All participants and parents/caregivers gave written informed consent.

Three adult patient groups were recruited: (1) severe non-smoking asthmatics); (2) smokers and ex-smokers with severe asthma; and (3) mild/moderate non-smoking asthmatics25. The four pediatric groups were: (1) severe school-aged asthmatics; (2) mild/moderate school-aged asthmatics; (3) severe pre-school wheezers; and (4) mild/moderate pre-school wheezers26. The groups were defined by strict clinical criteria and both the adult and pediatric participants with severe asthma had poorly controlled disease despite high-dose inhaled corticosteroids and other controller medication, with a proportion requiring maintenance oral corticosteroids25,26.

**Data sources/measurements**

Full cohort descriptions, inclusion and exclusion criteria and details of clinical assessment are shown in Tables S1 and S2. Initial characterization of allergic sensitization was by SPT and/or measurement of serum IgE specific to common allergens, with sensitization defined as a mean wheal diameter ≥3mm or IgE ≥0.35kUA/l respectively.

**Component-resolved diagnostics (CRD)**

We measured c-sIgE to 112 allergenic molecules (components) using ImmunoCAP ISAC® (ThermoFisher/Phadia AB, Uppsala, Sweden) where sufficient serum sample was available. The level of c-sIgEs was reported in ISAC Standardized Units (ISU). Depending on the nature of the analysis, we dichotomized c-sIgE using a binary threshold (positive ≥0.30 ISU) or used the continuous values.

**Statistical learning**

The overall approach to analysis is described in Figure 1 (see only supplementary materials for glossary). To account for zero-inflated variables, we analyzed data from participants with at least one c-sIgE >0.3 ISU and for components with a positive response by at least three subjects (described before as active components23,24). Clustering methods (see below) were applied **to explore how different** c-sIgEs **interact** (*component clusters* derived by clustering c-sIgEs) and to identify patterns of c-sIgE sensitization amongst participants (*sensitization clusters* derived by clustering participants). We used Chi-squared test to evaluate the associations between the identified sensitization and component clusters and asthma severity; all p-values were adjusted for false discovery rate (FDR). Network analysis techniques were then applied to explore the connectivity structure of c-sIgE, and differential network analysis to ascertain any differences in **c-sIgE interactions between severe and mild/moderate asthma**. Finally, classification models for asthma severity were investigated.

**Statistical grouping of c-sIgEs (component clusters)**

Co-occurrence patterns of c-sIgEs were analyzed by grouping the components into homogeneous clusters using Bayesian estimation of a mixture of Bernoulli distributions (Bernoulli Mixture Model-BMM)27 (Supplementary methods).

Patterns of sensitization among participants with asthma (**sensitization clusters)**

Participants with similar c-sIgE sensitization patterns were grouped by hierarchical clustering, using the Ward’s linkage algorithm on the Jaccard distance matrix computed between binary responses to c-sIgE profiles (Supplementary methods). **To determine the number of clusters,** dendrogram plots were **evaluated and suggested solutions assessed using a bootstrap approach28.**

**c-sIgE connectivity patterns**

To investigate the connectivity structure of c-sIgEs**, we used a** weighted co-occurrence network analysis (Supplementary methods). Nodes in the network represent c-sIgEs, and pairs of c-sIgEs simultaneously present in at least one participant are connected by an edge. Edges are weighted by the co-occurrence frequency of the component pairs. To compare connectivity structures between adults and children, we considered network size and density. To assess the importance and position of each c-sIgE (i.e. node) in the network we used three centrality indices, namely strength, closeness, and betweenness.

Difference in c-sIgE interactions between severe and mild/moderate asthma

We compared network structure between adults with severe and mild/moderate non-smoking asthma; in children, we focused on school-aged asthma. To reduce the noise due to c-sIgEs with no positive response, c-sIgEs with <3 positive responses were filtered out in the mild/moderate and severe asthma groups separately23, and we used those c-sIgEs present in both groups.

The networks were built by connecting c-sIgEs whose pairwise expression similarity, based on Spearman correlation coefficient, was statistically significant (p<0.05). **Centrality measures were used to compare the networks and the role of specific components**. 21To **investigate whether there are differences in c-sIgE connections between severe and mild/moderate asthma, we applied differential correlation analysis (Supplementary methods), which computes the correlations in each severity group and uses the difference in z-transformed correlation coefficients to assess statistical significance29. W**e assessed the stability of the inferred network using a bootstrap approach. We performed sensitivity analyses to evaluate comparability of the mild/moderate and severe networks and assess the impact of the different sample sizes **(Supplementary methods)**.

Classification model for severe asthma

We evaluated the ability of **c-sIgE pairwise interactions to classify asthma severity by applying the joint density-based nonparametric differential interaction network analysis and classification (JDINAC; Supplementary methods)30. This model identifies differential interaction patterns of network activation of c-sIgEs that are most closely related to asthma severity, and builds a classification model using the network biomarkers.** To ascertain whether c-sIgE pairwise interactions are more informative than individual components to classify asthma severity, we compared the performances of JDINAC with penalized logistic regression (pLR). To ensure robustness of the results, we run both models with 10-fold cross validation in 20 independent repetitions. In this analysis, sIgEs values were log-transformed (*log*(x + 1)). To reduce the effect of imbalanced data, we included class weights in both models.

**All statistical analyses were run in the programming language R. Allergen components clusters were estimated using** packages BayesBinMix27 and label.switching31. **We used *fpc32* for computing the number of clusters through bootstrap procedure, and *igraph33* for network visualizations. Differential correlation analysis** was carried out through package *DGCA29***. *JDINAC* scripts were made available by the authors at** https://github.com/jijiadong/JDINAC**.**

**Results**

**Participants**

Study population consisted of 780 participants (509 adults, 140 school-age and 131 pre-school children). Of these, 595 had severe asthma/wheeze, with the rest having mild/moderate disease. Characteristics of the study population are shown in Table 1.

Allergic sensitization (both SPT and IgE) was highly prevalent across the four cohorts. The proportion of sensitized participants was similar in the severe and mild/moderate groups (Table 1) with the exception of the adult smoking and ex-smoking severe asthma groups in which fewer participants were sensitized.

**Allergic sensitization to allergen components**

CRD data were available for 491/509 (96.5%) adults, and for 238/271 (87.8%) children (Figure 2). The percentage of participants with at least one positive c-sIgE in the adult, school-age and pre-school cohorts was 50.7% (249/491), 82.7% (105/127) and 27.1% (30/111), respectively (Table S1). The number of active components was similar among adults and school-age children (n=58 and n=64 respectively), and was significantly lower in the pre-school group (n=13; Table S3). The proportion of participants in each cohort with no (<0.3 ISU), low (0.3-1 ISU), medium (1-15 ISU) or high (>15 ISU) sensitization for each component is shown in Figure S1.

There were no obvious differences in c-sIgE to individual components across different asthma severity groups in adults and school-age children, with a high prevalence of positive responses to house dust mite (HDM) and grass components (Figure 3). In pre-school children, positive c-sIgEs were concentrated around the HDM components. The number of positive c-sIgEs did not vary with asthma severity in either adults or children (Table S3).

**Statistical grouping of allergen components: Component clusters**

The BMM inferred **three component clusters in adults, five in school-age and two in pre-school children (Figure 4).** Cluster membership mapped mostly to the structural homology of proteins and/or their biological sources. For each age-group, the algorithm revealed a “Broad” cluster, with sensitization to components from multiple sources, and “HDM” cluster, consisting of four HDM components (Der p 1 and Der p 2, Der f1 and Der f 2). In addition, a “Grass” cluster emerged in adults, consisting of two components (Phl p 1 and Phl p 5). In the school-age children, 3 additional clusters were found: “Grass/cat”, “Animal/pollen” and “Class-10 Pathogenesis-related proteins (PR-10)” cluster. There was no significant association between cluster membership, or sensitization to any individual component, and asthma severity (Table S4).

**Patterns of sensitization among participants with asthma: Sensitization clusters**

**The optimal solution for** **sensitization profiles among study participants identified** **four clusters in adults and school-age children, and two in pre-school age (Figure 5). Based on the profile of components to which sensitization occurred, we qualitatively labelled clusters in adults as**: Multiple (c-sIgE to multiple components across all component clusters); Predominantly HDM (c-sIgE mainly to HDM components); Predominantly grass pollen (c-sIgE to multiple grass components); and Lower-grade sensitization (sensitization to a few of a broad range of components). The same clusters were found in school-age children. In pre-school children, we observed only Multiple and HDM clusters.

No significant association was found between cluster membership and asthma severity, or allergic comorbidities (Table S5). Among adults, participants in Multiple and Predominantly HDM clusters had significantly earlier age of onset of symptoms and diagnosis, and those in the Predominantly grass cluster had better lung function (significantly higher FEV1 and FEV1/FVC); there was no significant association between cluster membership and bronchodilator response, FeNO, and number of exacerbations in the previous year in both children and adults (Table S6).

**c-sIgE connectivity patterns**

Investigation of the co-occurrence patterns of c-sIgEs revealed similar networks in adults and school-age children**.** Network density suggested a higher level of association in the school- and preschool-age networks (Table S7). Comparison of centrality measures revealed similar connectivity and role of components in the adult and school-age networks (Figure S3). IgE to HDM components, Phl p 1 and Fel d 1 showed similar behavior, and were among the 10 most important components in the network (Table S8). Network centrality measures showed that peanut-related components were more important and connected with a broader range of components in school-age children, indicating that a larger number of participants in this cohort was co-sensitized to peanut-related components and other c-sIgEs. Grass pollen components were more important in the adult network (Figure S3) while HDM components and Fel d 1 shared many connections also in the pre-school age group (Figure S4).

**Differential sIgE connectivity structure in mild and severe asthma**

After filtering, 32 active c-sIgE components were used in the analysis among adults with severe and mild/moderate non-smoking asthma, and 41 in school-age children. Table S9 lists components which were removed. Figure 6 shows the connectivity structure stratified by asthma severity in adults and children. There was a markedly different connectivity structure between the severe and mild/moderate asthma networks, with similar patterns of associations in adults and children. Severe asthma networks were characterized by a high number of connections among c-sIgEs, many of which, however, were weak (low values for the Spearman correlation coefficients). In contrast, mild/moderate networks had fewer but stronger connections (Figure 6, Table S10). In mild/moderate networks, we observed the tendency of c-sIgEs to components with structural homology to connect, while in severe networks, c-sIgEs interacted with a broader range of components.

The stability analysis (Figure S6) showed that connections were very stable for high edge weights. Slightly wider error bars were registered for moderate weights, while high variability around the bootstrap estimates was observed for low weights, making these connections unstable and difficult to interpret. In the inferred network, these weak connections were filtered out by the thresholding procedure based on the significance of the correlation coefficients. The sensitivity analysis excluded the effect of the different sample sizes on the findings (Figure S7). Both network structure and density measures suggested that more connections among c-sIgEs were established within the severe group, even when considering samples of reduced size (Figures S6 and S7, panels c and d).

We proceeded to investigate differences in the connectivity structure between mild/moderate and severe asthma using differential correlation analysis between all pairs of c-sIgEs. There was a significant difference in the connectivity structure between the severe and mild/moderate networks. In adults, 13 c-sIgEs showed significant differential correlations; in school-age children, significant differences were found amongst 12 c-sIgEs (Figure 7). Connections amongst animal components showed higher correlations in severe asthma in both adults and children, and connections between peanut allergens characterized mild/moderate asthma. Connections amongst remaining components also differed, with results suggesting that connections between pairs of components with no structural homology tend to occur in the severe group. The sensitivity analysis showed that results were independent from the different sample sizes of the mild/moderate and severe groups. The pairwise differential c-sIgEs connections were confirmed by the bootstrap analysis (Table S11).

To investigate whether individual c-sIgEs or their pairwise interactions were better discriminators between mild/moderate and severe asthma, we compared the performances of pLR and JDINAC. JDINAC was superior in classifying asthma severity in both adults and children: area under the curve (AUC) 0.95 vs. 0.54; and 0.93 vs. 0.53, JDINAC and pLR respectively). Both models had high sensitivity and were able to correctly classify the severe group, but had limits in classifying the mild/moderate group (Table S12).

**Discussion**

Our results suggest that amongst adults and school-age children with allergic asthma, key differences in sensitization between participants with mild/moderate and those with severe disease is different connectivity and interaction between allergen component-specific IgEs. Severe asthma is characterized by higher connectivity among components (i.e. higher number of co-occurrences of sensitization between any two components) but these connections were weaker than those in mild/moderate asthma. In contrast, in mild/moderate asthma, there are fewer connections between components, but the individual connections are stronger. Connectivity between components with no structural homology tends to co-occur among participants with severe asthma, whilst in those with mild/moderate asthma there is a tendency of IgE to structurally homologous components to connect. The classification results suggest that global connectivity is much more important than individual component-specific IgE in predicting asthma severity.

Our study provided an opportunity to compare sensitization amongst asthmatics across the life-course. Although the prevalence of CRD sensitization was lower in adults with asthma when compared to school-age children, amongst sensitized participants there were striking similarities in the structure of data between adults and school-children, with a similar number of active components (58 and 64 respectively), and four identical sensitization clusters. The number of active components in pre-school children was lower (only 13), with considerably fewer participants being sensitized. Severe wheeze in pre-school cohort was associated with the HDM cluster, but given the small number of pre-school children who were sensitized, it is difficult to make definitive conclusions about this age group. This and other studies suggest that specific sensitization patterns in pre-school children may be associated with severe wheezing and lung function decline34,35, but whether this can be used as a marker of future risk can only be answered through a long-term follow-up of larger cohorts. Our results suggest that among sensitized individuals, there are considerable similarities between adult and school-age severe asthma and that severe asthma in school-age children is more similar to severe asthma in adults than to troublesome pre-school wheezing.

In this study, in both adults and school-age children, there was no relationship between asthma severity and IgE responses to whole allergen extracts in standard allergy tests (SPT or serum IgE), or any single allergenic molecule tested using CRD**.** Previous studies have reported that severe asthma was associated with multi-sensitization to animal component36,37, and that children with asthma respond to more components than those who do not have asthma24. By carrying out a thorough assessment of component-specific sensitization using multiplex array in a large number of participants, we have shown no major difference in the number of positive component-specific IgEs between mild/moderate and severe asthma, but an increase in the strength of the connections (measured by the Spearman correlation coefficient) between proteins from furry animals in severe asthma.

**W**e described four sensitization clusters (Multiple, Predominantly HDM, Predominantly grasses and Lower-grade sensitization) amongst participants with asthma. **We have previously reported almost identical clusters in a population-based birth cohort (i.e. in general population) and have shown that** Multiple sensitization **cluster was strongly** associated with asthma diagnosis24. In the current study, these sensitization clusters were seen in both severe and mild/moderate asthma, with no difference between the severity groups. **Similarly, allergen component clusters which we describe here (Broad, HDM, Grass/cat, PR-10 etc.), are** comparable **to those uncovered in the previous population-level analysis23. These component clusters are biologically plausible and likely reflect the sources and structural homogeneity of components within protein families21. However, in contrast to our previous population-based studies in which sensitization to Broad component clusters was associated with asthma diagnosis21,24 and sensitization to Grass/cat cluster predicted asthma development23, in the current study** **of participants with asthma, component clusters were not associated with disease severity. I**t is reassuring that similar sensitization clusters and allergen component clusters were discovered using data-driven techniques in two different types of studies with different biases (population-based and study of participants with asthma), thereby allowing triangulation of findings38,39. Taken together, the results **suggest that mechanisms responsible for asthma development differ from those that impact upon disease severity11, and that specific sensitization patterns which are associated with asthma diagnosis22 and persistence21 are not informative about severity.**

Severe asthma was characterized by higher connectivity among c-sIgEs, but the connections were weaker; participants with mild/moderate asthma had fewer connections, but connections were stronger. Furthermore, connections between pairs of components with no structural homology tend to occur in the severe asthma, while in mild/moderate asthma the connectivity was mainly between structurally homologous components. It is tempting to speculate on potential mechanisms which may underpin these observations. Recent data-driven analysis of multiple interferon-related, pro-inflammatory and regulatory cytokines induced by rhinovirus described six clusters of anti-virus responses, one of which (characterized by a weak interferon and strong pro-inflammatory response) was linked with early-onset severe asthma with multiple early allergic sensitization40. IgE in these participants may be produced locally in the airways without cognate T-cell help, to a wide range of structurally unrelated proteins40, resulting in a sensitization pattern characterized by a high number of connections which are however weak. In participants with mild/moderate asthma, the main pathway to sensitization is via the adaptive Th2 immunity and allergen exposure, and sensitization is characterized by strong connectivity to structurally homologous proteins from similar sources. This links features of poor respiratory outcomes (impaired anti-virus immunity40, severe exacerbations35, low lung function trajectory34 and multiple early sensitization14,15,19,32,37) with the severe asthma network of IgE connectivity described here.

One limitation of our study is that some potentially important allergens were not included on the chip which we used, and it is possible that our analyses would provide different solutions if additional components were available. For example, there are relatively few fungal components, which may be important in severe asthma6,7. Our filtering process may have excluded some informative allergens but, since zero-inflated variables can reduce accuracy of cluster analyses and reliability of prediction models, this was a necessary step to account for the effect of measurement errors and noise.

Another limitation is the sparseness of data, and relatively small sample size in pre-school children. It is therefore difficult to infer results to the general context and our findings require external validation. However, the number of participants with severe asthma is small and U-BIOPRED is one of largest studies with a focus on severe disease. To the best of our knowledge, there are no other studies with meticulously defined mild/moderate and severe asthma in which CRD data is available. Of note, we demonstrated limited (internal) replication between the school-age and adult cohorts recruited from several European countries and assessed using identical protocols, strengthening our conclusions. It would be helpful to also demonstrate external replication.

We identified only pairwise interactions between component-specific IgEs, and the relation between asthma severity and IgE connectivity structure may be more complex. We suggest that higher-order interactions should be investigated in the future, but such analyses are currently too computationally demanding.

In conclusion, the pattern of connectivity and interactions between IgE to multiple allergenic proteins is a potentially important biomarker of asthma severity amongst school-age and adult patients with allergic asthma. This study provides new directions for the assessment and interpretation of sensitization in asthma and may help design more reliable diagnostic tools to differentiate clinically important sensitization patterns associated with asthma severity.

**Acknowledgments**

This paper is presented on behalf of the U-BIOPRED Study Group with input from the U-BIOPRED Patient Input Platform, Ethics Board and Safety Management Board. We thank Tom Cull who undertook the ISAC assays at University Hospital Southampton NHS Foundation Trust. The authors would like to acknowledge help in data and knowledge management from the eTRIKS project, which is funded by the Innovative Medicines Initiative. Additionally, the authors would like to acknowledge the support of the University Hospital Southampton NHS Foundation Trust and the NIHR-Wellcome Trust Clinical Research Facility, Southampton, UK.

**References**

1. Bahadori K, Doyle-Waters MM, Marra C, Lynd L, Alasaly K, Swiston J, et al. Economic burden of asthma: a systematic review. *BMC Pulm Med* 2009; **9**: 24.

2. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* 2014; **43**(2): 343-73.

3. Chipps BE, Szefler SJ, Simons FE, Haselkorn T, Mink DR, Deniz Y, et al. Demographic and clinical characteristics of children and adolescents with severe or difficult-to-treat asthma. *J Allergy Clin Immunol* 2007; **119**(5): 1156-63.

4. Holguin F, Cardet JC, Chung KF, Diver S, Ferreira DS, Fitzpatrick A, et al. Management of Severe Asthma: a European Respiratory Society/American Thoracic Society Guideline. *Eur Respir J* 2019.

5. Del Giacco SR, Bakirtas A, Bel E, Custovic A, Diamant Z, Hamelmann E, et al. Allergy in severe asthma. *Allergy* 2017; **72**(2): 207-20.

6. Castanhinha S, Sherburn R, Walker S, Gupta A, Bossley CJ, Buckley J, et al. Pediatric severe asthma with fungal sensitization is mediated by steroid-resistant IL-33. *J Allergy Clin Immunol* 2015; **136**(2): 312-22 e7.

7. Denning DW, O'Driscoll BR, Hogaboam CM, Bowyer P, Niven RM. The link between fungi and severe asthma: a summary of the evidence. *Eur Respir J* 2006; **27**(3): 615-26.

8. Bossley CJ, Fleming L, Gupta A, Bossley CJ, Fleming L, Gupta A, et al. Pediatric severe asthma is characterized by eosinophilia and remodeling without T(H)2 cytokines. *J Allergy Clin Immunol* 2012; **129**(4): 974-82 e13.

9. Szefler SJ, Wenzel S, Brown R, Erzurum SC, Fahy JV, Hamilton RG, et al. Asthma outcomes: Biomarkers. *J Allergy Clin Immun* 2012; **129**(3): S9-S23.

10. Roberts G, Ollert M, Aalberse R, Austin M, Custovic A, DunnGalvin A, et al. A new framework for the interpretation of IgE sensitization tests. *Allergy* 2016; **71**(11): 1540-51.

11. Saglani S, Custovic A. Childhood Asthma: Advances Using Machine Learning and Mechanistic Studies. *Am J Respir Crit Care Med* 2019; **199**(4): 414-22.

12. Oksel C, Custovic A. Development of allergic sensitization and its relevance to paediatric asthma. *Curr Opin Allergy Clin Immunol* 2018; **18**(2): 109-16.

13. Custovic A, Lazic N, Simpson A. Pediatric asthma and development of atopy. *Curr Opin Allergy Clin Immunol* 2013; **13**(2): 173-80.

14. Holt PG, Strickland D, Bosco A, Belgrave D, Hales B, Simpson A, et al. Distinguishing benign from pathologic TH2 immunity in atopic children. *J Allergy Clin Immunol* 2016; **137**(2): 379-87.

15. Lazic N, Roberts G, Custovic A, Belgrave D, Bishop CM, Winn J, et al. Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. *Allergy* 2013; **68**(6): 764-70.

16. Simpson A, Tan VY, Winn J, Svensen M, Bishop CM, Heckerman DE, et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med* 2010; **181**(11): 1200-6.

17. Sonntag HJ, Filippi S, Pipis S, Custovic A. Blood Biomarkers of Sensitization and Asthma. *Front Pediatr* 2019; **7**: 251.

18. Eiringhaus K, Renz H, Matricardi P, Skevaki C. Component-Resolved Diagnosis in Allergic Rhinitis and Asthma. *J Appl Lab Med* 2019; **3**(5): 883-98.

19. Wickman M, Lupinek C, Andersson N, Belgrave D, Asarnoj A, Benet M, et al. Detection of IgE Reactivity to a Handful of Allergen Molecules in Early Childhood Predicts Respiratory Allergy in Adolescence. *EBioMedicine* 2017; **26**: 91-9.

20. Prosperi MC, Belgrave D, Buchan I, Simpson A, Custovic A. Challenges in interpreting allergen microarrays in relation to clinical symptoms: a machine learning approach. *Pediatr Allergy Immunol* 2014; **25**(1): 71-9.

21. Simpson A, Lazic N, Belgrave DCM, Johnson P, Bishop C, Mills C, et al. Patterns of IgE responses to multiple allergen components and clinical symptoms at age 11 years. *J Allergy Clin Immun* 2015; **136**(5): 1224-31.

22. Custovic A, Sonntag H-J, Buchan IE, Belgrave D, Simpson A, Prosperi MCF. Evolution pathways of IgE responses to grass and mite allergens throughout childhood. *J Allergy Clin Immun* 2015; **136**(6): 1645-52.e8.

23. Howard R, Belgrave D, Papastamoulis P, Simpson A, Rattray M, Custovic A. Evolution of IgE responses to multiple allergen components throughout childhood. *J Allergy Clin Immunol* 2018; **142**(4): 1322-30.

24. Fontanella S, Frainay C, Murray CS, Simpson A, Custovic A. Machine learning to identify pairwise interactions between specific IgE antibodies and their association with asthma: A cross-sectional analysis within a population-based birth cohort. *PLoS Med* 2018; **15**(11): e1002691.

25. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015; **46**(5): 1308-21.

26. Fleming L, Murray C, Bansal AT, Hashimoto S, Bisgaard H, Bush A, et al. The burden of severe asthma in childhood and adolescence: results from the paediatric U-BIOPRED cohorts. *Eur Respir J* 2015; **46**(5): 1322-33.

27. Papastamoulis P, Rattray M. BayesBinMix: an R Package for Model Based Clustering of Multivariate Binary Data. *The R Journal* 2017; **9**(1): 403--20.

28. Fang Y, Wang J. Selection of the number of clusters via the bootstrap method. Comput. Stat. Data Anal. 2012; **56**(3): 468-77.

29. McKenzie AT, Katsyv I, Song W-M, Wang M, Zhang B. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC systems biology* 2016; **10**(1): 106-.

30. Ji J, He D, Feng Y, He Y, Xue F, Xie L. JDINAC: joint density-based non-parametric differential interaction network analysis and classification using high-dimensional sparse omics data. *Bioinformatics* 2017; **33**(19): 3080-7.

31. Papastamoulis P. label.switching: An R Package for Dealing with the Label Switching Problem in MCMC Outputs. *Journal of Statistical Software; Vol 1, Code Snippet 1* 2016.

32. Hennig C. Fpc: Flexible procedures for clustering. R package version 2010; 2(2): 0-3.

33. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006; 11; 1695(5): 1-9.

34. Belgrave DCM, Granell R, Turner SW, Curtin JA, Buchan IE, Le Souëf PN, et al. Lun g function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir Med* 2018; **6**(7): 526-34.

35. Belgrave DC, Buchan I, Bishop C, Lowe L, Simpson A, Custovic A. Trajectories of lung function during childhood. *Am J Respir Crit Care Med* 2014; **189**(9): 1101-9.

36. Konradsen JR, Nordlund B, Onell A, Borres MP, Gronlund H, Hedlin G. Severe childhood asthma and allergy to furry animals: refined assessment using molecular-based allergy diagnostics. *Pediatr Allergy Immunol* 2014; **25**(2): 187-92.

37. Nwaru BI, Suzuki S, Ekerljung L, Sjölander S, Mincheva R, Rönmark EP, et al. Furry Animal Allergen Component Sensitization and Clinical Outcomes in Adult Asthma and Rhinitis. *J Allergy Clin Immunol Pract* 2019; **7**(4): 1230-8 e4.

38. Custovic A, Henderson J, Simpson A. Does understanding endotypes translate to better asthma management options for all? *J Allergy Clin Immunol* 2019; **144**(1): 25-33.

39. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; **45**(6): 1866-86.

40. Custovic A, Belgrave D, Lin L, Bakhsoliani E, Telcian AG, Solari R, et al. Cytokine Responses to Rhinovirus and Development of Asthma, Allergic Sensitization, and Respiratory Infections during Childhood. *Am J Respir Crit Care Med* 2018; **197**(10): 1265-74.

 *[Figure 1 goes here (Figure\_1.tiff) ]*

**Figure 1. Approach to statistical analysis**

The data set was initially restricted to allergen component specific IgE (c-sIgE) where at least 3 participants showed evidence of sensitisation (≥0.30) and then to participants with sensitisation to at least one c-sIgE. Parallel work flows then focused on c-sIgE connectivity and allergen component sensitisation patterns.

*[Figure 2 goes here (Figure\_2.tiff) ]*

**Figure 2. Participants in each cohort and numbers included in the ISAC subgroup**

Participant were included in the ISAC subgroup if serum was available.

*[Figure 3 goes here (Figure\_3.tiff) ]*

**Figure 3. Patterns of sensitisation to each allergen components for individual participants stratified by age group**

Participants are represented in columns and allergen components (key ones labelled on right) in rows. Black square indicated that a participant is sensitised to a particular allergen component. Participants are divided into the cohorts. Green bars represent mild/moderate asthma, orange/red bars indicate severe asthma cohorts. Further details can be found in Table S3 in the supplementary materials.

[Figure 4 goes here (Figure\_4.tiff)]

|  |
| --- |
| **Allergen component connectivity clusters** |
| **Adult** | **School age** | **Pre-school age** |
| Cluster C1 (blue): Broad  | Cluster C1 (purple): Broad  | Cluster C1 (blue): Broad  |
| Cluster C2 (green): House dust mite  | Cluster C2 (green): House dust mite  | Cluster C2 (red): House dust mite  |
| Cluster C3 (red): Grass pollen  | Cluster C3 (blue): Grass/cat  |  |
|  | Cluster C4 (orange): Animal/pollen  |  |
|  | Cluster C5 (red): PR-10  |  |

**Figure 4. Bernoulli Mixture Model and Co-occurrence network analysis reveal connectivity structure between allergen components**

Colours (see table) represent allergen component cluster memberships, node diameter is proportional to the node relative strength, while edge (line) width is proportional to the number of participants sensitised to both components. Participants with no sensitisation to allergen components were not included in this step.

*[Figure 5 goes here (Figure\_5.tiff)]*

**Figure 5. Patterns of IgE responses to allergen components for individual participants within clusters**

Rows represent sIgEs, while columns indicate participants. Row colours represent sensitisation clusters’ membership (participants with no sensitisation to allergen components not shown in this and subsequent figures). A coloured square represents a c-sIgE>0.3 and colours represent asthma severity.

*[Figure 6 goes here (Figure\_6.tiff) ]*

**Figure 6. Correlation network for allergen component sensitisation in mild/moderate versus severe asthma in the (a-b) adult and (c-d) school age cohorts**

Connections are based on Spearman’s rank-correlation coefficient (r). Pairs of c-sIgE components are connected if their correlation is significant (p-value <0.05). Node colours represent allergen component cluster memberships (as per Figure 3), node diameter is proportional to the scaled connectivity of the specific component, while edge width represents the strength of correlation between pairs of c-sIgE components, while colours captures the differential connectivity. Different shades of grey indicate strength of associations: black edges (line) indicate r>0.8, dark grey 0.6<r≤0.8,light grey 0.4<r≤0.6 and light grey r≤0.4.

*[Figure 7 goes here (Figure\_7.tiff) ]*

**Figure 7. Differential correlation network analysis** **for allergen component sensitisation between severe and mild/moderate asthma in adult (A) and school age (B) cohorts**

Only significant changes in correlation between the mild\moderate and severe asthma are shown. Significance is assessed by controlling the false discovery rate at 5%. Edge (line) colours indicate correlation gain (difference in correlation) in the conditions: green indicates higher absolute correlation value in the mild\moderate asthma, while red indicates higher absolute correlation in the severe asthma.