Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY

Allergy

ORIGINAL ARTICLE EPIDEMIOLOGY AND GENETICS

Phenotyping asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic comorbidity cluster

J. Garcia-Aymerich[1,2,3], M. Benet[1,2,3], Y. Saeys[4], M. Pinart[1,2,3,5], X. Basagaña[1,2,3], H. A. Smit[6], V. Siroux[7,8], J. Just[9], I. Momas[10,11], F. Rancière[10,11], T. Keil[12,13], C. Hohmann[12], S. Lau[14], U. Wahn[14], J. Heinrich[15], C. G. Tischer[15], M. P. Fantini[16], J. Lenzi[16], D. Porta[17], G. H. Koppelman[18], D. S. Postma[19], D. Berdel[20], S. Koletzko[21], M. Kerkhof[22], U. Gehring[23], M. Wickman[24,25], E. Melén[24,25], J. Hallberg[24,25], C. Bindslev-Jensen[26], E. Eller[26], I. Kull[26], K. C. Lødrup Carlsen[27], K.-H. Carlsen[27], B. N. Lambrecht[4], M. Kogevinas[1,2,3,5,28], J. Sunyer[1,2,3,5], F. Kauffmann[29,30], J. Bousquet[31] & J. M. Antó[1,2,3,5]

[1]Centre for Research in Environmental Epidemiology (CREAL); [2]CIBER Epidemiología y Salud Pública (CIBERESP); [3]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; [4]VIB Inflammation Research Center and Department of Respiratory Medicine, Ghent University, Ghent, Belgium; [5]IMIM (Hospital del Mar Research Institute), Barcelona, Spain; [6]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; [7]Inserm, U823, Institut Albert Bonniot, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health; [8]Université Joseph Fourier, Grenoble; [9]Groupe Hospitalier Trousseau-La Roche-Guyon, Centre de l'Asthme et des Allergies, APHP, Université Paris 6; [10]Department of Public Health and Biostatistics, EA 4064, Paris Descartes University; [11]Paris Municipal Department of Social Action, Childhood, and Health, Paris, France; [12]Institute of Social Medicine, Epidemiology and Health Economics, Charité – Universitätsmedizin Berlin, Berlin; [13]Institute of Clinical Epidemiology and Biometry, University of Wuerzburg, Wuerzburg; [14]Department of Pediatrics, Charité – Universitätsmedizin Berlin, Berlin; [15]Institute of Epidemiology I, Helmholtz Zentrum, Munich, Germany; [16]Department of Biomedical and Neuromotor Sciences, Alma Mater Studiorum – University of Bologna, Bologna; [17]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy; [18]Pediatric Pulmonology and Pediatric Allergology, GRIAC Research Institute, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen; [19]Department of Pulmonology, GRIAC Research Institute, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; [20]Department of Pediatrics, Research Institute, Marien-Hospital Wesel, Wesel; [21]Division of Pediatric Gastroenterology and Hepatology, Dr von Haunersches Kinderspital, Ludwig-Maximilians-University of Munich, Munich, Germany; [22]Department of Epidemiology, GRIAC Research Institute, University Medical Center Groningen, University of Groningen, Groningen; [23]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; [24]Institute of Environmental Medicine, Karolinska Institutet; [25]Sach's Children's Hospital, Stockholm, Sweden; [26]Department of Dermatology and Allergy Centre, Odense University Hospital, Odense, Denmark; [27]Department of Paediatrics, Oslo University Hospital and University of Oslo, Oslo, Norway; [28]National School of Public Health, Athens, Greece; [29]Inserm (Institut National de la Santé et de la Recherche Médicale), CESP (Centre de recherche en Épidémiologie et Santé des Populations), U1018, Respiratory and Environmental Epidemiology Team; [30]UMRS 1018, Université Paris Sud 11, Villejuif; [31]Hôpital Arnaud de Villeneuve, University Hospital of Montpellier, Montpellier, France

To cite this article: Garcia-Aymerich J, Benet M, Saeys Y, Pinart M, Basagaña X, Smit HA, Siroux V, Just J, Momas I, Rancière F, Keil T, Hohmann C, Lau S, Wahn U, Heinrich J, Tischer CG, Fantini MP, Lenzi J, Porta D, Koppelman GH, Postma DS, Berdel D, Koletzko S, Kerkhof M, Gehring U, Wickman M, Melén E, Hallberg J, Bindslev-Jensen C, Eller E, Kull I, Lødrup Carlsen KC, Carlsen K-H, Lambrecht BN, Kogevinas M, Sunyer J, Kauffmann F, Bousquet J, Antó JM. Phenotyping asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic comorbidity cluster. *Allergy* 2015; **70**: 973–984.

Keywords
allergy; asthma; cluster analysis; eczema; rhinitis.

Correspondence
Judith Garcia-Aymerich, MD, PhD, Centre for Research in Environmental Epidemiology, Doctor Aiguader 88, 08003 Barcelona, Spain.
Tel.: +34932147350
Fax: +34932147302
E-mail: jgarcia@creal.cat

Abstract

Background: Asthma, rhinitis and eczema often co-occur in children, but their interrelationships at the population level have been poorly addressed. We assessed co-occurrence of childhood asthma, rhinitis and eczema using unsupervised statistical techniques.

Methods: We included 17 209 children at 4 years and 14 585 at 8 years from seven European population-based birth cohorts (MeDALL project). At each age period, children were grouped, using partitioning cluster analysis, according to the distribution of 23 variables covering symptoms 'ever' and 'in the last 12 months', doctor diagnosis, age of onset and treatments of asthma, rhinitis and eczema; immunoglobulin E sensitization; weight; and height. We tested the

Accepted for publication 24 April 2015

DOI:10.1111/all.12640

Edited by: Werner Aberer

sensitivity of our estimates to subject and variable selections, and to different statistical approaches, including latent class analysis and self-organizing maps.
Results: Two groups were identified as the optimal way to cluster the data at both age periods and in all sensitivity analyses. The first (reference) group at 4 and 8 years (including 70% and 79% of children, respectively) was characterized by a low prevalence of symptoms and sensitization, whereas the second (symptomatic) group exhibited more frequent symptoms and sensitization. Ninety-nine percentage of children with comorbidities (co-occurrence of asthma, rhinitis and/or eczema) were included in the symptomatic group at both ages. The children's characteristics in both groups were consistent in all sensitivity analyses.
Conclusion: At 4 and 8 years, at the population level, asthma, rhinitis and eczema can be classified together as an allergic comorbidity cluster. Future research including time-repeated assessments and biological data will help understanding the interrelationships between these diseases.

Allergy-related diseases, including asthma, rhinitis and eczema, are very common (1). Their characteristics and management are well established, but gaps exist in their causes, mechanisms, diagnosis and prevention (2–4).

Allergy-related diseases often co-occur in the same subjects as comorbidities (2), but this co-occurrence has been seldom studied at the population level. Firstly, using a classical approach defining the diseases by experts on symptom-based definitions and self-reported diagnoses, we studied 17 000 children from 12 ongoing European birth cohort studies participating in Mechanisms of the Development of ALLergy (MeDALL) (5). We showed that co-occurrence of asthma, rhinitis and eczema is more common (50% higher) than expected by chance, in both the presence and absence of immunoglobulin E (IgE) sensitization (6). Another approach applies unsupervised machine learning methods to several characteristics including symptoms to identify their distribution in a population. In the PARIS birth cohort, latent class and transition analyses at 4 years revealed four distinct phenotypes: 'transient rhinitis', 'transient wheeze', 'persistent cough/rhinitis' and 'persistent dermatitis', the latter two associated with IgE sensitization (7, 8). Most comorbidity of asthma, rhinitis and eczema was observed in the 'cough/rhinitis' phenotype. Another study assessed repeatedly 9801 children in two population-based British birth cohorts (ALSPAC and MAAS) using Bayesian machine learning methods to identify developmental profiles of symptoms over time (9). The study revealed eight latent classes, four of which (accounting for 16% of children) included comorbidity of asthma, rhinitis and/or eczema. However, the authors concluded that this comorbidity was likely due to chance, while IgE sensitization, the most frequently considered common mechanism of allergy-related diseases, was not assessed.

To further advance the understanding of comorbidities of allergy-related diseases at the population level, we assessed the presence of different allergic phenotypes using unsupervised (hypothesis-free) statistical techniques in children at 4 and 8 years of age from seven European population-based birth cohorts as part of the MeDALL project. Additionally, we tested whether IgE sensitization modified the classification of allergy-related symptoms in these children.

Abbreviations

BAMSE, The Stockholm Children Allergy and Environmental Prospective Birth Cohort Study; DARC, The Danish Allergy Research Centre; GINIplus, German Infant Study on the influence of Nutrition Intervention plus environmental and genetic influences on allergy development study; IgE, immunoglobulin E; LISAplus, The Influence of Lifestyle factors on the development of the Immune System and Allergies in West Germany plus the influence of environment and genetics study; MAS, Multicentre Allergy Study; MeDALL, Mechanisms of the Development of Allergy; PARIS, Pollution and Asthma Risk: an Infant Study; PIAMA, Prevention and Incidence of Asthma and Mite Allergy.

Methods

Design and study population

A cross-sectional analysis of birth cohort studies was carried out at 4 years (ranging from 3 to 5) and 8 years (ranging from 8 to 10). The children were selected from seven MeDALL cohorts [BAMSE (10), Sweden; DARC (11, 12), Denmark; GINIplus (13), LISAplus (14) and MAS (15), Germany; PARIS (16), France; and PIAMA (17), the Netherlands]. Inclusion and exclusion criteria are presented in Supporting Information. The sample size for each cohort and period ranged from 505 to 4299 children (Supporting Information). In all participating cohorts, parents gave written informed consent and local ethics review boards approved the studies.

Measurements

We pooled and harmonized the data collected from questionnaires on 20 variables covering symptoms 'ever' and 'in the last 12 months', doctor diagnosis, age of onset and treatment of allergy-related diseases (asthma, rhinitis and eczema) (Table S1). Children's weight and height were obtained from physical examination. Sensitization was defined by serum-specific IgE $\geq$0.35 kUA/l against at least one of the following aeroallergens and food allergens: house dust mite, cat dander, birch pollens, grass pollens, cow's milk and egg (6). We also

defined current asthma, rhinitis and eczema using the classical definitions (6). Asthma, rhinitis and eczema comorbidities were defined as the co-occurrence of two or three of these diseases in the same child (Supporting Information).

Statistical analysis

The number of subjects available was greater than required according to sample size calculations (Supporting Information). We assessed the presence and patterns of missing values and, assuming the missing-at-random hypothesis (18), we used multiple imputation (20 imputed data sets) with the method of chained equations (19). We compared the characteristics of children in the complete case to the imputed data sets (Supporting Information).

For the unsupervised analysis, we included 23 variables: all 20 features of allergy-related diseases, weight, height and IgE sensitization. All variables were standardized using *Z*-scores; in a secondary analysis, variables were scaled from 0 to 1. We did not perform any data reduction (e.g. factor analysis) prior to clustering, because there was not a high degree of colinearity in our variables (Fig. S2), to avoid losing the amount of variance that is not explained in such preprocessing and to facilitate the interpretation of clusters (20).

Primarily, we used *k*-means partitioning cluster analysis, which groups subjects according to the Euclidean distance between the included variables (21). This analysis was performed at both time periods, 4 and 8 years, for each of the 20 data sets generated by the multiple imputation method, following a method previously reported to integrate multiple imputation in cluster analyses (22). We selected the number of groups (clusters) that maximized the Calinski–Harabasz stopping rule if and only if it was in agreement with another stopping rule (the average silhouette width) and with consensus measures, and it could not be attributed to chance (Supporting Information). To test the longitudinal stability of the identified clusters, we compared the groups to which children belonged between 4 and 8 years using cross-tabulation. We evaluated the role of IgE sensitization in the cluster analysis by performing all analyses both including and excluding IgE from the cluster model and by stratifying the cluster analysis according to IgE sensitization.

As part of our sensitivity analysis, we tested whether alternative hypothesis-free grouping methods could have yielded different results, repeating all analyses using (i) hierarchical clustering with Ward's method, (ii) latent class analysis and (iii) self-organizing maps (Supporting Information). We also performed several secondary analyses to assess the sensitivity of our estimates against our assumptions regarding selection bias and information bias, as well as to test for model misspecification (Supporting Information).

For the graphical description of the groups identified by cluster analysis, we plotted the prevalence of each variable in each group with a colour intensity scale spanning from white (prevalence of 0%) to red (prevalence of 100%). We compared the distribution of all 23 variables across groups and calculated the relative relevance of each variable to the separation in cluster groups using *F* values [the ratio of the variance of the group means (between-group variance) over the overall variance of the variable, where higher values indicate higher relevance of the variable for separating cluster groups]. We also assessed the distribution of the classical definitions of current asthma, rhinitis and eczema, as well as their comorbidity, according to cluster groups.

All analyses were performed using Stata 12 (Stata Statistical Software: Release 12; StataCorp LP, College Station, TX, USA) and R 2.14.2 (R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/).

Table 1 Characteristics and symptoms of asthma, rhinitis and eczema in participating children at 4 and 8 years

	4 years *n* = 17 209 *n* (%)	8 years *n* = 14 585 *n* (%)
Sex: female	8354 (48.5)	7060 (48.4)
Age (months), M (SD)	46.9 (5.0)	106.4 (12.1)
Wheezing ever	5641 (32.8)	5767 (39.5)
Wheezing attacks in the last 12 months		
None	15 309 (89.0)	13 112 (89.9)
1–3 times	1289 (7.5)	1005 (6.9)
4–12 times	482 (2.8)	358 (2.5)
>12 times	129 (0.8)	110 (0.8)
Wheezing after exercise ever	1346 (7.8)	2345 (16.1)
Asthma ever	1410 (8.2)	2243 (15.4)
Asthma treatment in the last 12 months	1936 (11.3)	1371 (9.4)
Asthma onset before 2 years of age	924 (5.4)	879 (6.0)
Bronchitis or bronchiolitis ever	5794 (33.7)	5760 (39.5)
Cough at night (when no cold) ever	4948 (28.8)	6189 (42.4)
Sneezing or runny or blocked nose (when no cold) ever	5607 (32.6)	6392 (43.8)
Sneezing or runny or blocked nose (when no cold) in the last 12 months	2474 (14.4)	3400 (23.3)
Itchy watery eyes (when no cold) in the last 12 months	831 (4.8)	1845 (12.7)
Allergic rhinitis ever	648 (3.8)	2326 (15.9)
Rhinitis onset before 2 years of age	876 (5.1)	345 (2.4)
Itchy rash (coming and going for at least 6 months) ever	6290 (36.6)	6921 (47.5)
Itchy rash (coming and going for at least 6 months) in the last 12 months	3353 (19.5)	2126 (14.6)
Itchy rash affecting common areas	4820 (28.0)	1657 (11.4)
Itchy rash onset before 2 years of age	3734 (21.7)	3477 (23.8)
Eczema ever	4614 (26.8)	5049 (34.6)
Urticaria ever	3403 (19.8)	3043 (20.9)
Food allergy ever	1850 (10.7)	2699 (18.5)
IgE sensitization	3611 (21.0)	5680 (38.9)
Weight (kg), M (SD)	17.0 (2.7)	32.3 (7.7)
Height (cm), M (SD)	103.8 (6.0)	137.9 (9.4)

*A total of 14 383 children had data available at both age periods.

Results

A total of 17 209 children were included at 4 years [49% female, mean (SD) 46.9 (5.0) months] and 14 585 at 8 years [48% female, 106.4 (12.1) months] (Tables 1 and S4). Lifetime (ever) prevalences of asthma, rhinitis and eczema were 8.2%, 3.8% and 26.8% at 4 years, and 15.4%, 15.9% and 34.8% at 8 years.

Both the Calinski–Harabasz and the average silhouette width stopping rules (Fig. 1), as well as the consensus matrix (Fig. S3), showed that two groups was the most effective classification of children and was not due to chance; this is that the separation in two clusters resulted in groups of children homogeneous within and heterogeneous between them, while the classification in more than two groups provided mixed groups that moreover were poorly reproducible.

Figure 2 and Table S5 show how the 20 symptoms were distributed in the two cluster groups. At 4 years, Group 1 included 12 052 (70.0%) children with low symptoms prevalence; Group 2 included 30.0% children exhibiting a higher prevalence of most symptoms (22.9% of asthma ever, 10.4% of allergic rhinitis ever, 64.5% of eczema ever). IgE sensitization occurred in 16.6% of children in Group 1 and 31.2% of

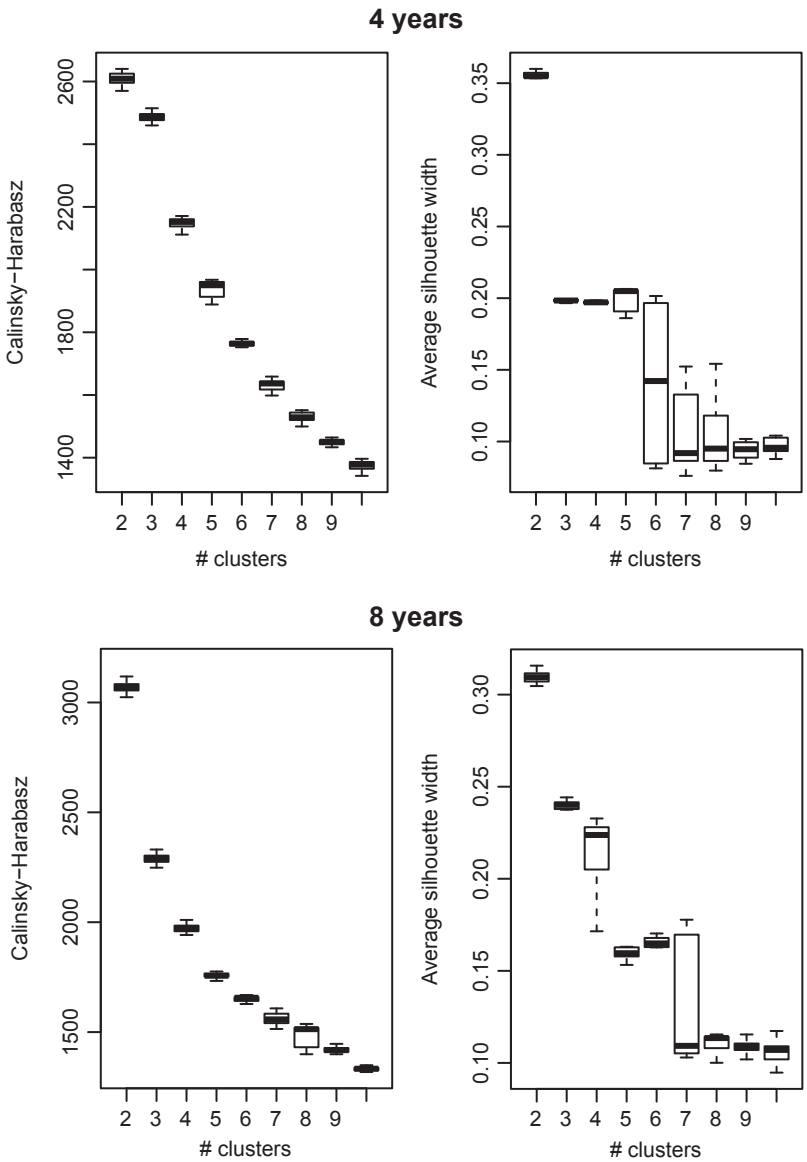

Figure 1 Distribution of values over 20 imputed data sets of the Calinski–Harabasz and average silhouette width stopping rules* across 2–10 cluster groups at 4 and 8 years. *Higher values indicate higher separation between groups and similarity within groups. The *P*-values for the observed values of both stopping rules being generated by their background distributions are 0, so not likely to be observed by chance.

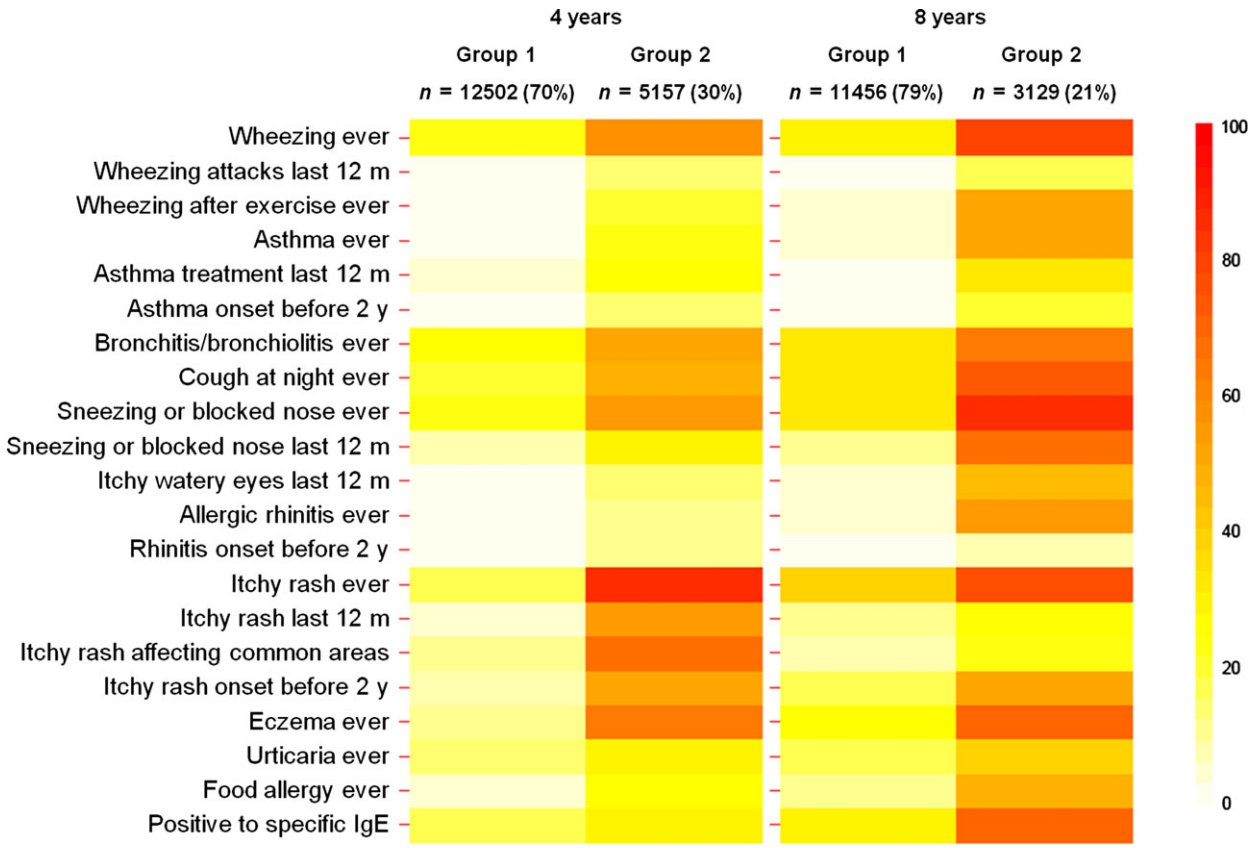

Figure 2 Prevalence* of symptoms of asthma, rhinitis and eczema according to the two groups identified in cluster analysis, at 4 and 8 years. *Each coloured line represents a variable, whose prevalence ranges from 0% (white colour) to 100% (red colour).

Table 2 Stability of group membership of children at 4 and 8 years

	Cluster analysis at 8 years		
	Group 1	Group 2	Total
Cluster analysis at 4 years			
Group 1			
n	8891	924	9815
Row %*	90.6	9.4	
Total %†	61.8	6.4	68.2
Group 2			
n	2624	1944	4568
Row %*	57.4	42.6	
Total %†	18.2	13.5	31.8
Total			
n	11 515	2868	14 383
Total %	80.1	19.9	100

*Proportion of children belonging to Group 1 or 2 at 8 years taking into account their belonging at 4 years.
†Total proportion of children assigned to each combination of cluster groups at 4 and 8 years.

children in Group 2. At 8 years, the results were similar although fewer children (21.5%) were classified in Group 2. A total of 10 835 (75.3%) of 14 383 children belonged to the same group at both 4 and 8 years, but Group 1 was more stable than Group 2 (Table 2).

The prevalences of current asthma, rhinitis and eczema according to classical definitions were higher in Group 2 than in Group 1 (at 8 years: 36.9%, 49.0% and 27.5% *vs* 1.1%, 1.8% and 8.0%) (Fig. 3; Table S6). Almost all children with comorbidity of asthma, rhinitis and eczema were included in Group 2 at both 4 and 8 years.

The classification in three groups (Fig. S4) showed a similar Group 1 (59.7% of children) with a low prevalence of symptoms and sensitization, a Group 2 (15.6%) with a higher prevalence of symptoms of asthma and rhinitis and a Group 3 (24.7%) with higher proportions of eczema symptoms.

The prevalences of symptoms in groups 1 and 2 were almost identical with and without including IgE sensitization in the cluster analysis. After stratifying the cluster analysis according to IgE sensitization, the pattern of differences between groups 1 and 2 was maintained, although prevalences of symptoms and diseases were higher in the IgE-sensitized children (Figs 4, S5 and S6; Tables S7 and S8).

Sensitivity analyses showed little change in response to changes in assumptions regarding statistical models, as well as selection and information biases (Figs 5 and S7–S20; Tables S9–S19). Alternative hypothesis-free grouping

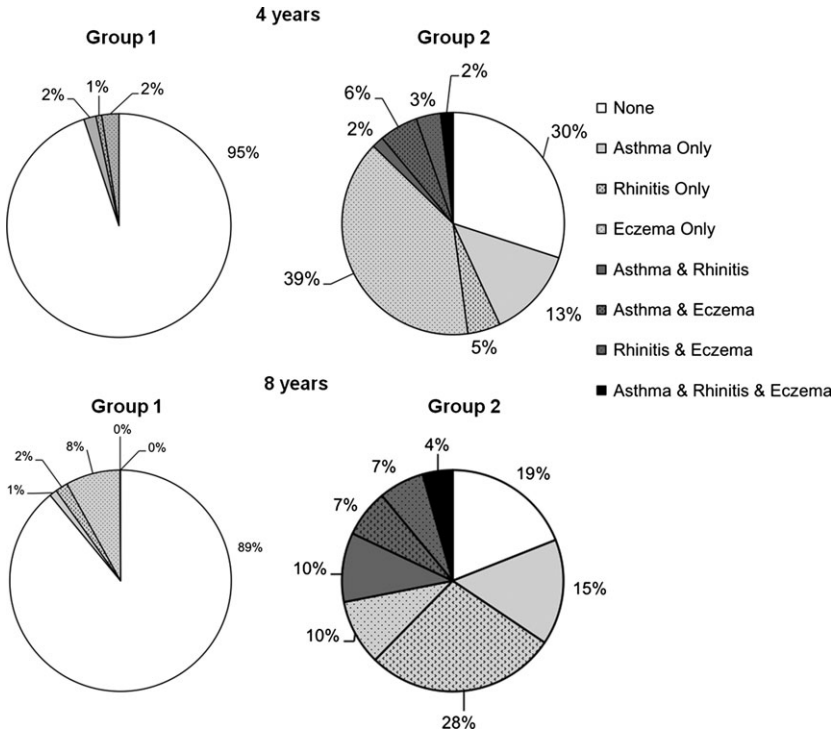

Figure 3 Distribution of classical definitions of current asthma, rhinitis, eczema and their comorbidity, according to the two groups identified in cluster analysis, at 4 and 8 years.

methods (hierarchical clustering with Ward's method, latent class analysis and self-organizing maps) also showed the organization of children into two groups as the best option. The same classification was obtained after considering variables with potential problems of measurement error, that is after excluding 'itchy rash ever' and 'food allergy ever' (one at a time), using a higher cut-off for IgE sensitization (≥3.5 kUA/l) and using body mass index instead of weight and height separately. Likewise, the inclusion of additional variables (spirometry, bronchial responsiveness, airway resistance (Rint), skin prick test, exhaled nitric oxide (FeNO) and others; Supporting Information) in a subset of two birth cohorts (PIAMA and BAMSE) also supported two groups, as well as did the stratification by birth cohort or by the proportion of missing data. In all sensitivity analyses, Group 2 showed a higher prevalence of allergy-related diseases and sensitization.

Discussion

Using hypothesis-free statistical analyses, we identified two groups of children at 4 and 8 years of age from seven population-based birth cohorts: a reference group (70.0% at 4 years and 78.5% at 8 years) with low frequencies of asthma, rhinitis and eczema symptoms and a symptomatic group (30.0% at 4 years and 21.5% at 8 years) with high frequencies of symptoms of the three diseases. The symptomatic group presented 99% comorbidity. While IgE sensitization was more prevalent in the symptomatic group (31.2% *vs* 16.6% at 4 years; 71.7% *vs* 30.0% at 8 years), the distribution of symptoms across cluster groups did not change according to the inclusion or exclusion of IgE sensitization in the model. The sensitivity analysis showed that classification in two groups was very stable in relation to changes in the selection of subjects and variables and the use of different clustering methods.

Strengths and limitations

Our study is based on a large network of European birth cohorts (7) including a large sample size, a wide geographical and environmental variability and the harmonization of standardized questionnaires. To avoid losing information due to missing values, we performed multiple imputations, a valid solution in cluster analysis (22). Information bias during questionnaire-based symptom assessment cannot be excluded although we do not expect this bias differentially distributed across cluster groups. Input data are restricted to information available in all seven cohorts, but our results did not change in the sensitivity analyses including lung function and biomarkers. Having only two time points and lacking data from the first year of life might have hampered the study of the development over time of such complex diseases, as elegantly performed in the unsupervised analysis of the ALSPAC and MAAS cohorts (9). Nevertheless, our assessment of changes in group membership from 4 to 8 years suggested substantial temporal stability. Unsupervised methods for the classification of subjects may differ according to the type of

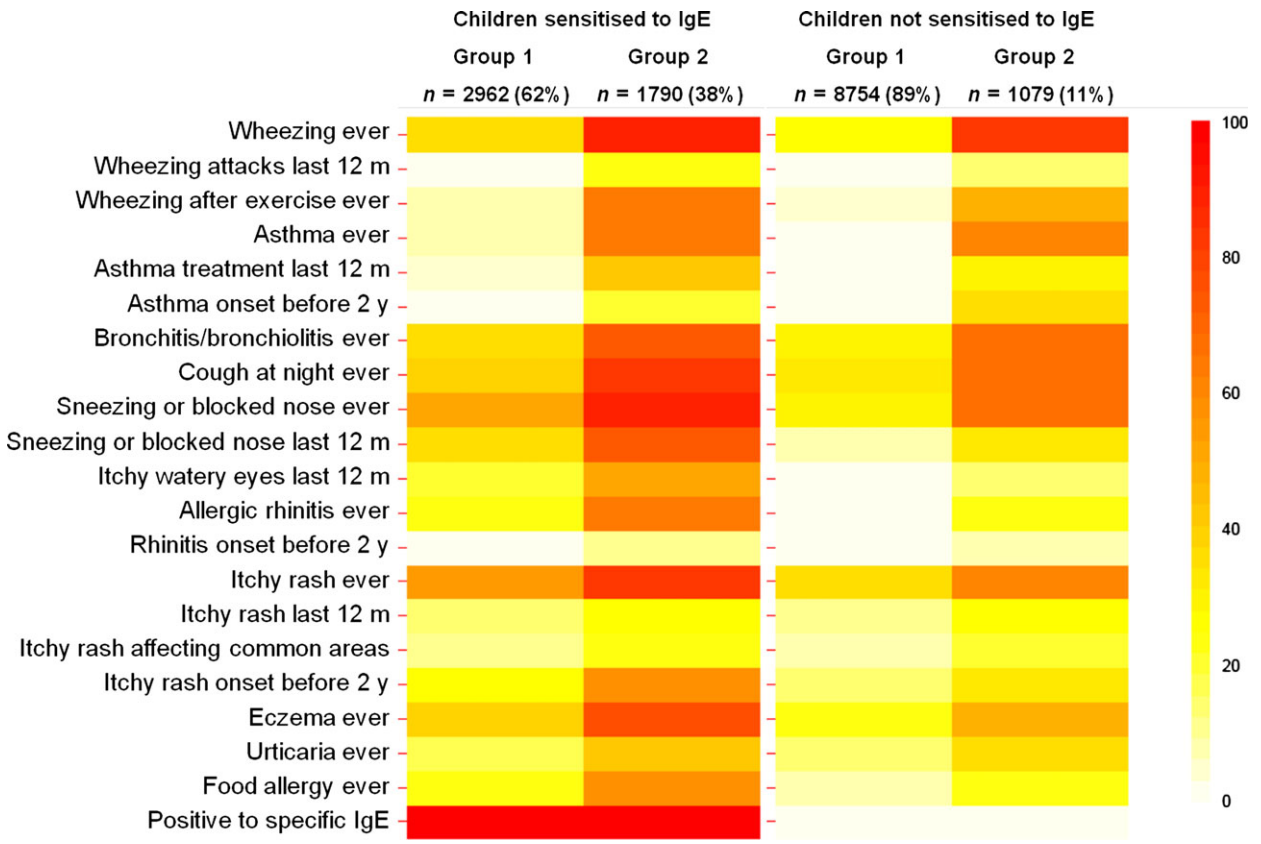

Figure 4 Prevalence* of symptoms of asthma, rhinitis and eczema according to the two groups identified in cluster analysis at 8 years, according to IgE sensitization. *Each coloured line represents a variable, whose prevalence ranges from 0% (white colour) to 100% (red colour).

modelling and stopping rules and are conditional to a number of assumptions and analytical decisions. We used a wide range of models and tested our assumptions in the sensitivity analyses, which indicated that our results are very stable.

Consistency with previous studies

Most previous studies using unsupervised methods have focused on a single allergic disease (23–36). Based on these results, it was expected that our study would identify several groups, separating asthma, rhinitis and eczema symptoms. However, our results show that, at the population level, most children with symptoms of asthma, rhinitis and eczema are better classified together, in a single symptomatic group. Moreover, 99% children with comorbidity were classified in this symptomatic group. Overall, results strongly suggest the existence of an allergic comorbidity cluster. Some studies focusing on wheezing/asthma in children, applying unsupervised methods, found similar prevalences of the other allergy-related diseases (rhinitis and eczema) among all identified wheezing groups, which is consistent with our results (25, 30).

Conversely, two previous unsupervised studies that assessed the joint distribution of asthma, rhinitis and eczema obtained several groups (7–9). Relevant methodological differences could explain the diverse results. First, previous studies included a reduced number of allergy-related symptoms or diseases diagnostics, while our analysis used a large diversity of the diagnostic signs, symptoms and biomarkers of three different organ-related diseases (lungs, nose and skin). Second, and as a consequence of the first, they used longitudinal clustering techniques (feasible for small number of variables) while we performed cross-sectional clustering at two time points. Interestingly, in our three-group solution, children with a high prevalence of rhinitis and asthma symptoms tended to remain in the same group, whereas a third group of children with the highest prevalence of eczema symptoms emerged. However, this separation was largely driven (according to F values, see Methods) by the symptom 'itchy rash ever', a variable with known potential problems of missclassification. Remarkably, our sensitivity analyses confirmed that two-group classification is a better option than classification into a larger number of groups. The additional analysis with the P-values of stopping rules showed that it is very unlikely to find such two groups clustering structure by chance. Therefore, and although unexpected, we conclude that childhood symptoms of asthma, rhinitis and eczema can be classified in two groups at the population level, which requires replication in further studies.

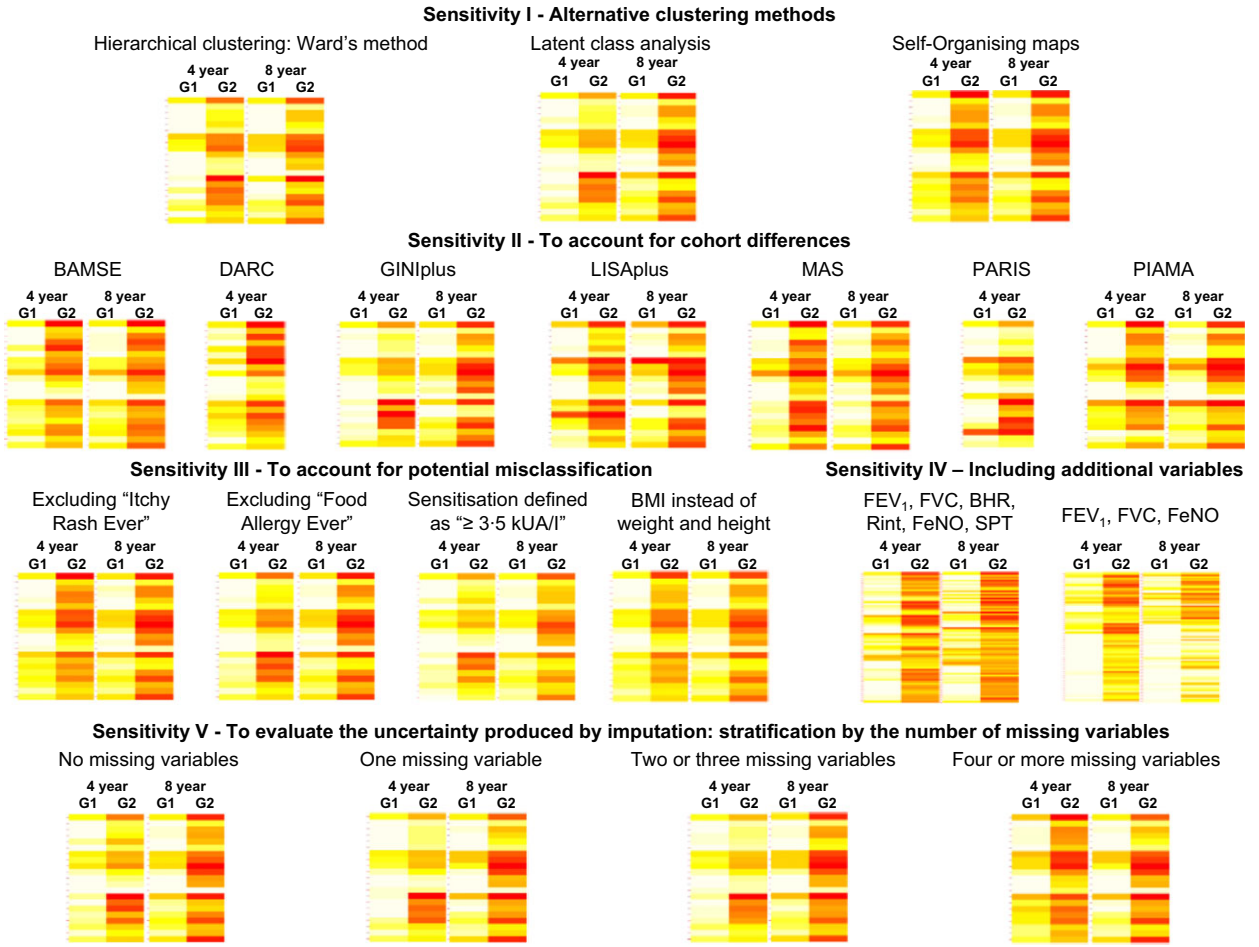

Figure 5 Prevalence* of symptoms of asthma, rhinitis and eczema according to the two groups identified in cluster analysis, at 4 and 8 years, in sensitivity analyses. *Each coloured line represents a variable, whose prevalence ranges from 0% (white colour) to 100% (red colour).

The influence of IgE sensitization appeared to be minor, as the classification was similar after stratification by IgE sensitization, as well as after including or excluding the IgE sensitization variable. The PARIS study had shown that the two persistent phenotypes (cough/rhinitis and dermatitis) were associated with IgE sensitization, whereas the two transient phenotypes (rhinitis and wheeze) were not (7). However, the two studies are not directly comparable due to dissimilar study design and analytical approaches. By contrast, the present findings are consistent with results from our previous MeDALL study which examined a larger number of birth cohorts and showed that the strong tendency of asthma, rhinitis and eczema to co-occur in the same children was independent of IgE sensitization (6). Both the present unsupervised and the previous supervised MeDALL studies strongly suggest that specific IgE contributes to the comorbidity cluster of asthma, rhinitis and eczema but should no longer be considered its dominant mechanism.

Several mechanisms other than IgE can be proposed as responsible for the identified comorbidity cluster. Some studies have reported common genetic determinants of asthma, rhinitis and eczema, such as the filaggrin gene (37), the leucine-rich repeat-containing 32 gene (*LRRC32*) (38) and the 17q21 locus (39). Comorbidity cluster could also be the clinical expression of the effects of common environmental factors, but few studies have focused on environmental determinants of allergy-related comorbidity. A bioinformatics approach to analysing allergy-related diseases in European children combined feature selection and machine learning to show that combinations of environmental and lifestyle factors were more frequently related to allergy-related diseases than combinations solely involving genes (40). However, the study did not assess comorbidity determinants. Overall, understanding the mechanisms of the identified allergic comorbidity cluster warrants further research.

Implications

The current emphasis on assessing the heterogeneity of allergy-related entities and searching for meaningful subgroups should be balanced with increased efforts to understand their interrelationships (41). Although comorbidities of allergy-related diseases are well known, the cluster found in the present paper suggests that undisclosed mechanisms

underlying the three diseases need to be investigated using new research approaches and concepts, such as the diseasome (42) or the integrative systems biology models (43). At the clinical level, the present study supports an integration of care pathways in children with allergy-related diseases.

Conclusion

Our study has shown that, at the population level, childhood asthma, rhinitis and eczema are more accurately classified together as an allergic comorbidity cluster, than as three independent diseases.

Future research including time-repeated assessments and biological data will help understanding the interrelationships between these diseases.

Acknowledgments

We are grateful to Andrea von Berg (GINIplus and LISAplus) and Alet H. Wijga (PIAMA) for providing data and to Isabella Annesi-Maesano, Maties Torrent, Renato T. Stein and Stefano Guerra for their comments to the study protocol and/or interpretation of results.

Funding

This work was supported by Mechanisms of the Development of ALLergy (MeDALL), a collaborative project conducted within the European Union under the Health Cooperation Work Programme of the 7th Framework programme (grant agreement No. 261357). The sponsor of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all study data and had final responsibility for the decision to submit for publication.

Conflicts of interest

The authors declare no conflict of interest. Dr Lau reports grants from the German Research Foundation during the conduct of the study. Prof. Koppelman reports grants from Dutch Lung Foundation, outside the submitted work. Prof. Postma has received money from Astra Zeneca, Chiesi, Boehringer Ingelheim, GSK, Takeda and TEVA for consultancies, outside the submitted work. Dr Koletzko reports personal fees from Centocor, MSD, Danone, Merck, Vifor, Nestle Nutrition Institute, Euroimmun, Thermo-Fischer (Phadia), Abbvie, Schär, Hipp, Falk and Grants or money for research collaboration for clinical trials from Nestle Nutrition Institute, Mead Johnson, Euroimmun, Eurospital, Inova, R-Biopharm, Schär, ThermoFisher, outside the submitted work. Professor M. Wickman has received a grant for research (reagents and analysis costs) from Thermo Fisher Scientific, Uppsala, Sweden. Prof. KH Carlsen reports personal fees from Meda, Boehringer Ingelheim and Nycomed, outside the submitted work. Prof. Kauffmann reports grants from EU during the conduct of the study. Prof Jean Bousquet has received money from Stallergènes for board membership; from Actelion, Almirall, Meda, Merck, Merck Sharp Dohme, Novartis, Sanofi-Aventis, Takeda, Teva and Uriach for consultancies; and from Almirall, AstraZeneca, Chiesi, GlaxoSmithKline, Meda, Merck, Merck Sharp Dohme, Novartis, OM Pharma, Sanofi-Aventis, Schering-Plough, Takeda, Teva and Uriach for lectures and/or educational presentations.

Author contributions

JG-A wrote the initial draft. MB prepared the common database. MB and YS carried out statistical analysis. JG-A, MB, MP, XB and JMA had full access to the data and participated in the interpretation of the findings. MW, EM, IK, JH (BAMSE); CBJ, EE (DARC); JH, SK, CGT, DB (GINIplus and LISAplus); TK, CH, SL, UW (MAS); IM, FR, JJ (PARIS); and HAS, MK, UG, GK (PIAMA) provided data. All authors (i) provided substantial contributions to the conception or design of the work, or the acquisition, analysis or interpretation of data for the work, (ii) revised the manuscript for important intellectual content, (iii) approved the final version and (iv) agreed to be accountable for all aspects of the work. JB and JMA coordinate the MeDALL project.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Distribution of the number of variables with missing values, per period.

Figure S2. Correlations between the 23 variables included in the cluster analysis (20 features of allergy-related diseases, weight, height, and age).

Figure S3. Consensus matrix* for 2 and 3 cluster groups, at 4 and 8 years.

Figure S4. Prevalence* of symptoms of asthma, rhinitis, and eczema according to three groups identified in cluster analysis, at 4 and 8 years.

Figure S5. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, without including IgE sensitisation as a variable in the cluster analysis.

Figure S6. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, after stratifying the cluster analysis according to IgE sensitisation.

Figure S7. Distribution of Calinsky–Harabasz and Average silhouette width stopping rules*, and graphical description† of the two groups identified by hierarchical analysis at 4 and 8 years.

Figure S8. Distribution of Calinsky–Harabasz stopping rule*, and graphical description† of the two groups identified by latent class analysis at 4 and 8 years.

Figure S9. Distribution of Calinsky–Harabasz and Average silhouette width stopping rules*, and graphical description† of the two groups identified by self-organising maps analysis at 4 and 8 years.

Figure S10. Distribution of average values over 20 imputed datasets of the Calinsky–Harabasz and Average silhouette width stopping rules* in several cluster strategies using variables scaled from 0 to 1, across 2–8 cluster groups at 4 and 8 years.

Figure S11. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, after stratifying cluster analysis according to cohort.

Figure S12. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, without including 'itchy rash ever' as a variable in the cluster analysis.

Figure S13. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, without including 'food allergy ever' as a variable in the cluster analysis.

Figure S14. Values of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, using a ≥3.5 kUA/l cut-off for IgE-sensitisation in the cluster analysis.

Figure S15. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, including BMI (instead of weight and height separately) as a variable in the cluster analysis.

Figure S16. Distribution of values over 20 imputed datasets of the Calinski-Harabasz stopping rule* across 2–8 cluster groups at 4 and 8 years, restricting analysis to a subset of the PIAMA cohort (including 67 variables in 480 children at 4 year and 76 variables in 774 children at 8 year).

Figure S17. Prevalence* of symptoms of asthma, rhinitis, and eczema according to the two groups identified in cluster analysis, at 4 and 8 years, restricting analysis to a subset of the PIAMA cohort (including 67 variables† in 480 children at 4 year and 76 variables† in 774 children at 8 year).

Figure S18. Distribution of values over 20 imputed datasets of the Calinski-Harabasz stopping rule* across 2–8 cluster groups at 4 and 8 years, restricting analysis to a subset of the BAMSE cohort (including 61 variables in 3993 children at 4 year and 86 variables in 457 children at 8 year).

Figure S19. Prevalence* of symptoms of asthma, rhinitis, and eczema according to the two groups identified in cluster analysis, at 4 and 8 years, restricting analysis to a subset of the BAMSE cohort (including 61 variables† in 3993 children at 4 year and 86 variables† in 457 children at 8 year).

Figure S20. Distribution of Calinsky–Harabasz stopping rule* and graphical description† of the two groups identified by cluster analysis at 4 and 8 years, after stratifying according to the number of missings.

Table S1. List of questions of allergy-related diseases.

Table S2. Number and proportion of missing values per variable, cohort and period.

Table S3. Characteristics of participating children at 4 and 8 years, using complete cases and imputed datasets.

Table S4. Characteristics of children at 4 and 8 years, by birth cohort.

Table S5. Description of the two groups identified by cluster analysis at 4 and 8 years.

Table S6. Distribution of classical phenotypes of current asthma, rhinitis, eczema, and their comorbidity, according to the two groups identified in cluster analysis, at 4 and 8 years.

Table S7. Description of the two groups identified by cluster analysis at 4 and 8 years, without including IgE sensitisation as a variable in the cluster analysis.

Table S8. Description of the two groups identified by cluster analysis at 4 and 8 years, after stratifying the cluster analysis according to IgE sensitisation.

Table S9. Description of the two groups identified by cluster analysis at 4 and 8 years, using hierarchical analysis as the clustering method.

Table S10. Description of the two groups identified by cluster analysis at 4 and 8 years, using latent class analysis as the clustering method.

Table S11. Description of the two groups identified by cluster analysis at 4 and 8 years, using self-organising maps as the clustering method.

Table S12. Description of the two groups identified by cluster analysis at 4 and 8 years after, after stratifying cluster analysis according to cohort.

Table S13. Description of the two groups identified by cluster analysis at 4 and 8 years after, without including 'itchy rash ever' as a variable in the cluster analysis.

Table S14. Description of the two groups identified by cluster analysis at 4 and 8 years after, without including 'food allergy ever' as a variable in the cluster analysis.

Table S15. Description of the two groups identified by cluster analysis at 4 and 8 years after, using a ≥3.5 kUA/l cut-off for IgE-sensitisation in the cluster analysis.

Table S16. Description of the two groups identified by cluster analysis at 4 and 8 years, including BMI (instead of weight and height separately) as a variable in the cluster analysis.

Table S17. Description of the two groups identified by cluster analysis at 4 and 8 years, restricting analysis to a subset of the PIAMA cohort (including 67 variables in 480 children at 4 year and 76 variables in 774 children at 8 year).

Table S18. Description of the two groups identified by cluster analysis at 4 and 8 years, restricting analysis to a subset of the BAMSE cohort (including 61 variables in 3993 children at 4 year and 86 variables in 457 children at 8 year).

Table S19. Description of the two groups identified by cluster analysis at 4 and 8 years after stratifying according to the number of missings.

References

1. Johansson SG, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF et al. Revised nomenclature for allergy for global use: report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol* 2004;**113**:832–836.
2. Bousquet J, Khaltaev N, Cruz AA, Denburg J, Fokkens WJ, Togias A et al. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008

update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). *Allergy* 2008;**63**(Suppl 86):8–160.
3. Bieber T. Atopic dermatitis. *N Engl J Med* 2008;**358**:1483–1494.
4. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J* 2008;**31**:143–178.
5. Bousquet J, Anto J, Auffray C, Akdis M, Cambon-Thomsen A, Keil T et al. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy* 2011;**66**:596–604.
6. Pinart M, Benet M, Annesi-Maesano I, von Berg A, Berdel D, Carlsen KC et al. Comorbidity of eczema, rhinitis, and asthma in IgE-sensitised and non-IgE-sensitised children in MeDALL: a population-based cohort study. *Lancet Respir Med* 2014;**2**:131–140.
7. Herr M, Just J, Nikasinovic L, Foucault C, Le Marec AM, Giordanella JP et al. Risk factors and characteristics of respiratory and allergic phenotypes in early childhood. *J Allergy Clin Immunol* 2012;**130**:389–396.
8. Rancière F, Nikasinovic L, Bousquet J, Momas I. Onset and persistence of respiratory/allergic symptoms in preschoolers: new insights from the PARIS birth cohort. *Allergy* 2013;**68**:1158–1167.
9. Belgrave DCM, Granell R, Simpson A, Guiver J, Bishop C, Buchan I et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med* 2014;**11**:e1001748.
10. Kull I, Melen E, Alm J, Hallberg J, Svartengren M, van Hage M et al. Breast-feeding in relation to asthma, lung function, and sensitization in young schoolchildren. *J Allergy Clin Immunol* 2010;**125**:1013–1019.
11. Jøhnke H, Vach W, Norberg LA, Bindslev-Jensen C, Høst A, Andersen KE. A comparison between criteria for diagnosing atopic eczema in infants. *Br J Dermatol* 2005;**153**:352–358.
12. Kjaer HF, Eller E, Høst A, Andersen KE, Bindslev-Jensen C. The prevalence of allergic diseases in an unselected group of 6-year-old children. The DARC birth cohort study. *Pediatr Allergy Immunol* 2008;**19**:737–745.
13. Berg AV, Krämer U, Link E, Bollrath C, Heinrich J, Brockow I et al. Impact of early feeding on childhood eczema: development after nutritional intervention compared with the natural course—the GINIplus study up to the age of 6 years. *Clin Exp Allergy* 2010;**40**:627–636.
14. Heinrich J, Bolte G, Hölscher B, Douwes J, Lehmann I, Fahlbusch B et al. Allergens and endotoxin on mothers' mattresses and total immunoglobulin E in cord blood of neonates. *Eur Respir J* 2002;**20**:617–623.
15. Bergmann RL, Bergmann KE, Lau-Schadensdorf S, Luck W, Dannemann A, Bauer CP et al. Atopic diseases in infancy. The German multicenter atopy study (MAS-90). *Pediatr Allergy Immunol* 1994;**5**(6 Suppl):19–25.
16. Clarisse B, Nikasinovic L, Poinsard R, Just J, Momas I. The Paris prospective birth cohort study: which design and who participates? *Eur J Epidemiol* 2007;**22**:203–210.
17. Brunekreef B, Smit J, de Jongste J, Neijens H, Gerritsen J, Postma D et al. The prevention and incidence of asthma and mite allergy (PIAMA) birth cohort study: design and first results. *Pediatr Allergy Immunol* 2002;**13**(Suppl 15):55–60.
18. Schafer JL. *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall/CRC, 1997.
19. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–694.
20. Dolnicar S, Lazarevski K. Methodological reasons for the theory/practice divide in market segmentation. *J Marketing Manag* 2009;**25**:357–373.
21. Steinley D. K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 2006;**59**:1–34.
22. Basagaña X, Barrera-Gómez J, Benet M, Antó JM, Garcia-Aymerich J. A framework for multiple imputation in cluster analysis. *Am J Epidemiol* 2013;**177**:718–725.
23. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008;**63**:974–980.
24. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol* 2011;**127**:1505–1512.
25. Smith JA, Drake R, Simpson A, Woodcock A, Pickles A, Custovic A. Dimensions of respiratory symptoms in preschool children: population-based birth cohort study. *Am J Respir Crit Care Med* 2008;**177**:1358–1363.
26. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J* 2008;**31**:974–981.
27. Clarisse B, Demattei C, Nikasinovic L, Just J, Daures J-P, Momas I. Bronchial obstructive phenotypes in the first year of life among Paris birth cohort infants. *Pediatr Allergy Immunol* 2009;**20**:126–133.
28. Simpson A, Tan VY, Winn J, Svensén 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**:1200–1206.
29. Mahut B, Peyrard S, Delclaux C. Exhaled nitric oxide and clinical phenotypes of childhood asthma. *Respir Res* 2011;**12**:65.
30. Rancière F, Clarisse B, Nikasinovic L, Just J, Momas I. Cough and dyspnoea may discriminate allergic and infectious respiratory phenotypes in infancy. *Pediatr Allergy Immunol* 2012;**23**:367–375.
31. Weinmayr G, Keller F, Kleiner A, du Prel JB, Garcia-Marcos L, Batllés-Garrido J et al. Asthma phenotypes identified by latent class analysis in the ISAAC phase II Spain study. *Clin Exp Allergy* 2013;**43**:223–232.
32. Garden FL, Simpson JM, Marks GB; CAPS Investigators. Atopy phenotypes in the Childhood Asthma Prevention Study (CAPS) cohort and the relationship with allergic disease: clinical mechanisms in allergic disease. *Clin Exp Allergy* 2013;**43**:633–641.
33. Spycher BD, Silverman M, Pescatore AM, Beardsmore CS, Kuehni CE. Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts. *J Allergy Clin Immunol* 2013;**132**:1058–1067.
34. 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**:764–770.
35. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V et al. Clinical and epidemiologic phenotypes of childhood asthma. *Am J Respir Crit Care Med* 2014;**189**:129–138.
36. Just J, Saint-Pierre P, Gouvis-Echraghi R, Laoudi Y, Roufai L, Momas I et al. Childhood allergic asthma is not a single phenotype. *J Pediatr* 2014;**164**:815–820.
37. van den Oord RA, Sheikh A. Filaggrin gene defects and risk of developing allergic sensitisation and allergic disorders: systematic review and meta-analysis. *BMJ* 2009;**339**:b2433.
38. Weidinger S, Willis-Owen SA, Kamatani Y, Baurecht H, Morar N, Liang L et al. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet* 2013;**22**:4841–4856.
39. Fuertes E, Söderhäll C, Acevedo N, Becker A, Brauer M, Chan-Yeung M et al. Associations between the 17q21 region and allergic rhinitis in 5 birth cohorts. *J Allergy Clin Immunol* 2015;**135**:573–576.
40. Bornelöv S, Sääf A, Melén E, Bergström A, Torabi Moghadam B, Pulkkinen V et al. Rule-based models of the Interplay between

genetic and environmental factors in childhood allergy. *PLoS ONE* 2013;**8**:e80080.
41. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;**380**:37–43.
42. Barabási AL. Network medicine—from obesity to the "diseasome". *N Engl J Med* 2007;**357**:404–407.
43. Auffray C, Adcock IM, Chung KF, Djukanovic R, Pison C, Sterk PJ. An integrative systems biology approach to understanding pulmonary diseases. *Chest* 2010;**137**:1410–1416.