

ORIGINAL PAPER

Factor analysis in the Genetics of Asthma International Network family study identifies five major quantitative asthma phenotypes

S. G. Pillai^{*}, Y. Tang^{*,†}, E. van den Oord[‡], M. Klotzman^{*}, K. Barnes[§], K. Carlsen[¶], J. Gerritsen^{||}, W. Lenney^{**}, M. Silverman^{††}, P. Sly^{‡‡}, J. Sundy^{§§}, J. Tsanakas^{¶¶}, A. von Berg^{|||}, M. Whyte^{***}, H. G. Ortega^{†††}, W. H. Anderson^{*} and P. J. Helms^{†††}

^{*}Medical Genetics, GlaxoSmithKline, Research Triangle Park, NC, USA, [†]Department of Psychiatry, SUNY Downstate Medical Center, Brooklyn, NY, USA, [‡]Virginia Institute for Psychiatric and Behavioral Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA, USA, [§]Departments of Medicine & Epidemiology, Johns Hopkins University, Baltimore, MD, USA, [¶]Ullevaal University Hospital, Oslo, Norway, ^{||}University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, ^{**}Academic Department of Pediatrics, North Staffordshire Hospital, Stoke on Trent, UK, ^{††}Division of Child Health, University of Leicester, Leicester, UK, ^{‡‡}Center for Child Health Research, University of Western Australia, Perth, Australia, ^{§§}Duke University Medical Center, Durham, NC, USA, ^{¶¶}Pediatric Respiratory Unit, Hippokraton General Hospital, Thessaloniki, Greece, ^{|||}Abt. Fuer Kinderheilkunde Forschungsinstitut zur Praevention von Allergien und Atemwegserkrankungen im Kindesalter, Wesel, Germany, ^{***}Academic Unit of Respiratory Medicine, University of Sheffield, Sheffield, UK, ^{†††}Respiratory Medicine Development Center, Glaxo SmithKline, Research Triangle Park, NC, USA and ^{†††}Department of Child Health, University of Aberdeen Royal Aberdeen Children's Hospital, Aberdeen, UK

Clinical and Experimental Allergy

Summary

Background Asthma is a clinically heterogeneous disease caused by a complex interaction between genetic susceptibility and diverse environmental factors. In common with other complex diseases the lack of a standardized scheme to evaluate the phenotypic variability poses challenges in identifying the contribution of genes and environments to disease expression.

Objective To determine the minimum number of sets of features required to characterize subjects with asthma which will be useful in identifying important genetic and environmental contributors.

Methods Probands aged 7–35 years with physician diagnosed asthma and symptomatic siblings were identified in 1022 nuclear families from 11 centres in six countries forming the Genetics of Asthma International Network. Factor analysis was used to identify distinct phenotypes from questionnaire, clinical, and laboratory data, including baseline pulmonary function, allergen skin prick test (SPT).

Results Five distinct factors were identified: (1) baseline pulmonary function measures [forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC)], (2) specific allergen sensitization by SPT, (3) self-reported allergies, (4) symptoms characteristic of rhinitis and (5) symptoms characteristic of asthma. Replication in symptomatic siblings was consistent with shared genetic and/or environmental effects, and was robust across age groups, gender, and centres. Cronbach's α ranged from 0.719 to 0.983 suggesting acceptable internal scale consistencies. Derived scales were correlated with serum IgE, methacholine PC₂₀, age and asthma severity (interrupted sleep). IgE correlated with all three atopy-related factors, the strongest with the SPT factor whereas severity only correlated with baseline lung function, and with symptoms characteristic of rhinitis and of asthma.

Conclusion In children and adolescents with established asthma, five distinct sets of correlated patient characteristics appear to represent important aspects of the disease. Factor scores as quantitative traits may be better phenotypes in epidemiological and genetic analyses than those categories derived from the presence or absence of combinations of +ve SPTs and/or elevated IgE.

Keywords atopy, FEV₁, IgE, PC₂₀, rhinitis

Submitted 18 June 2007; **revised** 2 October 2007; **accepted** 9 November 2007

Correspondence:

Sreekumar G. Pillai, Medical Genetics,
5 Moore Drive, GlaxoSmithKline,
Research Triangle Park, NC 27709, USA.
E-mail: Sreekumar.g.pillai@gsk.com

Introduction

The phenotypic variability of asthma provides a challenge in the identification of major environmental and genetic contributors to disease initiation and expression. Asthma is defined as a chronic inflammatory disorder of the airways, in which many cells play a role, resulting in episodic coughing, wheezing, and shortness of breath that vary spontaneously and with treatment [1]. However, this broad definition encompasses a set of heterogeneous conditions that share clinical features, but may have different underlying causes. These subtypes range from the transient wheezing frequently seen in young children through moderate disease in children and adolescents, mainly associated with atopy [2], to severe persistent disease in adults that may or may not be associated with allergy [3, 4], and that merges with progressive and largely unresponsive chronic obstructive airways disease [5]. The identification of different asthma phenotypes, both at an individual and population level, is therefore critical in understanding causation prognosis and guiding therapy.

Factor analysis is a statistical tool that may be used to disentangle heterogeneous phenotypes such as seen in asthma. The statistical framework uses correlations between variables to identify a smaller set of latent or unmeasured factors to explain the interrelationship among a larger set of observed features. Subsets of variables that have relatively high correlations with each other (but low correlations with other subsets of variables) tend to load on the same factor. In other words, factor analysis reduces a large number of disease features to a smaller, more manageable number of independent and analyzable features or 'factors'. The underlying assumption is that any observed features that correlate with each other are likely to be associated with the same underlying disease process. The derived factors can then be used to construct measurement instruments that are more reliable and valid than each of the individual disease features used independently. Derived factor scores not only reduce the dimensions of the data, but can also help to refine phenotype definitions used in clinical trials and in epidemiological, and genetic research.

The challenges of defining asthma are well known. Currently, a 'top down' approach is used, in which a clinician or researcher determines what constitutes the various subtypes or phenotypes that define 'asthma'. In contrast, the empirical dimensional approach used in factor analysis assumes that disorders may not fall into clear-cut diagnostic categories, but rather, span a range of quantitative, variable phenotypes. This approach follows a 'bottom-up' strategy that allows the empirical data, rather than the disease expert, to determine how patients are classified. It also has the advantage of producing a score, reflecting the importance of the particular factor that can be used as a quantitative variable [6].

Factor analyses have been applied in asthma using a range of data in various combinations [7–20]. To date, most published reports are of relatively small size and there is a paucity of familial information. The latter point is of particular relevance to genetic studies because homogenous patient samples increase the likelihood of identifying subtle genetic effects. We therefore sought to determine the minimum number of sets of features required to characterize subjects with asthma in anticipation that these could be useful in identifying important genetic and environmental contributors. Herein, we report the results of such an analysis based on a large international asthma family collection (1022 families from 11 centres) recruited from Europe, Australia, and the United States, and in which the same methods of ascertainment and outcomes were used.

Materials and methods

Data from 1022 nuclear families recruited to the Genetics of Asthma International Network (GAIN) were available for analysis (Table 1). The ascertainment procedures and data acquisition have been described elsewhere [21]. In brief, families were identified through probands aged 7–35 years with physician diagnosed asthma, with at least one sibling who had symptoms of asthma for a minimum of two continuous years since the age of 7 years, but not necessarily currently, and with both biological parents available for study. A common protocol was used including respiratory questionnaires for children modified from the International Study of Asthma and Allergies in Childhood (ISAAC) and for adults from the ATS and European Community Respiratory Health Study (ECRHS) instruments that had been validated in several studies [22, 23] and with translation into the required language. Baseline spirometry [24], methacholine challenge using the cockcroft protocol [25], and skin prick test (SPT) to a common panel of seven aero allergens were performed (Table 2) with an additional local allergen (e.g. Birch in the Norwegian Center and Olive in the Greek Center). All subjects were instructed to omit antihistamines 72 h before testing and histamine dihydrochloride was used as the positive control with normal saline as negative control. All allergens with the exception of cockroach were supplied as SOLUPRICK[®] SQ by Alk Abelló, A/S (Bøge, Hørsholm, Denmark). The cockroach allergen was supplied by Greer Laboratories Inc. (Lic.308, Lenoir, NC, USA). Total serum IgE was measured using the UniCAP total IgE fluoroenzymeimmunoassay (Pharmacia Upjohn Diagnostics AB, Uppsala, Sweden) using the instrument Unicap 100.

Informed consents were obtained from the study participants and/or their parents before collecting these data. Study protocols were reviewed and approved by the appropriate Institutional Review Boards.

Table 1. Details of the family structure of the subjects recruited in the Genetics of Asthma International Network (GAIN)

Center	Families (n)	Probands (n) [1]	Siblings (n)
Aberdeen, UK	101	100	167
Barbados	100	98	113
North Carolina, US	64	51	111
Groningen, the Netherlands	75	60	142
Leicester, UK	87	82	119
Oslo, Norway	102	99	179
Perth, Australia	100	93	162
Sheffield, UK	99	96	149
Stoke-on-Trent, UK	91	87	127
Thessaloniki, Greece	101	102	112
Wesel, Germany	102	57	182
Total	1022	925	1563

A total of 1022 families were ascertained. In 97 out of 1022 families, either the proband designation was not available or one of the siblings did not meet the proband criteria.

Statistical analysis

All analyses were performed with SAS software (SAS institute, Cary, NC, USA). Variables that had more than 20% missing values or that had close to a uniform response across all probands and their siblings were excluded. Differences in quantitative traits between probands and siblings were evaluated with *t*-tests, and differences in qualitative traits were evaluated by χ^2 tests.

Factor analysis was conducted on the variables from the respiratory questionnaire, pulmonary function tests (PFT) and SPT, using SAS PROC FACTOR procedure. Probands were used in the first step and the number of factors to be retained determined by the scree plot, by Eigenvalue larger than one criterion, whether or not the derived scales had a satisfactory internal consistency, and the face validity of the solution. PROMAX (oblique) rotation was used to rotate the retained factors to improve interpretation. To assess the robustness of the derived factor structure across sub-samples, the analysis was re-run on the retained variables using one other affected sibling from each family, siblings from Caucasian families only, and sub-samples across age and gender. In all analyses, one member was randomly selected from each family since the correlation matrix could have resulted in bias if multiple sibs from each family were used. For each factor, a scale score was defined as the sum of the variables that loaded on that factor >0.45 . Variables that loaded >0.45 on more than one factor were not included in any scale. Internal consistency, the extent to which variables included in the scale measure the same underlying factor(s), was determined by calculating Cronbach's α [26]. To assess whether the scales captured meaningful but different aspects of asthma (external validity) correlations with demographic variables, with clinical characteristics and intra-class correlations were determined between siblings for each factor. Age, gender, PC₂₀ and total serum

IgE were not included in the factor analysis because we chose to keep them as 'stand-alone' variables to enable comparisons with other studies and to study their relationship with the derived scales. To remove the effects of covariates on the correlations between the studied variables, all procedures were repeated after adjustment for these covariates. The results from the raw measures are reported unless the results from adjusted and raw measures led to different conclusions. The SAS PROC MIXED procedure was then used to fit stepwise linear models with the centres, age group (or age, age²), sex, height group (or height, height²) as fixed effects and compound symmetry as the variance-covariance structure within each family. Only significant fixed effects were retained, and the residuals were used in the external validity procedure.

Results

Families, recruited from 11 centres, had an average of 2.5 children per family (Table 1). Proband designation was not available for 97 families, thus reducing the number to 925 and 1563 informative probands and siblings, respectively (Table 2). Compared with their siblings, probands had characteristics consistent with significantly more severe asthma such as a lower PC₂₀, lower baseline pulmonary function [forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC)], higher total serum IgE, greater SPT reactivity, and a higher proportion of self-reported symptoms (Table 2).

Included variables were reduced to five primary factor loadings (Table 3) comprising: (1) allergy assessed by 'SPT'; (2) baseline pulmonary function measurements of FEV₁ and FVC ('PFT'); (3) self-reported allergies ('SRA'); (4) rhinitis symptoms ('rhinitis'); and (5) respiratory symptoms ('symptoms').

Table 2. Demographic and clinical characteristics of the probands and siblings from the Genetics of Asthma International Network (GAIN)

	Proband	Siblings
<i>n</i>	925	1563
Male gender (%)	56.9	53.8
Age (years)	13.71 ± 4.93	13.84 ± 5.03
Age of onset (years)	4.62 ± 3.97	5.05 ± 4.08
Runny nose (%)**	60.5	51.3
Sneezing (%)**	64.6	52.8
Watery eyes (%)**	55.1	47.6
Blocked nose (%)**	63.1	52.0
Eczema (%)	45.8	42.9
Wheezing (%)**	96.4	79.8
Common cold induces wheeze (%)**	80.1	60.8
Shortness of breath (%)**	86.7	64.2
Triggers other than common cold (%)	89.5	69.1
Speech limited by wheeze (%)**	21.0	10.3
Sleep interrupted (%)**	61.0	38.5
Normal between wheeze episodes (%)	78.9	71.6
Exercise induced asthma (%)**	69.1	50.7
Allergy to animals (%)**	46.8	34.7
Allergy to birds (%)**	24.4	14.8
Allergy to dust (%)**	57.1	45.8
Allergy to food (%)	36.4	30.6
Allergy to other (%)**	21.3	14.2
Allergy to pollen (%)**	58.8	51.7
Allergy to detergent (%)**	33.4	23.5
IgE (IU)	570.69 ± 1069.02	443.52 ± 836.05
Alternaria SPT (mm)	0.50 ± 1.16	0.40 ± 1.07
Cat SPT (mm)**	2.58 ± 3.36	2.03 ± 3.00
Cockroach SPT (mm)	0.66 ± 1.35	0.62 ± 1.36
Dog SPT (mm)**	2.02 ± 2.54	1.47 ± 2.17
Dust far SPT (mm)	1.98 ± 2.31	1.68 ± 2.25
Dust Pt SPT (mm)	2.79 ± 2.79	2.41 ± 2.72
Grass spt (mm)	2.78 ± 3.13	2.54 ± 3.00
PC ₂₀ (mg/L)**	9.48 ± 12.88	12.40 ± 13.75
FEV ₁ (L)**	2.48 ± 0.90	2.65 ± 0.99
FVC (L)	3.01 ± 1.09	3.14 ± 1.20

Means ± SD for quantitative traits; percentage for qualitative traits.

***P*-value < 0.001.

SPT, skin prick tests; FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity.

SPT to alternaria and cockroach and positive responses to the following questionnaire items (a) 'in the last 12 months have your symptoms ever been severe enough to limit your speech to only one or two words at a time', (b) 'is your sleep interrupted by episodes of cough, wheezing or shortness of breath' (c) 'have you ever been told you had eczema by a physician' had loadings of <0.45. Positive responses to the question 'are you allergic to pollen' loaded equally high on factors 3 ('SPT') and 4 ('rhinitis'). Loadings of reactivity to cockroach and alternaria were inconsistent, as unlike other variables, their loading scores varied in sub-sample analyses between Caucasian families, age groups, and gender (data not shown). All the above variables were therefore dropped from the model in a stepwise manner. Cronbach's α

(internal consistency) for each of the five scales were 0.798 for 'SPT', 0.983 for 'PFT', 0.736 for 'SRA', 0.814 for 'rhinitis' and 0.719 for 'symptoms'.

The patterns of factor loadings in sub-sample analysis of siblings, including those from Caucasian families only and sub-samples across age and gender were consistent with the patterns seen in probands (data not shown). One exception was that the loadings for 'symptoms' in the sub-sample analyses were slightly higher in siblings. The factor structure for probands were not statistically different from that for sibling. A multi-group confirmatory factor analysis also suggested that the factor structures between proband and siblings are not statistically different (*P*-value > 0.05). Similarly, the factor structures were not statistically different between male and female groups and

Table 3. Five factor solution and factor loadings after oblique rotation in the proband data

Phenotype	SPT	PFT	SRA	Rhinitis	Symptoms
Dust far SPT	0.771				
Dust Pt SPT	0.768				
Dog SPT	0.720		0.269		
Cat SPT	0.708		0.241		
Grass SPT	0.638	0.125	0.160	0.119	
FEV ₁	0.106	0.980			
FVC	0.181	0.949	0.103		
Allergy to detergent			0.661		
Allergy to animals	0.402		0.646	0.167	
Allergy to birds	0.218		0.644		
Allergy to dust	0.214		0.598	0.183	
Allergy to other		0.118	0.590		
Allergy to food			0.568		
Runny nose				0.844	
Sneezing				0.840	
Blocked nose				0.757	0.100
Watery eyes	0.163		0.161	0.692	
Wheezing					0.850
Triggers other than common					
Cold induces wheeze			0.122	0.101	0.765
Shortness of breath	0.122				0.665
Exercise induced asthma			0.129	0.121	0.547
Common cold induces wheeze		− 0.166	0.147		0.518
Normal between wheeze episodes			− 0.175		0.464

SPT, skin prick tests; PFT, pulmonary function test; SRA, self-reported allergies; 'rhinitis', rhinitis symptoms; 'symptoms', respiratory symptoms; FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity.

Table 4. Correlation coefficients between scale scores

	SPT	PFT	SRA	Rhinitis	Symptoms
SPT		0.212**	0.332**	0.272**	0.168**
PFT			0.204**	0.102*	− 0.035
SRA				0.313**	0.178**
Rhinitis					0.153**

*P-value: 0.05–0.001.

**P-value <0.001.

SPT, skin prick tests; PFT, pulmonary function test; SRA, self-reported allergies; 'rhinitis', rhinitis symptoms; 'symptoms', respiratory symptoms.

between young and adult groups. The correlation coefficients between the five factors are given in Table 4 where it can be seen that factors representing different features of atopy including +ve SPTs, symptoms of rhinitis and self-reported allergies were moderately correlated (r 0.27–0.33). Correlations between 'PFT' and the atopy factors although statistically significant were weak and the 'symptoms' factor although significantly correlated with all three atopy factors was not significantly correlated with the 'PFT' factor.

The external validity of the five factors was assessed by examining correlations of the respective scale scores with key demographic and clinical characteristics (Table 5). In this context, a non significant correlation does not mean

that the criterion variable is not a risk factor for the presence or absence of asthma as all probands and 85% of the siblings had a physician confirmed diagnosis of asthma at the assessment visit. Factor scores for 'PFT' were significantly higher in males than in females, while scores for 'SPT', 'rhinitis' and 'symptoms' were significantly higher in females. Increasing scores of all factors, i.e. the strength of the association of their components, were significantly positively associated with increasing age with the exception of the 'symptoms' factor. PFT scores increased with age up to around 25 years and then declined. Total serum IgE was significantly positively correlated with scores for 'SPT', 'SRA', and 'rhinitis', but not with 'PFT' or 'symptoms'. Methacholine PC₂₀ was

Table 5. External validity: correlations between scale scores and clinical and demographic variables

Phenotype	SPT	PFT	SRA	Rhinitis	Symptoms
Age	0.207**	0.724**	0.317**	0.120**	0.044
Age of onset	-0.029	0.247**	-0.094*	-0.029	-0.041
Gender	-0.042	-0.076*	0.111*	0.074*	0.085*
Height	0.189**	0.900**	0.238**	0.102*	-0.017
BHR	0.299**	-0.066	0.084*	0.098*	0.145**
IgE	0.454**	-0.035	0.194**	0.169**	0.064
PC ₂₀	-0.315**	0.103*	-0.108*	-0.083*	-0.140**
Skin test positive	0.769**	0.184**	0.208**	0.270**	0.099*
Sleep interrupted	0.029	-0.138**	-0.003	0.079*	0.225**

BHR-PC₂₀ ≤ 8 mg/mL of methacholine.

Skin test positive: at least one skin test positive (>3 mm).

**P*-value: 0.05–0.001.

***P*-value <0.001.

SPT, skin prick test; PFT, pulmonary function test; SRA, self-reported allergies; 'rhinitis', rhinitis symptoms; 'symptoms', respiratory symptoms; BHR, bronchial hyperresponsiveness.

negatively correlated with all factor scores other than a significant positive correlation with the 'PFT' factor, an observation consistent with reduced baseline lung function being associated with increased BHR. The correlation matrix based on the data from the siblings of probands showed similar patterns (data not shown).

After adjusting for the effects of covariates between-siblings correlations for the factor scores were significant for all factors except for symptoms (SPT: $r=0.281$, $P<0.001$; PFT: $r=0.236$, $P<0.001$; SRA: $r=0.227$, $P<0.001$; rhinitis: $r=0.152$, $P<0.001$). Assuming shared environmental effects among siblings and that the genetic effects were additive, twice the full sibling correlation provides an estimate of the heritability of the identified factors [27]. Making these assumptions the heritability estimates were 56.2% for SPT, 47.2% for PFT, 45.4% for atopy:SR, 30.4% for rhinitis but only 5.6% for symptoms, estimates that were likely to underestimate the true heritability due to the small number of asymptomatic individuals in the sample.

Discussion

We identified five factors describing different components of asthma and associated features in children and young adults namely: atopy characterized by SPTs, atopy characterized by self-reported allergies, symptoms of rhinitis, lung function, and respiratory symptoms. These five factors provided a succinct summary of the information contained in a large number of individual variables. The internal consistencies, as measured by Cronbach's α , for the five scales were at acceptable levels and consistent findings across the subgroup analyses (stratified by age, gender and ethnicity) in both probands and their siblings, served to further validate the loading scores. An important finding was that clinical hallmarks of atopy including

questions on allergic status, symptoms of rhinitis and skin prick allergy tests, loaded on three distinct factors. This indicates that interpretation of allergen exposure and rhinitis symptoms differ from each other and from specific allergen sensitization. For example, self-reported allergy to animals had a cross-loading with SPT of only 0.4, while self-reported allergy to dust had an even lower cross-loading of <0.2 (Table 3) suggesting either that these items measure different aspects of features commonly associated with atopic asthma or that self-reported allergy is unreliable. The observations that cockroach and *Alternaria* cross loaded onto more than one factor and were inconsistent in population sub-groups were not entirely surprising in view of the significant variation in responsiveness to these specific allergens in family collections from different countries in the GAIN sample. None of the probands from the Leicester (UK) families were sensitized to cockroach whereas 24% of the probands from the North Carolina (US) families were. For *Alternaria* a similar range was noted with no probands positive for this allergen in the Sheffield (UK) compared with 16%, in North Carolina (US) and 18% in Thessaloniki (Greece). This is not to say that individual allergens are not important but rather that local condition and exposures need to be taken into account when generalizing results from one country or population to another. However our international sample demonstrates that features that are held in common between populations can be identified, and hence, could be used without prejudice in the identification of common genetic and environmental contributors to disease expression. The correlation coefficient between SPT and SRA had the highest magnitude (0.33) of any cross-factor comparisons whereas symptoms characteristic of rhinitis only loaded on one factor supporting the conclusion that subjects with asthma and with rhinitis may be a discrete subset, a conclusion that supports the associations of

perennial rhinitis and asthma in non-atopic adults seen in the ECRHS survey [28].

The correlations between the factors and with key demographic variables were not unexpected. As anticipated, the lung function measures had a quadratic association with age and were significantly higher in males than females, both relationships that are well established in the literature [29–32]. The skin test responses, self-reported allergies and rhinitis symptoms showed positive associations with age, which were found to be non-linear and were mainly due to higher values in probands and siblings above 20 years of age.

Atopic asthma has generally been defined using clinical history, symptoms, IgE (total and specific), and/or responses to allergen SPTs. Each of these measurements have inherent limitations, for example it has been shown that a weal size of up to 5.5 mm may be necessary to obtain a 99% specificity, while 2 or 3 mm above negative control is more frequently used to define a positive test [33]. Previous reports in families have shown that total IgE was the least important in determining severity of atopy [34], that subjects who report clinical symptoms of asthma can have normal IgE [35], and that many adult asthmatic subjects are non-atopic [3, 4]. The present study suggests that the use of factor scores as quantitative traits would be better phenotypes in epidemiological and genetic analyses than definitions of atopy based on one of, or a combination of +ve SPTs and/or elevated IgE. Serum IgE has been suggested as a valid intermediate phenotype in the search for genetic candidates relevant to asthma, particularly in view of its quantitative nature [36, 37]. However, the correlations between serum total IgE and lung function and respiratory symptom factors were not significant in our study indicating that although IgE may reflect susceptibility to symptoms, other than those associated directly with asthma, it may not be a valid 'intermediate' phenotype for asthma. Although atopy is frequently associated with asthma, particularly in children, the overlapping and separate biological mechanisms in asthma and atopy remain to be identified. The association between total serum IgE levels and specific IgE measurements is also debated. It is generally considered that the total IgE and specific IgE are distinct phenotypes. PC₂₀ to methacholine was found to be correlated with all five factors with the strongest correlation to 'SPT' and 'symptoms'. Dichotomizing PC₂₀ into two categories, bronchial hyper-reactivity (PC₂₀ ≤ 8 mg) and no bronchial hyperresponsiveness (BHR) (PC₂₀ > 8 mg), preserved the correlation structure, but using PC₂₀ as a quantitative trait was much more powerful. BHR has been shown to be associated with atopic status [38] and total serum IgE, specific IgE, baseline airway caliber, and asthma symptoms are the main independent factors influencing BHR [39]. However, not all atopic individuals have BHR and not all those with BHR are allergic [40].

An important goal of the factor analysis is to reduce the large number of disease symptoms to a smaller set of reliable measures that can be used in subsequent clinical, epidemiological and genetic research. PC₂₀ and IgE were initially not included in the factor analyses, because they are standard variables in asthma research. We want to keep analysing these variables separately in subsequent research, to help facilitate comparisons across studies. However, it can be argued that when we do factor analyses we need to incorporate all possible variables to get the best possible solution. We therefore conducted factor analyses again after including PC₂₀ and IgE. The results (not shown) indicated that the original factor structure remained the same as before. IgE and PC₂₀ loaded in to the SPT factor. The external validity analyses reported on Table 5 highlights this relationship.

Factor analysis helps to reduce the variable dimensions in complex diseases by using composite variables in which a number of different but related symptoms and signs can be combined through use of the correlations in the empirical data. This result in fewer, less correlated dimensions that may prove to be more useful in subsequent studies and point to different mechanisms contributing to the overall asthma phenotype. By using the factor scores as quantitative phenotypes, the probability of identifying susceptibility genes representing these factors is likely to be increased as indeed shown in linkage analyses of asthma [6], diabetes [41, 42] and the metabolic syndrome [41].

This study has several limitations with its cross sectional nature arguably the most relevant. Longitudinal population based studies are required in order to determine at which period of life the features defined by different factors become relevant. Another potential problem was that some variables with substantial cross loading had to be eliminated from the analysis and that rather than being unimportant, could be indicators of relevant subtypes that are indistinguishable in factor analyses. Eliminated variables could reflect causal factors leading to symptoms through a latent factor with their effects constrained to cause a similar clustering of the items. The substantive interpretation of this constraint is that 'phenotypic' factor analyses essentially assume that different types of causal factors affect the disease via similar pathways. Cross-loadings may also be the result of averaging of pathways or represent aetiological pathways that have smaller effects than can be detected in phenotypic factor analyses. However the asthma sample used in the present study is part of a larger initiative aimed at the identification of susceptibility genes that will enable target selection in drug discovery [43] and by including measured genotypes in factor models [44], may provide further opportunities to refine the factor scales. The samples used in this study are from 11 clinical centres and arguably there is considerable phenotypic

heterogeneity. In order to identify the centre effect, we conducted factor analyses within each centre. The factor structure was found to be very similar in most of the centres (data not shown). The PFT factor was consistent in all the centres while cross-loadings noticed in several of the other variables, but the solution was very similar to the analysis of the full data set. A factor analyses using a random sample of subjects from the general population (aged 20–44 years), from 35 centres in 15 countries from the European Community Respiratory Health Study (ECRHS) addressed this question. In the confirmatory factor analysis of a structure specifying not only the same form but also the factor loadings and the factor covariances, all countries showed an adequate fit, except for one country [19]. Our exploratory analysis also shows similar results though the sample size per centre is not high enough to make meaningful conclusions.

In conclusion we have identified five factors in children adolescents and young adults with physician diagnosed asthma, which reflect important objective and subjectively reported features of the disease. Factors that expressed as quantitative traits may be better phenotypes in epidemiological and genetic exploration of asthma causation and susceptibility rather than definitions based on one of, or combination of features such as +ve SPT's elevated IgE or BHR.

Acknowledgement

K. C. B. was supported in part by the Mary Beryl Patch Turnbull Scholar Programme.

References

- Global Initiative for Asthma. *Global strategy for asthma treatment and prevention*. Bethesda, MD: NIH Publication 02-3659, 2004.
- Martinez FD, Helms PJ. Types of asthma and wheezing. *Eur Resp J* 1998; 12:S3–8.
- Douwes J, Gibson P, Pekkanen J, Pearce N. Non-eosinophilic asthma: importance and possible mechanisms. *Thorax* 2002; 57:643–8.
- Court CS, Cook DG, Strachan DP. Comparative epidemiology of atopic and non-atopic wheeze and diagnosed asthma in a national sample of English adults. *Thorax* 2002; 57:951–7.
- Anonymous. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *Am Rev Respir Dis* 1986; 136:225–44.
- Holberg CJ, Halonen M, Solomon S *et al*. Factor analysis of asthma and atopy traits shows 2 major components, one of which is linked to markers on chromosome 5q. *J Allergy Clin Immunol* 2001; 108:772–80.
- Allen RM, Jones MP. The validity and reliability of an asthma knowledge questionnaire used in the evaluation of a group asthma education self-management program for adults with asthma. *J Asthma* 1998; 35:537–45.
- Baiardini I, Pasquali M, Giardini A *et al*. Rhinasthma: a new specific QoL questionnaire for patients with rhinitis and asthma. *Allergy* 2003; 58:289–94.
- Gronke L, Kannies F, Holz O, Jorres RA, Magnussen H. The relationship between airway hyper-responsiveness, markers of inflammation and lung function depends on the duration of the asthmatic disease. *Clin Exp Allergy* 2002; 32:57–63.
- Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol* 1997; 50:233–8.
- Juniper EF, Wisniewski ME, Cox FM, Emmett AH, Nielsen KE, O'Byrne PM. Relationship between quality of life and clinical status in asthma: a factor analysis. *Eur Respir J* 2004; 23: 287–91.
- Powell CV, McNamara P, Solis A, Shaw NJ. A parent completed questionnaire to describe the patterns of wheezing and other respiratory symptoms in infants and preschool children. *Arch Dis Child* 2002; 87:376–9.
- Terada M, Ishioka S, Hozawa S, Yasumatsu Y, Nakamura K, Yamakido M. A statistical investigation of the influence of allergic factors on intractable asthma by multiple factor analysis. [Japanese]. *Arerugi – Jpn J Allergol* 1991; 40:1289–96.
- Rosier MJ, Bishop J, Nolan T, Robertson CF, Carlin JB, Phelan PD. Measurement of functional severity of asthma in children. *Am J Respir Crit Care Med* 1994; 149:1434–41.
- Bailey WC, Higgins DM, Richards BM, Richards JM Jr. Asthma severity: a factor analytic investigation. *Am J Med* 1992; 93:263–9.
- Gronke L, Kannies F, Holz O, Jorres RA, Magnussen H. The relationship between airway hyper-responsiveness, markers of inflammation and lung function depends on the duration of the asthmatic disease. *Clin Exp Allergy* 2002; 32:57–63.
- Grazzini M, Scano G, Foglio K *et al*. Relevance of dyspnoea and respiratory function measurements in monitoring of asthma: a factor analysis. *Respir Med* 2001; 95:246–50.
- Rodrigo G, Rodrigo C. Assessment of the patient with acute asthma in the emergency department. A factor analytic study. *Chest* 1993; 104:1325–8.
- Sunyer J, Basagana X, Burney P, Anto JM. International assessment of the internal consistency of respiratory symptoms. European Community Respiratory Health Study (ECRHS). *Am J Respir Crit Care Med* 2000; 162:930–5.
- Rosi E, Ronchi MC, Grazzini M, Duranti R, Scano G. Sputum analysis, bronchial hyperresponsiveness, and airway function in asthma: results of a factor analysis. *Allergy Clin Immunol* 1999; 103:232–7.
- Pillai SG, Chiano MN, White NJ *et al*. A genome-wide search for linkage to asthma phenotypes in the genetics of asthma international network families: evidence for a major susceptibility locus on chromosome 2p. *Eur J Hum Genet* 2006; 14:307–16.
- Palta M, Sadek-Badawi M, Sheehy M *et al*. Respiratory symptoms at age 8 years in a cohort of very low birth weight children. *Am J Epidemiol* 2001; 154:521–9.
- Kauffmann F, Dizier MH, Pin I *et al*. Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. *Am J Respir Crit Care Med* 1997; 156:S123–9.
- Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med* 1995; 152:1107–36.

- 25 Cockcroft DW, Killian DN, Mellon JJ, Hargreave FE. Bronchial reactivity to inhaled histamine: a method and clinical survey. *Clin Allergy* 1977; 7:235–43.
- 26 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 2004; 16:297–333.
- 27 Neale MC, Cardon LR. *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Press, 1992.
- 28 Leynaert B, Bousquet J, Neukirch C, Liard R, Neukirch F. Perennial rhinitis: an independent risk factor for asthma in nonatopic subjects: results from the European Community Respiratory Health Survey. *J Allergy Clin Immunol* 1999; 104:301–4.
- 29 Gustafsson PM, Kjellman B. Asthma from childhood to adulthood: course and outcome of lung function. *Respir Med* 2000; 94:466–74.
- 30 McDonnell WF, Enright PL, Abbey DE *et al*. Spirometric reference equations for older adults. *Respir Med* 1998; 92:914–21.
- 31 Hopper JL, Hibbert ME, Macaskill GT, Phelan PD, Landau LI. Longitudinal analysis of lung function growth in healthy children and adolescents. *J Appl Physiol* 1991; 70:770–7.
- 32 Sherrill D, Holberg CJ, Lebowitz MD. Differential rates of lung growth as measured longitudinally by pulmonary function in children and adolescents. *Pediatr Pulmonol* 1990; 8:145–54.
- 33 Nelson HS, Rosloniec DM, McCall LI, Ikle D. Comparative performance of 5 commercial skin prick tests. *J Allergy Clin Immunol* 1993; 92:750–6.
- 34 Rorke S, Barton SJ, Clough JB, Halloway JW, Keith T, Van Eerdewegh P. Development of asthma and atopy severity scores in an asthma-enriched population. *Am J Respir Crit Care Med* 2002; 165:B44.
- 35 Rorke S, Holgate ST. The atopy phenotype revisited. *Rev Française D'Allergologie* 2004; 44:436–44.
- 36 Mansur AH, Bishop DT, Markham AF, Morton NE, Holgate ST, Morrison JF. Suggestive evidence for genetic linkage between IgE phenotypes and chromosome 14q markers. *Am J Respir Crit Care Med* 1999; 159:1796–802.
- 37 Zhang Y, Leaves NI, Anderson GG *et al*. Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat Genet* 2003; 34:181–6.
- 38 Schwartz J, Schindler C, Zemp E *et al*. Predictors of methacholine responsiveness in a general population. *Chest* 2002; 122:812–20.
- 39 Anonymous. Determinants of bronchial responsiveness in the European Community Respiratory Health Survey in Italy: evidence of an independent role of atopy, total serum IgE levels, and asthma symptoms. *Allergy* 1998; 53:673–81.
- 40 Woolcock AJ, Peat J. What is the relationship between airway hyperresponsiveness and atopy? *Am J Respir Crit Care Med* 2000; 161:S215–7.
- 41 Tang WH, Miller MB, Rich SS *et al*. Linkage analysis of a composite factor for the multiple metabolic syndrome – The National Heart, Lung, and Blood Institute Family Heart Study. *Diabetes* 2003; 52:2840–7.
- 42 Austin MA, Edwards KL, McNeely MJ *et al*. Heritability of multivariate factors of the metabolic syndrome in nondiabetic Japanese americans. *Diabetes* 2004; 53:1166–9.
- 43 Roses AD, Burns DK, Chisoe S, Middleton L, St Jean P. Disease-specific target selection: a critical first step down the right road. *Drug Discov Today* 2005; 10:177–89.
- 44 Van den Oord EJCG, Snieder H. Including measured genotypes in statistical models to study the interplay of multiple factors affecting complex traits. *Behav Genet* 2002; 32:1–22.