Beyond Atopy
Multiple Patterns of Sensitization in Relation to Asthma in a Birth Cohort Study

Angela Simpson[1]*, Vincent Y. F. Tan[2]*, John Winn[3], Markus Svensén[3], Christopher M. Bishop[3], David E. Heckerman[4], Iain Buchan[5], and Adnan Custovic[1]

[1]The University of Manchester, Manchester Academic Health Science Centre, NIHR Translational Research Facility in Respiratory Medicine, University Hospital of South Manchester NHS Foundation Trust, Manchester, United Kingdom; [2]Stochastic Systems Group, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts; [3]Microsoft Research Cambridge, Cambridge, United Kingdom; [4]eScience Research Group, Microsoft Research, Redmond, Washington; [5]The University of Manchester, Northwest Institute for Bio-Health Informatics, Manchester, United Kingdom

***Rationale*: The pattern of IgE response (over time or to specific allergens) may reflect different atopic vulnerabilities which are related to the presence of asthma in a fundamentally different way from current definition of atopy.**

***Objectives*: To redefine the atopic phenotype by identifying latent structure within a complex dataset, taking into account the timing and type of sensitization to specific allergens, and relating these novel phenotypes to asthma.**

***Methods*: In a population-based birth cohort in which multiple skin and IgE tests have been taken throughout childhood, we used a machine learning approach to cluster children into multiple atopic classes in an unsupervised way. We then investigated the relation between these classes and asthma (symptoms, hospitalizations, lung function and airway reactivity).**

***Measurements and Main Results*: A five-class model indicated a complex latent structure, in which children with atopic vulnerability were clustered into four distinct classes (*Multiple Early* [112/1053, 10.6%]; *Multiple Late* [171/1053, 16.2%]; *Dust Mite* [47/1053, 4.5%]; and *Non-dust Mite* [100/1053, 9.5%]), with a fifth class describing children with *No Latent Vulnerability* (623/1053, 59.2%). The association with asthma was considerably stronger for Multiple Early compared with other classes and conventionally defined atopy (odds ratio [95% CI]: 29.3 [11.1–77.2] versus 12.4 [4.8–32.2] versus 11.6 [4.8–27.9] for Multiple Early class versus Ever Atopic versus Atopic age 8). Lung function and airway reactivity were significantly poorer among children in Multiple Early class. Cox regression demonstrated a highly significant increase in risk of hospital admissions for wheeze/asthma after age 3 yr only among children in the Multiple Early class (HR 9.2 [3.5–24.0], $P < 0.001$).**

***Conclusions*: IgE antibody responses do not reflect a single phenotype of atopy, but several different atopic vulnerabilities which differ in their relation with asthma presence and severity.**

Clinical trial registered with www.controlled-trials.com (ISRCTN72673620).

Keywords: asthma; atopy; unsupervised clustering; Bayesian inference; machine learning in epidemiology

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

In epidemiologic studies and clinical practice, children are classified as atopic if they have a positive IgE or skin prick test.

What This Study Adds to the Field

By adopting a machine learning approach, we have identified that IgE antibody responses do not reflect a single phenotype of atopy, but rather multiple different atopic vulnerabilities. We have demonstrated that only one of these atopic classes (multiple early atopic vulnerability) predicts asthma.

The term "atopy" describes the tendency to become IgE-sensitized to common allergens to which most people are exposed but do not have a prolonged IgE antibody response (1, 2). In most literature, atopic sensitization is defined as a positive allergen-specific serum IgE (sIgE) test or skin prick test (SPT) to any common food or inhalant allergens, and atopic sensitization thus defined remains the single strongest risk factor for asthma in the western world (3–5). Although evidence from twin and family studies suggests a strong genetic component of atopy (6), more than a decade of intensive work has failed to identify causal associations with genetic variants that are consistently replicated (7). Similarly, the increase in prevalence of atopy since the 1960s suggests an important environmental component, but no environmental exposure has consistently been associated with the development of atopy (8). We propose that one reason for this is phenotypic heterogeneity, because the diagnostic label of "atopy" may encompass many different phenotypes with different etiologies, not all of which are associated with symptomatic disease. The conventional epidemiologic approach does not reflect the complexities of disease; consequently, reproducible genetic and environmental studies remain elusive.

We speculate that the presence of positive allergy test (either sIgE or SPT) does not equate to the atopic phenotype associated with symptomatic allergic disease. We hypothesize that more useful information may be obtained by identifying common underlying statistical clusters that are characterized by IgE responses. Several recent publications have demonstrated the usefulness of a clustering approach in multidimensional data to identify different asthma phenotypes (9–12). Results of latent class analysis on a large dataset collected annually over a 7-year period identified six childhood wheezing phenotypes, two of which had not been described previously (10). Unsupervised hierarchical cluster analysis identified five distinct clinical

(*Received in original form July 21, 2009; accepted in final form February 17, 2010*)

* Joint first authorship.

Supported by Asthma UK Grant 04/014, Moulton Charitable Trust, the James Trust, and Microsoft Research.

Correspondence and requests for reprints should be addressed to Angela Simpson, M.D., University of Manchester, ERC Building, Second Floor, Wythenshawe Hospital, Manchester M23 9LT, UK. E-mail: angela.simpson@manchester.ac.uk

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 181. pp 1200–1206, 2010
Originally Published in Press as DOI: 10.1164/rccm.200907-1101OC on February 18, 2010
Internet address: www.atsjournals.org

phenotypes of adult asthma, emphasizing the need for new approaches for classification of disease phenotypes (11). We conducted a principle component analysis using answers to multiple questions relating to wheeze to identify five syndromes of coexisting symptoms that are likely to reflect different underlying pathophysiologic processes (12).

Ideally, one should aim to model all available data (i.e., multiple measurements at multiple time points) to identify latent variables that best describe the structure of the data. Such models need to be tailored to individual datasets, to encode prior knowledge precisely and to scale up to large volumes of data. A machine learning approach using Bayesian inference for unsupervised learning of latent variables to identify structure within the data is used commonly by computer scientists for problems in many other fields, and is ideally suited to this task. We applied this approach to a large complex data-set from a population-based birth cohort in which measures of allergic sensitization (both sIgE and SPT) to multiple inhalant and food allergens have been taken throughout childhood, to assign children to atopic latent classes in an unsupervised way, thus avoiding constraints placed by prespecified ideas of the nature and number of such classes. We sought to investigate whether these different latent atopic classes were related to the presence or absence of asthma in ways that are fundamentally different from current diagnostic categories.

METHODS

Study Design, Setting, and Participants

Manchester Asthma and Allergy Study is a population-based birth cohort (*see* online supplement) (13–16). Participants were recruited prenatally and followed prospectively, attending review clinics at ages 1, 3, 5, and 8 years. The study is registered as ISRCTN72673620 and approved by the Local Research Ethics Committee (04/Q1403/45). Written informed consent was obtained from all parents, and children gave their assent.

Definition of Variables

Atopic sensitization. We ascertained atopic sensitization by SPT (Hollister-Stier, Spokane, WA) and measurement of sIgE (ImmunoCAP; Phadia, Uppsala, Sweden) at each time point to a panel of inhalant and food allergens (*see* Table E2 in the online supplement). We defined allergen-specific sensitization as mean wheal diameter at least 3 mm greater than the negative control or specific IgE: 0.35 kU/l. The conventional definition considered a child to be atopic if he or she had allergen-specific sensitization to at least one allergen. Children with any positive test (SPT or sIgE) at any time point were considered to be "atopic ever."

Wheeze. A validated questionnaire was interviewer-administered to collect information on parentally reported symptoms, physician-diagnosed illnesses, and treatments received. Current wheeze was defined as wheeze in the past 12 months.

Based on prospectively collected data, children were assigned to the following wheeze phenotypes: no wheezing—no wheezing ever at any follow-up by age 8 years; transient early wheezing—wheezing only during the first 3 years; late-onset wheezing—wheezing started after age 3 years; persistent wheezing—wheezing during the first 3 years or wheezing in the previous 12 months at ages 5 and 8 years; and intermittent wheezing—wheezing at one time point during the first 5 years or wheezing at age 8 years.

Lung function. We measured specific airway conductance (sGaw) using whole-body plethysmography at ages 3 and 5 years (15, 17) and FEV_1 using spirometry at age 8 years (*see* online supplement).

Airway hyperreactivity (methacholine challenge). Airway hyperreactivity (AHR) was assessed at age 8 years in a five-step protocol using quadrupling doses of methacholine (*see* Table E1) according to American Thoracic Society guidelines (18). A dose–response ratio (DRR) was calculated and transformed as previously described (19).

Asthma. We used a stringent epidemiologic definition of asthma at age 8 years as symptomatic AHR (i.e., presence of current wheeze and positive methacholine challenge) (20).

Hospital admission for asthma and wheeze. A trained physician reviewed the written and computerized primary care medical records and extracted the data on hospitalizations for wheeze or asthma (21).

Data Analysis

We took a machine learning approach to the data analysis. Using a hidden Markov model (22), all available SPTs and sIgEs (collected at review clinics at ages 1, 3, 5, and 8 years) were used to infer one multinomial latent variable per child so as to cluster the children in an unsupervised manner into different sensitization classes (Figure 1). At the core of the model are the four dichotomous latent acquired sensitization variables for each allergen, which are linked together in a Markov chain across the four time points. We inferred time-dependent transition probabilities (i.e., the probabilities of gaining and losing sensitization at each age), which were assumed to be shared by all children in each sensitization class, but were allowed to differ between classes.

Inference. Inference was performed using Infer.NET (http://research.microsoft.com/infernet), a Microsoft-owned library for large-scale Bayesian inference, which is now freely available for research purposes. We used Infer.NET to infer the false- and true-positive rates of the SPTs and sIgEs; missing values; the class-specific state-transition probabilities; the observation (emission) probabilities; the acquired sensitization variables; and the sensitization class for each child. An approximate Bayesian inference method (variational message passing) (23) was used to perform the inference in an efficient manner.

Robustness and reproducibility. Robust and reproducible clustering was achieved by training the sensitization hidden Markov model multiple times on different subsets of children and selecting the clustering that gave good predictions on the remaining children and that was robust to the subset of children selected. Reproducibility was confirmed by computing confusion matrices between different replications of the clustering process (*see* online supplement).

Handling the missing data. Variables corresponding to missing data values were included in the model but treated as unobserved. Distributions over these missing data values were computed using variational message passing based on the available measurements.

Sensitization class. This is a multinomial variable indicating to which sensitization class each child belongs (out of between two and five classes). The model assumes that each child belongs to one of these classes. We investigated a two-class and a five-class model (*see* online supplement). No assumptions were made about the nature of each class. During inference, a distribution was computed for each child giving the probability of their belonging in each class. For further analysis, we assumed the child belonged to the highest probability class.

We then investigated the association between the classes we had inferred in a completely unsupervised manner and the clinical outcomes using appropriate statistical methods (chi-square test, logistic regression, Kaplan-Meier univariate estimates, and Cox regression multivariate estimates of survival and clinical status). Results are presented as the main effect with 95% confidence intervals (CI).

RESULTS

Of the 1,186 participants with any evaluable data, 133 who were randomized into the primary prevention study (24, 25) were excluded from the analysis of the association between clinical outcomes and inferred sensitization class. All remaining children with available clinical outcomes were included at each time point (*see* Table E2). There was no difference in parental history of allergic disease between children with or without missing data on clinical outcomes (data available on request).

At age 8 years, 18% (163 of 905) of children had current wheeze; 13.7% (124 of 905) were persistent wheezers; and 8.1% (45 of 555) had asthma (symptomatic AHR). Data collected

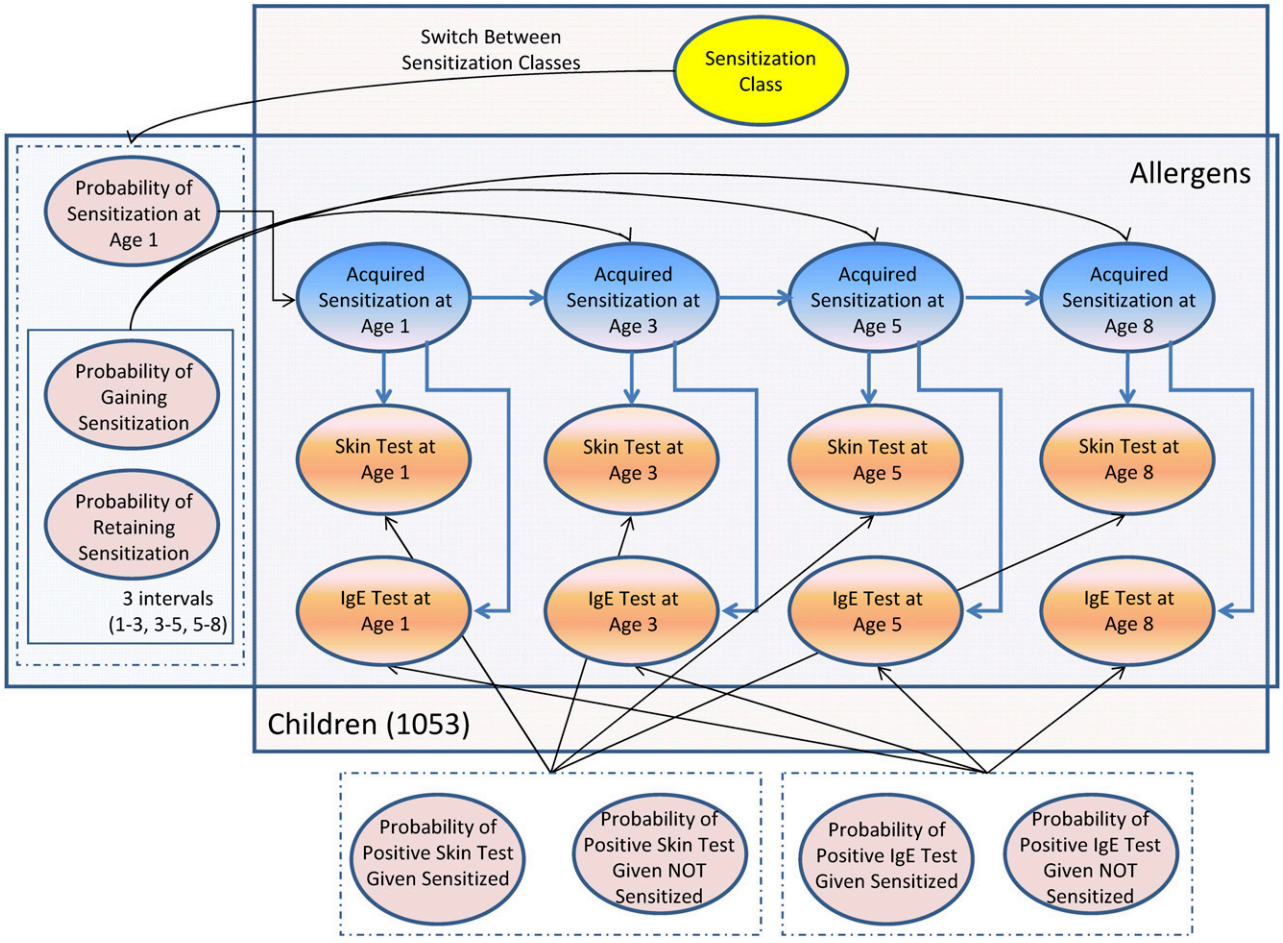

Figure 1. Graphic representation of a hidden Markov model. All available skin prick tests and serum IgE were used to infer one multinomial latent variable per child to cluster the children in an unsupervised manner into different sensitization classes.

from primary care records revealed that 16.7% (136 of 814) of children had been admitted to the hospital with wheezing or asthma on at least one occasion during the first 8 years of life. Using conventional definitions, of 827 children who had either SPT or sIgE measured at age 8 years, 322 (38.9%) were considered atopic; 1,029 children had at least one assessment of atopic status throughout the duration of the follow-up, of whom 441 (42.9%) were considered to be atopic ever.

Sensitization Class

The structure of the classes was inferred in a completely unsupervised manner using all data (SPT and sIgE) from all four time points with missing data inferred using variational message passing (23) (i.e., we did not assume beforehand how the children will be clustered, the unsupervised learning algorithm automatically discovered the latent structure) under the assumption that data were missing at random.

We present the results with the sensitization state being considered to have two classes (best reflecting a conventional assignment to atopy or no atopy) and five classes (which better captured the underlying structure of the data).

Two-class model. The children were assigned as having either a latent atopic vulnerability (280 [26.6%] of 1053) or no latent atopic vulnerability (773 [73.4%] of 1,053) (*see* Figure E1); 161 (36.6%) of 440 children who were sensitized on at least one occasion were classified as not vulnerable. Compared with conventional definitions, there was complete agreement in 86% (atopy age 8 years) and 84% of cases (atopy ever).

Five-class model. This model indicated a more complex latent structure incorporating time-varying probabilities of the gain and the loss of sensitization (*see* Figure E2). The children with latent atopic vulnerability were clustered into four distinct sensitization classes, which were assigned as the following based on our interpretation of the characteristics of each class: (*1*) non–dust mite atopic vulnerability (100 [9.5%] of 1053); (*2*) dust mite atopic vulnerability (47 [4.5%] of 1053); (*3*) multiple late atopic vulnerability (171 [16.2%] of 1053); and (*4*) multiple early atopic vulnerability (112 [10.6%] of 1053).

The final class comprised children with no latent vulnerability (623 [59.2%] of 1053). In this model, 61 (13.9%) of 440 children who were atopic ever were classified as having no latent vulnerability; among 322 children who were atopic at age 8 years, 36 (11.2%) were classified as having no latent vulnerability. All but one child in the multiple early class were atopic at age 8 years using conventional definition, but the multiple early class comprised only 28% of those atopic at age 8 years (*see* Table E3).

To determine the appropriate number of classes, differing numbers of clusters were tested as to their ability to predict the sensitization state of children where that state was artificially made missing. This imputation process suggests that between three and five clusters are justified (*see* Figure E3), and so a five-class model was selected because it exposed the most informa-

tion about the structure of the dataset. The choice of five classes was also validated when considering the confusion matrices found when the clustering process was replicated (*see* Tables E4 and E5). For the five-class case, there was very little confusion between different clusterings, indicating that the five-class clustering is robust. For example, for the multiple early class 111 of the 112 children assigned to this class in the reference clustering were repeatedly assigned to the same class in other five-class clusterings.

Sensitization Class and Clinical Outcomes

We ascertained relationships among atopy defined conventionally (atopic ever, atopic at age 8 years), the novel latent classes (two-class and five-class models), and clinical phenotypes associated with asthma (current wheeze, persistent wheeze, symptomatic AHR, and hospital admission with asthma or wheeze), adjusting for gender. The results are presented in Figures 2 and E4 and Table E6. The relationships with clinical outcomes for ever atopic, atopic at age 8 years, and the two-class model were not materially different. However, for the five-class model, it was apparent that there were marked differences between the four classes of atopic vulnerability, in that the associations with clinical outcomes were considerably stronger for multiple early compared with other classes, the two-class model, and conventionally defined atopy (e.g., for symptomatic AHR, odds ratio 29.3 [95% CI, 11.1–77.2] versus 12.4 [95% CI, 4.8–32.2] versus 11.6 [95% CI, 4.8–27.9] versus 9.2 [95% CI, 4.5–18.9] for multiple early class versus ever atopic versus atopic age 8 versus latent atopic vulnerability two-class model) (*see* Table E6). There was a very strong association between multiple early class and persistent wheezing (odds ratio 12.9; 95% CI, 6.8–24.4).

These finding indicated that IgE antibody responses do not reflect a single phenotype of atopy, but rather several atopic vulnerabilities that differ in their relationship with asthma. To further test this we investigated the relationship between markers of asthma severity (objective measures of lung function and airway reactivity and hospital admissions) within the five-class model.

Lung Function, Airway Reactivity, and Hospital Admissions in the Five-Class Model

In the univariate analysis we found a significant association between sGaw at ages 3 and 5 years, FEV_1, FEV_1/FVC ratio, and DRR at age 8 years and five-class latent variable (*see* Table E7). Multiple comparison test (Tukey) revealed that for all measures of lung function and airway reactivity, lung function was significantly poorer among children in the multiple early class compared with those with no latent atopic vulnerability, with little differences between the other three classes and the no latent vulnerability class (*see* Table E7 and Figures E5–E9).

In the multiple analysis of variance models adjusted for gender, maternal smoking, and wheezing (sGaw, FEV_1/FVC ratio, and DRR), and gender, wheezing, maternal smoking, and height (FEV_1), children in the multiple early class had significantly poorer lung function compared with those in the no latent vulnerability class (sGaw age 3 years, $P = 0.02$; sGaw age 5 years, $P = 0.01$; age 8 years FEV_1, FEV_1/FVC ratio, and DRR, $P < 0.001$) (Table 1). There were no significant differences in lung function between the other three classes and the no latent vulnerability class, apart from airway reactivity (DRR) being significantly higher in the multiple late class ($P = 0.05$) (Table 1).

Kaplan-Meier plots demonstrating the age of the first hospital admission with wheeze or asthma in relation to the five-class model are presented in Figure 3A. The results of a Cox regression that included the five classes, gender, and maternal smoking indicated a highly significant association between the risk of hospital admission and five-class model ($P < 0.0001$), with a risk of hospital admission increasing among children in the multiple early class (hazard ratio [HR] 5.1; 95% CI, 2.8–9.3; $P < 0.001$), dust mite class (HR 3.4; 95% CI, 1.4–8.2.7; $P = 0.004$) and non–dust mite (HR 2.5; 95% CI, 1.2–5.3; $P = 0.02$), but not those in the multiple late class (HR 1.3; 95% CI, 0.6–2.9; $P = 0.4$). To remove the effect of hospital admission for wheeze caused only by early life virus infections, we have reanalyzed the data on the time to the first hospital admission with wheeze or asthma among children who had a hospital admission after age 3 years (Kaplan-Meier plot, Figure 3B). Cox regression demonstrated

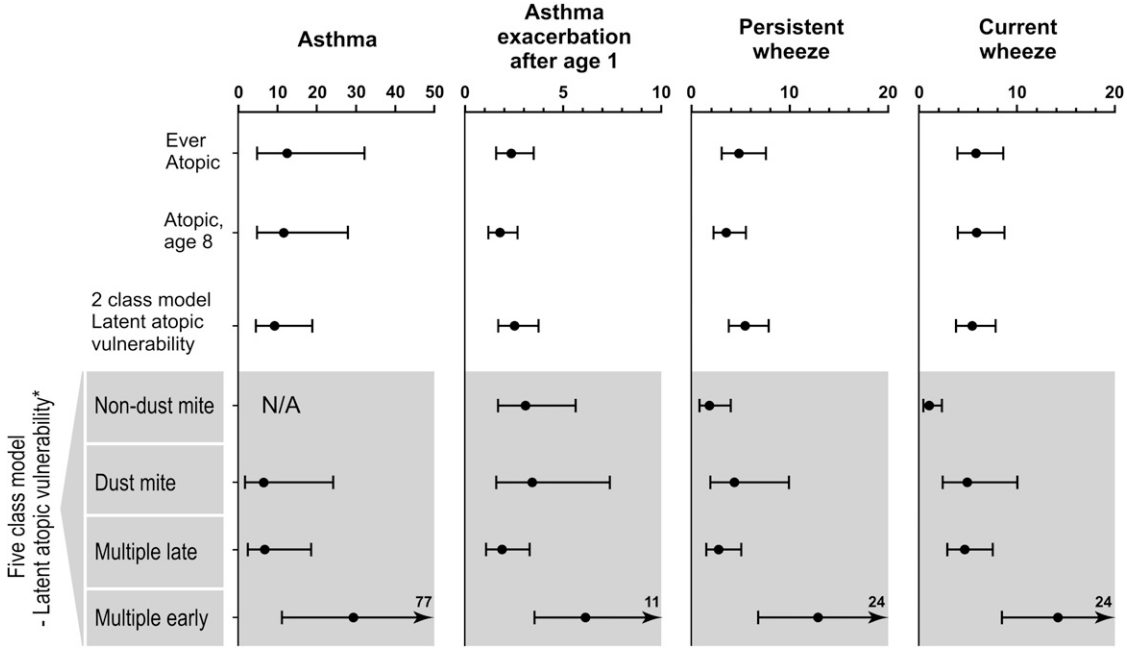

Figure 2. Association between atopy defined conventionally (atopic ever, atopic at age 8 yr), the novel latent classes (two-cluster and five-cluster models), and clinical phenotypes associated with asthma ascertained by age 8 years: regression analysis adjusted for gender. Results expressed as adjusted odds ratios and 95% confidence intervals.

TABLE 1. LUNG FUNCTION AND AIRWAY REACTIVITY IN CHILDREN WITH DIFFERENT LATENT ATOPIC VULNERABILITIES IN THE FIVE-CLASS MODEL

Lung Function Measure	No Latent Atopic Vulnerability	Non–Dust Mite	Dust Mite	Multiple Late	Multiple Early
sGaw age 3, kPa/s	0.90	0.90	0.97	0.88	0.83
	0.87–0.92	0.83–0.98	0.89–1.07	0.83–0.94	0.78–0.88
sGaw age 5, kPa/s	0.84	0.82	0.84	0.85	0.81
	0.82–0.87	0.79–0.86	0.79–0.90	0.82–0.88	0.78–0.84
FEV_1 age 8, L/sec	1.58	1.62	1.60	1.57	1.47
	1.54–1.62	1.54–1.70	1.54–1.66	1.53–1.61	1.42–1.52
FEV_1/FVC age 8, %	86.01	86.86	85.97	85.32	83.82
	84.76–87.25	84.38–89.34	84.02–87.92	84.08–86.56	82.40–85.24
Dose-response ratio age 8	8.40	6.33	7.64	9.73	13.32
	6.61–10.49	3.61–10.42	5.07–11.28	7.55–12.52	9.96–18.03

Estimated marginal means and 95% confidence intervals from multiple analysis of variance models adjusted for gender, maternal smoking and wheezing (sGaw and FEV_1/FVC ratio) and gender, wheezing, maternal smoking and height (FEV_1).

a highly significant increase in risk only among children in the multiple early class (HR 9.2; 95% CI, 3.5–24; $P < 0.001$).

DISCUSSION

Principal Findings

We have demonstrated that genuinely novel phenotypes of atopy can be revealed by adopting a machine learning approach that takes full advantage of the data-intensive environment provided by a birth-cohort study. Machine learning techniques identified latent structures within the data that may accurately reflect "unbiased" phenotypes of atopy and avoid constraints of investigator-imposed classifications. Our results suggest that IgE antibody responses do not reflect a mere presence or absence of atopy, but rather multiple atopic vulnerability classes. The validity of these classes was tested by examining their relations to the presence and severity of asthma and measures of lung function, which demonstrated that various atopic vulnerabilities (i.e., different phenotypes of atopy) differ markedly in their relationship with asthma. It is not the presence or absence of specific IgE antibodies, but rather the pattern of the response (age at development, type and number of specific allergens involved) that has a fundamental effect on the clinical expression of asthma. It is of note that less than a third of children who would have been considered atopic at age 8 years using conventional diagnostic criteria were in the class most strongly associated with asthma (multiple early), whereas there was little appreciable increase in risk of asthma among those in the other classes. We propose that positive specific IgE or positive SPTs do not equate to atopy, but should be viewed as intermediate phenotypes of a true atopic vulnerability. This may be analogous to asthma, where a collection of intermediate phenotypes can objectively be measured (e.g., peak flow variability, AHR, or an obstructive spirometric pattern), but individually their presence does not equate to a diagnosis of asthma (26).

Strengths and Limitations

We recognize that Bayesian learning applied to a longitudinal dataset is exploratory and hypothesis generating, rather than confirmatory. However, the classes we identified seem intuitively correct (i.e., have face validity), and we have demonstrated significant relationships with asthma, lung function, and airway reactivity (i.e., have content validity).

We acknowledge the computational complexity and intensity of this analysis. It is important to emphasize that this is not a simple "black box" or the "data-mining" approach; the analysis is informed by and capitalizes on the wealth of knowledge that already exist on the problem. Once determined, the classes may become clinical outcome variables in their own right and can be used in further analyses. Such dimensionality reduction reduces the need for repeated cross-sectional analysis, as often seen in longitudinal datasets, and reduces the need for multiple testing.

A strength of our model is that is generative, enabling missing measurements to be handled meaningfully. A further strength of the study is that the prevalence of atopic sensitization among the parents of the children in our cohort (27, 28) is similar to that of young adults in the United Kingdom (29), suggesting that the cohort is representative of the general population. However, it would be of great value and importance to explore similar approaches in the other large birth cohort studies. We recognize that the number of relevant classes might be different to the five reported here, and further replications would be desirable.

We acknowledge that our findings do not have an immediate impact on clinical practice. However, we argue that our approach to data analysis will advance the understanding of the etiology of asthma.

Interpretation of the Study

The study of asthma at the population level to date has been predominantly hypothesis driven, often focusing on ill-defined, oversimplified phenotypes, using reductionist approaches to causality. Although identifying some major independent determinants of disease, this approach does not fully reflect the complexity of disease. Furthermore, it fails to take full advantage of the richness of the available datasets collected in birth cohort studies. We propose that one of the reasons for contradictory findings reported by a number of genetic and environmental studies aiming to elucidate the mechanisms of asthma is phenotypic heterogeneity and poor phenotype definition.

In epidemiologic studies of allergic diseases investigators collect large volumes of information, often at multiple time points. Data on sensitization collected over a time series may be used to assign a phenotype based on distinctive patterns of results (e.g., early, late, or very late IgE sensitization [30]; monosensitization or polysensitization [30]; remission or persistence [30]; and declining, flat, or increasing pattern [31]). These categories are often imposed by the investigators, and do not necessarily reflect the substructure within the dataset. Ideally, one should aim to model all the data to identify a single multinomial latent variable that best describes the structure of the data. By using a machine learning approach, we have demonstrated that diagnostic label of "atopy" encompasses several different phenotypes that may have different etiologies.

Because these classes better reflect the presence and severity of disease, we propose that further efforts be made to develop new diagnostic tests that allow clinicians to differentiate better

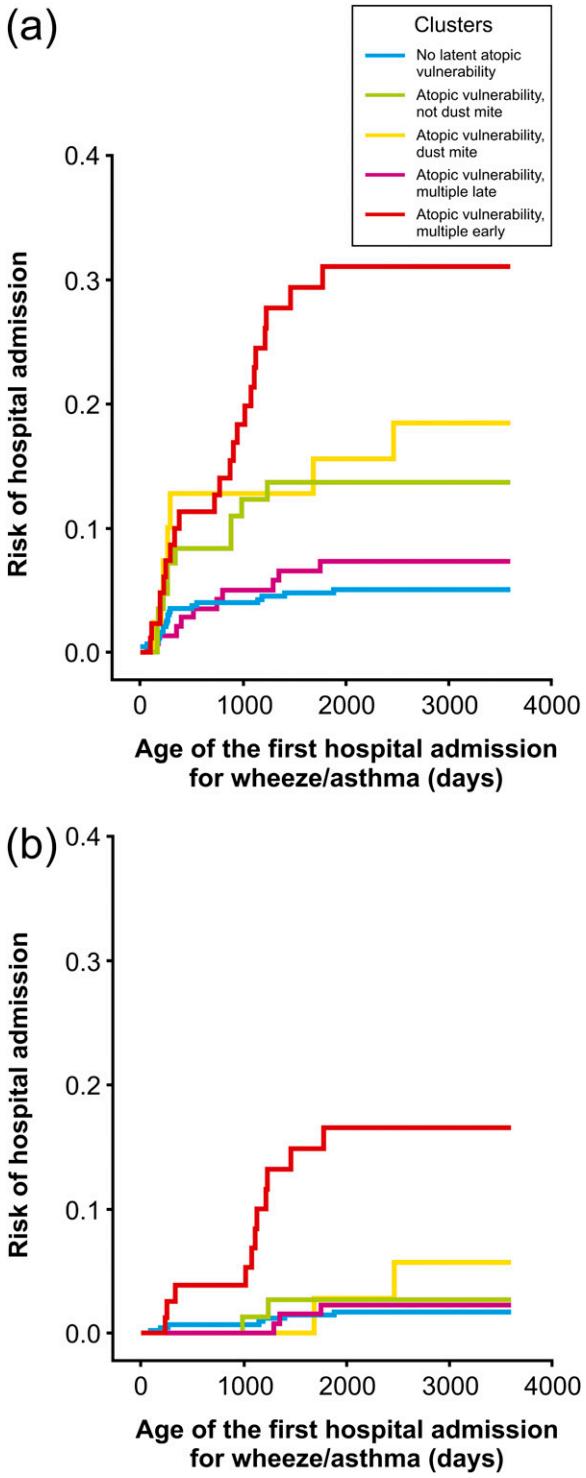

Figure 3. Kaplan-Meier estimates of cumulative risk of hospital admission with wheeze or asthma during the first 8 years of life stratified on five-class model. (*A*) Age at first hospital admission for children with hospital admission with wheeze or asthma at any age. (*B*) Age at first hospital admission among children who had a hospital admission after age 3 years.

between the true atopic classes than the currently available tests. Current reagents for skin testing and specific IgE measurement are based on whole extracts containing multiple proteins, many of which are recognized by IgE antibodies (32) (e.g., for dust mite *Dermatophagoides pteronyssinus* there are > 20 recognized allergens [33]). We speculate that response to different individual proteins within an allergen may be associated with different atopic classes (and consequently different clinical phenotypes). Use of this component-based approach may offer novel diagnostic possibilities and improve the value of allergy diagnosis, allowing practicing physicians more accurate diagnosis based on a single measurement at the time of presentation.

We have previously extended the observation that sensitization to inhalant allergens is a risk factor for wheezing by demonstrating that the level of specific IgE antibodies offers more information than just the presence of IgE (34). The current paper introduces the concept of different atopic vulnerabilities with distinct characteristics in terms of their association with disease. We have demonstrated that only one of the atopic classes (multiple early) predicts asthma. This may in part explain the huge variability in the relationship between "atopy" and asthma observed in the different parts of the world (e.g., the fraction of wheeze attributable to sensitization ranges from 0% in Turkey to 94% in China [35]), because the relative contribution of different atopic vulnerabilities to "atopy" may differ in each location consequent to differences in genetic predisposition and environmental exposures.

CONCLUSIONS

Viewing atopic sensitization as a dichotomous trait in its relationship to asthma may be an oversimplification. Our data suggest that IgE antibody responses do not reflect a single phenotype of atopy, but rather several different atopic vulnerabilities that differ in their relationship with asthma. One of these atopic vulnerability classes (multiple early, comprising approximately one quarter of children who would be considered atopic using conventional definition) predicts not only the presence, but also persistence and severity of childhood asthma.

Conflict of Interest Statement: A.S. received up to $1,000 in lecture fees for CME activity from GlaxoSmithKline and $5,001–$10,000 from Phadia Research in industry-sponsored grants for collaborative research. V.Y.F.T. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.W. is a full-time employee of Microsoft Research and has various patents through normal course of work within Microsoft Research. M.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. C.M.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. D.E.H.'s dependent is an employee of Microsoft, has received or is pending a patent from Microsoft in algorithm patents, and holds more than $100,001 in stock ownership or options in Microsoft. I.B. received more than $100,001 from Microsoft in funding to develop methods for genetic epidemiology. A.C. received $1,001–$5,000 from Pfizer and $1,001–$5,000 from UCB Pharma in consultancy fees, $1,001–$5,000 from ALK and $1,001–$5,000 from GlaxoSmithKline in advisory board fees, $5,001–$10,000 from GlaxoSmithKline in promotional lecture fees, $1,001–$5,000 from AstraZeneca, up to $1,000 from Phadia, up to $1,000 from ALK, and up to $1,000 from Novartis in lecture fees, and $5,001–$10,000 from Phadia in collaborative industry-sponsored grants.

Acknowledgment: The authors thank the children and their parents for continued support and enthusiasm. They greatly appreciate the commitment they have given to the project. They also acknowledge the hard work and dedication of the study team (research fellows, nurses, physiologists, technicians, and clerical staff). The authors thank Dr. Aida Semic-Jusufagic, who collected the primary care data. They thank Dr. John Guiver for his assistance with the use of Infer.NET.

References

1. Johansson SG, Hourihane JO, Bousquet J, Bruijnzeel-Koomen C, Dreborg S, Haahtela T, Kowalski ML, Mygind N, Ring J, van Cauwenberge P, *et al.* A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* 2001;56:813–824.
2. Johansson SG, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, Motala C, Ortega Martell JA, Platts-Mills TA, Ring J, *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.
3. Gergen PJ, Turkeltaub PC. The association of individual allergen reactivity with respiratory disease in a national sample: data from the second National Health and Nutrition Examination Survey, 1976–80 (NHANES II). *J Allergy Clin Immunol* 1992;90: 579–588.
4. Litonjua AA, Sparrow D, Weiss ST, O'Connor GT, Long AA, Ohman JL Jr. Sensitization to cat allergen is associated with asthma in older men and predicts new-onset airway hyperresponsiveness. The Normative Aging Study. *Am J Respir Crit Care Med* 1997;156:23–27.
5. Sears MR, Herbison GP, Holdaway MD, Hewitt CJ, Flannery EM, Silva PA. The relative risks of sensitivity to grass pollen, house dust mite and cat dander in the development of childhood asthma. *Clin Exp Allergy* 1989;19:419–424.
6. Duffy DL. A population-based study of bronchial asthma in adult twin pairs. *Chest* 1992;102:654.
7. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 2006;7:95–100.
8. von Mutius E. Allergies, infections and the hygiene hypothesis: the epidemiological evidence. *Immunobiology* 2007;212:433–439.
9. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008;178:218–224.
10. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, Strachan DP, Shaheen SO, Sterne JA. 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.
11. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino Jr R, Castro M, Curran-Everett D, Fitzpatrick AM, *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2009;181:315–323.
12. 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.
13. Custovic A, Simpson A, Woodcock A. Manchester cohort. *Pediatr Pulmonol Suppl* 2004;26:12–13.
14. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatr Allergy Immunol* 2002;13:32–37.
15. Lowe L, Murray CS, Custovic A, Simpson BM, Kissen PM, Woodcock A. Specific airway resistance in 3-year-old children: a prospective cohort study. *Lancet* 2002;359:1904–1908.
16. Nicolaou NC, Simpson A, Lowe LA, Murray CS, Woodcock A, Custovic A. Day-care attendance, position in sibship, and early childhood wheezing: a population-based birth cohort study. *J Allergy Clin Immunol* 2008;122:500–506 e5.
17. Lowe LA, Woodcock A, Murray CS, Morris J, Simpson A, Custovic A. Lung function at age 3 years: effect of pet ownership and exposure to indoor allergens. *Arch Pediatr Adolesc Med* 2004;158:996–1001.
18. Crapo RO, Casaburi R, Coates AL, Enright PL, Hankinson JL, Irvin CG, MacIntyre NR, McKay RT, Wanger JS, Anderson SD, *et al.* Guidelines for methacholine and exercise challenge testing-1999. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. *Am J Respir Crit Care Med* 2000;161:309–329.
19. Chinn S. Choosing a transformation. *J Appl Stat* 1996;23:395–404.
20. Woolcock AJ, Dowse GK, Temple K, Stanley H, Alpers MP, Turner KJ. The prevalence of asthma in the South-Fore people of Papua New Guinea. A method for field studies of bronchial reactivity. *Eur J Respir Dis* 1983;64:571–581.
21. Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, Casale TB, Chanez P, Enright PL, Gibson PG, *et al.* An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med* 2009;180:59–99.
22. Rabiner LR, Juang BH, Levinson SE, Sondhi MM. Some properties of continuous hidden Markov model representations. *At&T Technical Journal* 1985;64:1251–1270.
23. Winn J, Bishop CM. Variational message passing. *J Mach Learn Res* 2005;6:661–694.
24. Custovic A, Simpson BM, Simpson A, Kissen P, Woodcock A. Effect of environmental manipulation in pregnancy and early life on respiratory symptoms and atopy during first year of life: a randomised trial. *Lancet* 2001;358:188–193.
25. Woodcock A, Lowe LA, Murray CS, Simpson BM, Pipis SD, Kissen P, Simpson A, Custovic A. Early life environmental control: effect on symptoms, sensitization, and lung function at age 3 years. *Am J Respir Crit Care Med* 2004;170:433–439.
26. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M, Gibson P, Ohta K, O'Byrne P, Pedersen SE, *et al.* Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J* 2008;31:143–178.
27. Simpson BM, Custovic A, Simpson A, Hallam CL, Walsh D, Marolia H, Campbell J, Woodcock A. NAC Manchester Asthma and Allergy Study (NACMAAS): risk factors for asthma and allergic disorders in adults. *Clin Exp Allergy* 2001;31:391–399.
28. Custovic A, Simpson BM, Simpson A, Hallam CL, Marolia H, Walsh D, Campbell J, Woodcock A. Current mite, cat, and dog allergen exposure, pet ownership, and sensitization to inhalant allergens in adults. *J Allergy Clin Immunol* 2003;111:402–407.
29. Broadfield E, McKeever TM, Scrivener S, Venn A, Lewis SA, Britton J. Increase in the prevalence of allergen skin sensitization in successive birth cohorts. *J Allergy Clin Immunol* 2002;109:969–974.
30. Matricardi PM, Bockelbrink A, Keil T, Gruber C, Niggemann B, Hamelmann E, Wahn U, Lau S. Dynamic evolution of serum immunoglobulin E to airborne allergens throughout childhood: results from the Multi-Centre Allergy Study birth cohort. *Clin Exp Allergy* 2009;39:1551–1557.
31. Matricardi PM, Bockelbrink A, Gruber C, Keil T, Hamelmann E, Wahn U, Lau S. Longitudinal trends of total and allergen-specific IgE throughout childhood. *Allergy* 2009;64:1093–1098.
32. Harwanegg C, Laffer S, Hiller R, Mueller MW, Kraft D, Spitzauer S, Valenta R. Microarrayed recombinant allergens for diagnosis of allergy. *Clin Exp Allergy* 2003;33:7–13.
33. Chapman MD. Allergen nomenclature. *Clin Allergy Immunol* 2008;21: 47–58.
34. Simpson A, Soderstrom L, Ahlstedt S, Murray CS, Woodcock A, Custovic A. IgE antibody quantification and the probability of wheeze in preschool children. *J Allergy Clin Immunol* 2005;116:744–749.
35. Weinmayr G, Weiland SK, Bjorksten B, Brunekreef B, Buchele G, Cookson WO, Garcia-Marcos L, Gotua M, Gratziou C, van Hage M, *et al.* Atopic sensitization and the international variation of asthma symptom prevalence in children. *Am J Respir Crit Care Med* 2007;176: 565–574.