

Allergic Polysensitization Clusters: Newly Recognized Severity Marker in Urban Asthmatic Adults

Brian J. Patchett^a Bede N. Nriagu^{a, b} Granit Mavraj^a Ruchi R. Patel^a
Christopher MacLellan^c Tushar Thakur^a Edward S. Schulman^a

^aDivision of Pulmonary, Department of Medicine, Critical Care, and Sleep Medicine, Drexel University College of Medicine, Philadelphia, PA, USA; ^bDrexel University Dornsife School of Public Health, Philadelphia, PA, USA;

^cDrexel University College of Computing and Informatics, Philadelphia, PA, USA

Keywords

Asthma · Cluster analysis · Polysensitization · Allergy

Abstract

Introduction: While reliable, quantitative in vitro testing for sensitivity to aeroallergens has been available for decades, such information has largely been ignored in clustering analyses of asthma. Our aim is to explore allergic polysensitization as a possible marker of asthma severity and, as such, to be considered as an integral marker in future asthma clustering analyses. **Methods:** We constructed a database of sensitizations to the 25 aeroallergens in our geographic area (zone 1, Northeastern US) using the ImmunoCAP[®] in vitro assay. We used the Scikit-Learn[®] machine learning library for model-based clustering to identify allergic polysensitization clusters. Clusters were compared for differences in common office-based clinical markers of asthma. **Results:** The database consisted of 509 patients. Unbiased machine learning identified ten clusters of increasing allergic polysensitization of varying sizes ($n = 1\text{--}339$) characterized by significant increases in mean serum immunoglobulin E ($p < 0.001$), peripheral blood eosinophil count ($p < 0.001$), and D_{LCO} ($p = 0.02$). There was a significant decline in mean age at presen-

tation ($p < 0.001$), FEV_1/FVC ($p = 0.01$), and FEF_{25-75} ($p = 0.002$) with increasing allergic polysensitization. Finally, we identified two divergent paths for the poly-atopic march, one driven by perennial and the other by seasonal allergens. **Conclusion:** This pilot study showed that allergic polysensitization, using readily available qualitative and quantitative in vitro sensitization data, largely ignored in cluster analyses to date, may add further clinical precision in cluster analyses of asthma. We suggest the methods used here can be applied and tested using larger databases and aeroallergens present in diverse geographic regions.

© 2022 S. Karger AG, Basel

Introduction

Asthma is known as both one of the most common chronic respiratory conditions as well as a highly heterogeneous disease [1]. An active area of research has been characterizing the heterogeneity of asthma and assessing how cohorts with similar characterizations can predict more precision treatments [2, 3]. In recent years, machine

Edited by: H.-U. Simon, Bern.

Table 1. Major severe asthma cluster analyses and how atopy was defined

Cluster analysis	Definitions of atopy
Haldar et al. (2008) [4]	Presence or absence of atopy if a weal developed that was 2 mm larger than saline negative control for at least one allergen tested: cat fur, dog dander, grass pollen, and <i>D. pteronyssinus</i>
Moore et al. (2010) [5]	Number of positive skin prick test (12 aeroallergens, weal sized not specified)
Amelink et al. (2013) [6]	Presence or absence of atopy by at least one positive specific IgE test
Kim et al. (2013) [7]	Presence or absence of atopy based on at least one positive skin prick test to 12 common allergens (weal size unspecified)
Silkoff et al. (2015) [8]	Atopy not assessed
Schatz et al. (2014) [10]	Presence or absence of self-reported allergic triggers (pollen, animals, mold, dust)
Newby et al. (2014) [11]	Presence or absence of perennial or seasonal rhinitis, presence or absence of atopy via skin prick test
Zaihra et al. (2016) [12]	Presence or absence of atopy by skin prick test if at least one weal was >3 mm

learning-derived cluster analyses have identified subgroups that share asthma phenotypes and endotypes. Haldar et al. [4] from the UK was one of the first groups to identify severe asthma clusters and would be soon followed by the Severe Asthma Research Program [5] (SARP) from the NIH/NHLBI. As research groups explore findings across large and diverse cohorts [6–8], a goal is to prospectively identify markers useful to classify patients within well-characterized clusters [9–13]. Atopic asthmatics have been identified by independent analyses for their high disease burden and medication usage [14, 15].

In atopic asthma, immunoglobulin E (IgE)-mediated sensitization to an allergen(s) and subsequent allergen re-exposures can serve as a major driver of asthmatic symptoms and exacerbations [16, 17]. Between 51–80% of patients will have sensitization to more than one allergen, a state referred to as polysensitization [18]. Moreover, poly-sensitized patients can experience more severe allergy and asthma symptoms than mono-sensitized patients [19]. To date, however, cluster analyses have largely dis-

regarded laboratory quantitation of atopic sensitization, or used self-reported atopy, or included atopy as a simple touchpoint in their analysis if one or more skin prick tests were positive over control (Table 1). These analyses have not considered the number of positive sensitizations, the quantitative level of sensitization that can be measured by allergen-specific IgE, or co-sensitization/synergy between specific allergens as either variables in the clustering or as a biomarker to predict cluster assignment. Thus, an individual with one weakly positive skin test or IgE-allergen sensitization and an individual that is highly poly-sensitized are both labeled equally atopic. The focus of this study is the number of sensitizations, co-sensitizations, and clustering of specific IgE levels in the aeroallergen panel [20, 21] from patients in our urban asthma clinic. Moreover, we assess how demographic and phenotypic markers of asthma might be predictably influenced in patients matched for similar positive allergens, co-sensitizations, and levels of sensitivity to those allergens.

Methods

Chart Review

We performed a retrospective chart review of adults with moderate to severe asthma who presented to the Drexel University outpatient Pulmonary/Asthma Clinic between 2008 and 2019 and analyzed IgE levels for each positive allergen in the widely used ImmunoCAP® 25 aeroallergen panel. In our geographic region, we selected the zone 1 Northeastern US panel (see online suppl. Table E1 for list; for all online suppl. material, see www.karger.com/doi/10.1159/000526706) for identification of allergen sensitizations. All tests were performed by either Labcorp or Quest Diagnostics. Multiple studies have shown the ImmunoCAP® assay to be quantitative, accurate, and reproducible [20–22]. Though we ordered the zone 1 profile, specific allergen profiles are available for twenty different US regions including Puerto Rico. The research was approved by the Drexel Institutional Review Board in an expedited review. Written informed consent was waived because of de-identification. Diagnosis of moderate or severe asthma was made consistent with Step 4 or 5 GINA and/or ERS/ATS guidelines [23, 24]. Information extracted included self-identified demographics, smoking history, BMI, asthma and allergy medications, comorbid conditions (e.g., COPD, rhinitis), total IgE level, pulmonary function tests (PFT), and earliest recorded peripheral blood eosinophil count (PBEC). All extracted information came from or was ordered at the patient's initial specialist visit. Missing data were not imputed. Our exclusion criteria included patients <18 years old, those receiving allergen-specific immunotherapies, systemic corticosteroid use within a month of blood draw, hematologic malignancies, and allergic bronchopulmonary aspergillosis. Patients with asthma-COPD overlap were included if there was a history of childhood asthma or physician documented asthma before age 30. A high poverty zip code was defined as having a poverty rate higher than the national average, 10.5% [25, 26], from the City of Philadelphia Health Dashboard [27].

Preprocessing

During the study period, the ImmunoCAP® lower detection limit changed from “<0.35 IU/mL” to “<0.10 IU/mL” from laboratory providers. In either case, when the test reported less than the lower limit, we entered a “0” into our database. If the upper limit of detection (>100 IU/mL) was reported, we entered “100” into our database. Though normalization is a common step in machine learning data processing, because the sensitization data was on the same scale (i.e., all variables were specific IgE with the same range), this was not necessary and therefore was not done. We used the elbow method with a scree plot [28] (online suppl. Fig. E1) for dimensionality reduction; the elbow occurred between 3 and 6 principal components. We selected 4 principal components because the resulting clustering best balanced explaining the data without overfitting.

Dirichlet Process Mixture Models

To identify allergic polysensitization (APS) profiles, we performed cluster analysis in the form of Dirichlet Process Mixture Modeling (DPMM) [29] on the ImmunoCAP® 25 aeroallergen-specific IgE panel by utilizing the Scikit-Learn machine learning library [30] in Python version 3.9.0. This library uses a Variational Bayes (VB) [31, 32] approach with a Dirichlet process prior to determine the cluster number, means, and covariance matrices as well as cluster assignment. Unlike other clustering algorithms where the user must predetermine the appropriate cluster number, with a DPMM, the user is only required to input an upper bound for the appropriate cluster number. With this upper bound, some of the clusters will be used while others will be given a weight of zero and contain no observations/patients [33]. In this work, a cluster number of 20 was estimated based on preliminary clustering trials to be a valid upper limit. All other parameters required for DPMMs were kept at the Scikit-Learn program-calculated defaults.

Assessment of Cluster Correlations

After visualizing the clustering results, it was hypothesized sensitization to each allergen did not increase uniformly between clusters. A correlation circle was plotted with the contribution of each allergen to the first two principal components [34]. Viewing these findings, we further hypothesized there would exist clusters of the clusters, here called “superclusters.” To evaluate this, we created a correlation matrix of the clusters, then ordered the matrix by hierarchical clustering. Correlation matrix plotting was done with corplot [35], and factor analysis was done with the factoextra [36] and FactoMineR [37] packages in RStudio version 1.3 [38].

Between-Cluster Comparison of Clinical Markers

Given the known relationship between IgE and asthma, we hypothesized clusters associated with higher levels of sensitization would be predictive of more perturbed asthmatic clinical markers. To assess this, we first calculated an APS score for each cluster defined by the overall average allergen sensitization. Simple linear regression was then used to assess the relationship between the cluster APS scores and common clinical markers of allergic asthma, including serum IgE, PBEC, PFT results, and demographic data. The data for serum total IgE and PBEC were linearized prior to regression analysis. Plotting was done with ggplot2 [39] and dplyr [40].

Table 2. Demographics of subjects in the study

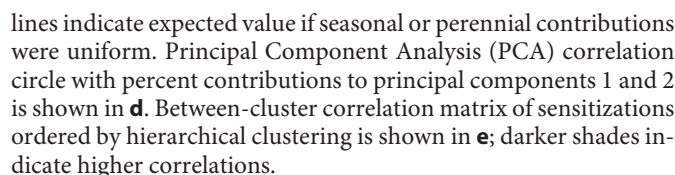
Demographics of subjects	
N	509
Age (SD)	50.48 (13.9)
Age groups (%)	
18–35 years	84 (16.6)
36–64 years	349 (69.0)
≥65 years	73 (14.4)
Missing	3 (0.58)
Male:female ratio ^a	1:3.4
Percent male, percent female	22.9%, 77.1%
BMI kg/m ² (SD)	33.0 (9.3)
BMI groups (%)	
<18.5 kg/m ²	12 (2.4)
18.5–24.9 kg/m ²	89 (17.5)
25.0–29.9 kg/m ²	93 (18.2)
≥30 kg/m ²	267 (52.4)
Missing	48 (9.4)
Race (%)	
Black	315 (61.9)
White	131 (25.7)
Latino	26 (5.1)
Asian	23 (4.5)
Other/not specified	14 (2.8)
In zip codes of high poverty (%)	76.2%
Black	270 (87.8)
White	61 (48.0)
Latino	22 (88.3)
Asian	18 (78.3)
Other/not specified	3 (50)
Smoking ≥10 pack-years (%)	198/494 (40.1)

^a Only the latter 179 patients had their gender recorded.

Results

Chart Review

A demographic review of the 509 patients who met inclusion criteria is shown in Table 2. The predominant racial group was black/African American (61.9%). Most patients identified as female (77.1%). The predominant age range was 36–64 years old (mean = 50.4 ± 13.9, range of 20–88 years). BMI average was 33.0 ± 9.3 kg/m² (range of 16.3–72.1 kg/m²). A large majority (76.2%) resided in high-poverty zip codes. This was particularly evident among black (87.8%) and Latino (88.3%) patients. Current and past smoking history of >10 pack-years was documented in 40.1% of patients.



The spread of the sensitizations for each cluster was right-skewed; therefore, the mean and standard deviation (Table 4a–c) as well as the median and interquartile range (Table 4d–f) are reported. Broadly, cluster assignment was determined by both which allergens the patient was sensitive to and their specific IgE levels. Assignment to

Table 3. Demographics of each cluster

Demographic data	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Total, <i>n</i>	339	60	51	25	6	16	8	2	1	1
Age (SD)	52.3 (13.9)	46.9 (12.9)	48.9 (13.7)	48.7 (11.0)	52.8 (9.7)	45.8 (12.4)	32.6 (14.3)	30 (8.5)	43	21
Missing, <i>n</i> (%)	1 (0.3)	0 (0)	1 (4.2)	0 (0)	0 (0)	0 (0)	0 (0)	0	0	0
BMI (SD)	33.0 (9.5)	33.9 (9.1)	34.2 (9.1)	30.4 (8.5)	33.6 (7.6)	33.3 (10.7)	29.5 (7.5)	31.1 (4.8)	28.0	21.7
Missing, <i>n</i> (%)	28 (8.9)	13 (18.1)	2 (8.3)	5 (14.3)	1 (8.3)	0 (0)	1 (10)	0 (0)	0 (0)	0 (0)
High poverty zip code, <i>n</i> (%)	241 (73.3)	46 (78.0)	42 (82.4)	23 (100)	3 (60)	13 (81.6)	8 (100)	2 (100)	0 (0)	0 (0)
Missing, <i>n</i> (%)	10 (2.9)	1 (1.7)	0 (0)	2 (8.0)	1 (16.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Race										
Black	190	43	40	20	1	12	6	1	1	1
White	100	14	6	3	4	3	0	1	0	0
Latino	21	2	0	2	1	0	0	0	0	0
Asian	19	0	2	0	0	0	2	0	0	0
Other	9	1	3	0	0	1	0	0	0	0

Mean and standard deviation are given where appropriate.

cluster 1 appeared to be driven by a median of one positive sensitization (IQR 0–3, online suppl. Table E2). The APS score was only 0.11 IU/mL. The overall mean specific IgE increased with respect to cluster number, but not every allergen increased stepwise with each cluster. Cluster 2 patients had an APS of 1.39 IU/mL with notably high sensitizations to dogs and especially cats. Cluster 3 had a similar APS to cluster 2 of 1.69 IU/mL and tended to have high sensitization to *D. pteronyssinus* and *D. farinae* with comparatively modest sensitizations to the other allergens. Cluster 4 had a higher APS of 4.25 IU/mL with high sensitizations to Timothy and Bermuda grass, birch, oak, maple trees, as well as cats and dogs. Cluster 5 had an APS of 7.07 IU/mL. Assignment appeared driven by high-level sensitizations to cats and dogs, as well as a general increase in all seasonal allergies but low sensitization to molds. Similar to cluster 5, cluster 6 had an APS of 7.35 IU/mL, driven by high-level sensitizations to *D. pteronyssinus* and *D. farinae*, as well as cats and dogs. The APS of cluster 7 was 13.55 IU/mL, exposing prominent sensitizations to birch and oak trees, as well as cats and dogs.

Cluster 8 only had two patients with very high sensitizations to perennial allergens and one of the largest readings to cockroaches. Cluster 9 and cluster 10 were both single patients. The patient in cluster 9 had an APS of 28.59 IU/mL, driven by reaching the upper limit of detection for *D. pteronyssinus* and *D. farinae*, nearly so for timothy grass, and severe sensitizations to Bermuda grass, sycamore, oak, and white ash tree. The cluster 10 patient,

despite sensitizations to *D. pteronyssinus*, *D. farinae*, *C. herbarum*, and *P. chrysogenum*, comparable to cluster 1 patients, displayed the upper limit for detection for five allergens (combined mean = 44.71 IU/mL), approximately 400-fold higher than the average cluster 1 patient. Moreover, the patient's highest sensitizations exceeded all other database patients for 12 of the 25 allergens.

Seasonal versus Perennial Divergence of Clusters

The first principal component (horizontal axis) of the correlation circle explained 47.2% of the variance and was comprised of seasonal allergens (Fig. 1b). The second principal component (vertical axis) explained 14.4% of the data variance and was comprised of perennial allergens (Fig. 1c). Interestingly, sensitization to birch, oak, and white mulberry trees appeared to violate this trend. Factor analysis of the sensitizations was plotted on a correlation circle. Two groups of allergens emerged as orthogonal to each other (Fig. 1d). This indicated that within the groups, the allergens were strongly and positively correlated and relatively uncorrelated to the allergens in the other group.

Knowing these group separations, we hypothesized this would translate into polarized clusters that were individually homogenous for either perennial or seasonal allergens. To examine this possibility, we created a correlation matrix of the clusters and identified three “super-clusters” (demonstrated in Fig. 1e). The first was comprised of clusters 1, 3, 6, 8, and 9; the second was made up

Table 4. Sensitizations for each cluster

a Perennial	Cluster 1 (SD)	Cluster 2 (SD)	Cluster 3 (SD)	Cluster 4 (SD)	Cluster 5 (SD)	Cluster 6 (SD)	Cluster 7 (SD)	Cluster 8 (SD)	Cluster 9 (SD)	Cluster 10 (SD)
APS	0.11 (0.19)	1.39 (0.89)	1.69 (0.90)	4.25 (1.60)	7.07 (2.51)	7.35 (3.18)	13.55 (5.66)	17.24 (5.53)	28.59 (0)	44.71 (0)
<i>D. pteronyssinus</i>	0.26 (0.63)	1.05 (2.24)	6.47 (7.61)	2.78 (3.20)	2.38 (3.09)	49.23 (40.31)	6.02 (5.93)	100 (0)	100	0.37
<i>D. farinae</i>	0.27 (0.70)	1.66 (3.47)	7.81 (8.06)	2.90 (4.18)	4.75 (6.68)	54.18 (52.87)	8.44 (9.14)	100 (0)	100	0.39
<i>Alternaria</i>	0.35 (1.55)	2.86 (7.69)	1.75 (3.63)	4.33 (8.96)	2.98 (4.29)	0.26 (0.76)	13.33 (23.87)	2.44 (1.63)	1.77	1.16
<i>Aspergillus</i>	0.10 (0.62)	0.42 (1.47)	0.69 (1.95)	1.55 (3.14)	0.62 (0.68)	0.50 (1.25)	6.84 (7.33)	6.96 (5.56)	11.6	1.56
<i>Penicillium</i>	0.07 (0.40)	0.13 (0.47)	0.43 (0.95)	0.66 (1.56)	0.51 (0.79)	0.38 (0.54)	1.46 (1.03)	5.77 (4.02)	0.84	0
<i>C. herbarum</i>	0.03 (0.18)	0.20 (0.71)	0.40 (1.08)	1.31 (3.53)	0.73 (0.80)	0.11 (0.20)	3.73 (5.20)	4.28 (2.58)	0	0
Cat dander	0.14 (0.42)	18.75 (16.23)	1.54 (2.46)	6.75 (7.19)	79.11 (28.52)	53.94 (46.60)	66.18 (38.83)	21.57 (20.40)	19.1	0.86
Dog dander	0.08 (0.28)	3.41 (3.71)	0.90 (1.68)	4.38	13.06 (16.34)	42.83 (22.69)	8.62 (9.34)	76.73 (35.08)	76.65 (33.02)	22.5
Cockroach	0.25 (1.30)	0.91 (1.71)	2.91 (5.81)	2.68 (5.34)	1.45 (3.01)	4.25 (5.39)	13.98 (18.67)	46.29 (63.65)	11.5	2.18
b Grasses/weeds	Cluster 1 (SD)	Cluster 2 (SD)	Cluster 3 (SD)	Cluster 4 (SD)	Cluster 5 (SD)	Cluster 6 (SD)	Cluster 7 (SD)	Cluster 8 (SD)	Cluster 9 (SD)	Cluster 10 (SD)
Ragweed	0.23 (0.98)	1.13 (1.96)	2.92 (6.44)	6.85 (12.87)	17.46 (35.10)	1.24 (2.33)	6.55 (9.22)	35.75 (50.55)	16.6	100
Rough pigweed	0.02 (0.09)	0.22 (0.62)	0.46 (0.87)	1.34 (2.25)	1.19 (1.26)	0.97 (2.22)	3.13 (4.25)	2.66 (3.76)	1.64	64.4
Mugwort	0.01 (0.08)	0.34 (1.32)	0.97 (2.37)	1.58 (2.51)	1.93 (3.39)	0.07 (0.13)	6.90 (10.28)	0.19 (0.26)	8.75	68.4
Sheep sorrel	0.01 (0.06)	0.07 (0.18)	0.28 (0.79)	0.94 (1.44)	0.56 (0.72)	0.51 (1.32)	2.18 (2.63)	0.24 (0.34)	1.92	35.5
Bermuda grass	0.07 (0.40)	0.50 (1.14)	1.59 (3.67)	4.41 (7.16)	1.56 (1.88)	1.14 (2.31)	3.92 (3.85)	2.11 (2.09)	20.4	100
Timothy grass	0.13 (0.74)	0.91 (2.89)	3.17 (4.99)	7.92 (15.84)	4.15 (7.87)	0.32 (0.51)	5.34 (7.32)	1.69 (1.84)	98.6	100
c Trees	Cluster 1 (SD)	Cluster 2 (SD)	Cluster 3 (SD)	Cluster 4 (SD)	Cluster 5 (SD)	Cluster 6 (SD)	Cluster 7 (SD)	Cluster 8 (SD)	Cluster 9 (SD)	Cluster 10 (SD)
White mulberry	0.006 (0.06)	0.02 (0.11)	0.16 (0.51)	0.28 (0.63)	0.49 (0.56)	0.70 (1.81)	1.42 (2.05)	3.41 (4.82)	0	5.06
Cottonwood	0.03 (0.35)	0.07 (0.21)	0.35 (0.84)	1.59 (2.40)	1.42 (1.31)	0.11 (0.22)	3.08 (2.92)	0.97 (0.79)	5.63	45.5
Mountain cedar	0.05 (0.44)	0.08 (0.23)	0.38 (0.80)	0.79 (0.91)	0.67 (0.65)	1.03 (1.59)	2.01 (1.86)	4.65 (5.85)	11.6	31.9
Walnut tree	0.03 (0.18)	0.14 (0.41)	0.63 (1.33)	3.47 (6.00)	1.84 (2.18)	0.37 (1.16)	3.75 (5.27)	5.15 (6.69)	13.6	100
Sycamore	0.03 (0.19)	0.10 (0.29)	0.60 (1.37)	2.13 (3.18)	1.45 (1.30)	1.26 (3.25)	5.80 (6.81)	0.63 (0.21)	19.4	48.9
Maple	0.07 (0.38)	0.23 (0.53)	1.53 (4.69)	6.00 (11.17)	1.30 (1.23)	0.72 (1.61)	4.11 (3.10)	1.06 (0.82)	12.7	79.4
Birch	0.09 (0.47)	0.39 (1.20)	1.84 (3.15)	14.41 (21.27)	1.50 (1.79)	1.13 (2.30)	35.49 (32.19)	4.55 (5.86)	83	83.8
Elm	0.02 (0.11)	0.22 (0.69)	1.22 (3.27)	3.30 (5.82)	1.41 (1.39)	1.03 (2.08)	9.14 (18.08)	0.95 (0.67)	20.1	63.3
White ash	0.12 (0.68)	0.30 (1.03)	0.89 (1.86)	4.60 (6.76)	2.51 (2.69)	0.45 (0.94)	3.59 (4.90)	1.61 (1.71)	70.9	100
Oak	0.09 (0.49)	0.55 (1.56)	2.35 (4.11)	10.72 (14.55)	1.89 (1.74)	1.23 (2.51)	45.69 (38.35)	1.57 (1.61)	62.8	80.9

Table 4 (continued)

d Perennial	Cluster 1 (IQR)	Cluster 2 (IQR)	Cluster 3 (IQR)	Cluster 4 (IQR)	Cluster 5 (IQR)	Cluster 6 (IQR)	Cluster 7 (IQR)	Cluster 8 (IQR)	Cluster 9 (IQR)	Cluster 10 (IQR)
Median APS	0 (0–0)	0 (0–0.43)	0 (0–1.15)	0.61 (0–4.03)	0.87 (0–3.02)	3.21 (0–0.95)	3.41 (0.47–8.6)	17.24 (5.53)	28.59 (0)	44.71 (0)
<i>D. pteronyssinus</i>	0 (0–0.1)	0 (0–1.1)	4.31 (0–10.4)	1.81 (0.3–3.9)	0.86 (0.4–3.7)	53.4 (8.4–84.1)	5.43 (1.6–7.2)	100 (100–100)	100	0.37
<i>D. farinae</i>	0 (0–0.1)	0.13 (0–1.3)	5.34 (0–12.6)	0.98 (0.2–3.0)	0.82 (0.5–8.6)	73.1 (8.5–91.9)	5.79 (0.9–12.1)	100 (100–100)	100	0.39
<i>Alternaria</i>	0 (0–0)	0 (0–1.8)	0 (0–0.9)	0 (0–3.7)	0.45 (0–6.1)	0 (0–0.2)	3.96 (2.2–9.8)	2.44 (1.9–3.0)	1.77	1.16
<i>Aspergillus</i>	0 (0–0)	0 (0–0)	0 (0–0.3)	0.38 (0–1.4)	0.52 (0–1.3)	0 (0–0.1)	4.5 (0–14)	6.96 (5.0–8.9)	11.6	1.56
<i>Penicillium</i>	0 (0–0)	0 (0–0)	0 (0–0.3)	0 (0–0.2)	0 (0–1.1)	0.14 (0–0.6)	1.76 (0–2.4)	5.77 (4.4–7.2)	0.84	0
<i>C. herbarum</i>	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0.6)	0.71 (0–1.5)	0 (0–0.2)	0.31 (0–7.7)	4.28 (2.58)	0	0
Cat dander	0 (0–0)	13 (6.1–27.5)	0.38 (0–1.7)	4.48 (1.1–10.4)	92.55 (65–100)	72.5 (3.2–98.2)	76.1 (39.5–100)	21.57 (20.40)	19.1	0.86
Dog dander	0 (0–0)	2.05 (0.8–4.6)	0.14 (0–0.8)	2.85 (0.9–23.1)	44.15 (31.4–57.1)	4.96 (0.7–18.6)	99.1 (57.6–100)	76.65 (65.0–88.3)	22.5	4.38
Cockroach	0 (0–0)	0 (0–0.4)	0.13 (0–2.3)	0.6 (0–2.7)	0 (0–0.9)	1.76 (0.4–6.4)	5.9 (3.6–14.2)	46.29 (23.8–69.8)	11.5	2.18
e Grasses/weeds	Cluster 1 (IQR)	Cluster 2 (IQR)	Cluster 3 (IQR)	Cluster 4 (IQR)	Cluster 5 (IQR)	Cluster 6 (IQR)	Cluster 7 (IQR)	Cluster 8 (IQR)	Cluster 9 (IQR)	Cluster 10 (IQR)
Ragweed	0 (0–0)	0.22 (0–1.4)	0.58 (0–3.0)	2.5 (0.3–8.4)	3.9 (2.4–5.6)	0 (0–0.8)	2.27 (1.3–8.0)	35.75 (17.9–53.6)	16.6	100
Rough pigweed	0 (0–0)	0 (0–0.1)	0 (0–0.5)	0.25 (0–1.3)	0.91 (0.1–2.3)	0 (0–0.4)	1.26 (0.7–3.0)	2.66 (1.3–4.0)	1.64	64.4
Mugwort	0 (0–0)	0 (0–0.02)	0 (0–0.6)	0.1 (0–2.5)	0.75 (0.1–1.3)	0 (0–0.1)	5.00 (0–6.9)	0.19 (0.1–0.3)	8.75	68.4
Sheep sorrel	0 (0–0)	0 (0–0)	0 (0–0.2)	0.42 (0–0.9)	0.29 (0–0.9)	0 (0–0.4)	1.67 (0–2.7)	0.24 (0.1–0.4)	1.92	35.5
Bermuda grass	0 (0–0)	0 (0–0.4)	0 (0–1.2)	0.82 (0–8.6)	1.2 (0.2–1.9)	0.07 (0–0.4)	1.91 (1.4–6.6)	2.11 (1.4–2.9)	20.4	100
Timothy grass	0 (0–0)	0 (0–0.03)	0.54 (0–4.9)	1.4 (0–6.3)	1.32 (0.2–2.1)	0.07 (0.1)	1.61 (1.1–7.1)	1.69 (1.0–2.3)	98.6	100
f Trees	Cluster 1 (IQR)	Cluster 2 (IQR)	Cluster 3 (IQR)	Cluster 4 (IQR)	Cluster 5 (IQR)	Cluster 6 (IQR)	Cluster 7 (IQR)	Cluster 8 (IQR)	Cluster 9 (IQR)	Cluster 10 (IQR)
White mulberry	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0.2)	0.39 (0–0.9)	0 (0–0.2)	0.79 (0.2–1.6)	3.41 (1.7–5.1)	0	5.06
Cottonwood	0 (0–0)	0 (0–0)	0 (0–0.2)	0.47 (0–2.2)	1.16 (0.7–1.6)	0 (0–0.1)	3.21 (0–5.3)	0.97 (0.7–1.3)	5.63	45.5
Mountain cedar	0 (0–0)	0 (0–0)	0 (0–0.5)	0.57 (0–0.9)	0.64 (0.1–1.1)	0.29 (0–1.2)	1.7 (1.1–2.4)	4.65 (2.6–6.7)	11.6	31.9
Walnut tree	0 (0–0)	0 (0–0)	0 (0–0.4)	0.61 (0–4.8)	1.43 (0.2–2.3)	0 (0–0.2)	1.69 (0–4.6)	5.15 (2.8–7.5)	13.6	100
Sycamore	0 (0–0)	0 (0–0)	0 (0–0.4)	0.39 (0–3.1)	0 (0–0.4)	0 (0–0.4)	4.42 (0.3–8.2)	0.63 (0.6–0.7)	19.4	48.9
Maple	0 (0–0)	0 (0–0.3)	0 (0–1.0)	1.12 (0–7.2)	0.16 (0–0.4)	0.16 (0–0.4)	4.23 (1.5–6.6)	1.06 (0.8–1.4)	12.7	79.4
Birch	0 (0–0)	0 (0–0.1)	0.14 (0–2.2)	4.55 (0.8–24.9)	0 (0–0.3)	0 (0–0.3)	31.75 (13.6–49.3)	4.55 (2.5–6.6)	83	83.8
Elm	0 (0–0)	0 (0–0.1)	0 (0–1.2)	0.92 (0–4.6)	0.78 (0.6–2.6)	0.07 (0–0.4)	2.65 (1.3–4.6)	0.95 (0.7–1.2)	20.1	63.3
White ash	0 (0–0)	0 (0–0.03)	0 (0–0.7)	1.12 (0.1–7.7)	1.91 (0.8–3.1)	0 (0–0.3)	2.11 (0–4.0)	1.61 (1.0–2.2)	70.9	100
Oak	0 (0–0)	0 (0–0)	0.17 (0–4.2)	4.67 (0.5–18.5)	1.58 (0.6–3.1)	0 (0–0.4)	43.65 (14.1–78.6)	1.57 (1.0–2.1)	62.8	80.9

a–c Means and standard deviations (in parentheses) of each cluster's allergic polysensitization (APS) score, calculated by the average specific immunoglobulin E (IgE) for each patient in the cluster, and within cluster specific IgE. **d–f** Median and interquartile range (in parentheses).

of clusters 2, 4, 5, and 7; and the third was just the single patient in cluster 10. The sensitizations in the first supercluster tended to be low for the perennial allergens but progressively higher for seasonal allergens. In contrast, the second supercluster showed progressively higher se-

lect perennial allergen sensitizations (i.e., mold, dust, cockroach) while comparatively modest sensitizations to seasonal allergens. While the patient in cluster 10 was in their own supercluster, this person was qualitatively like the first supercluster (i.e., seasonal > perennial), though

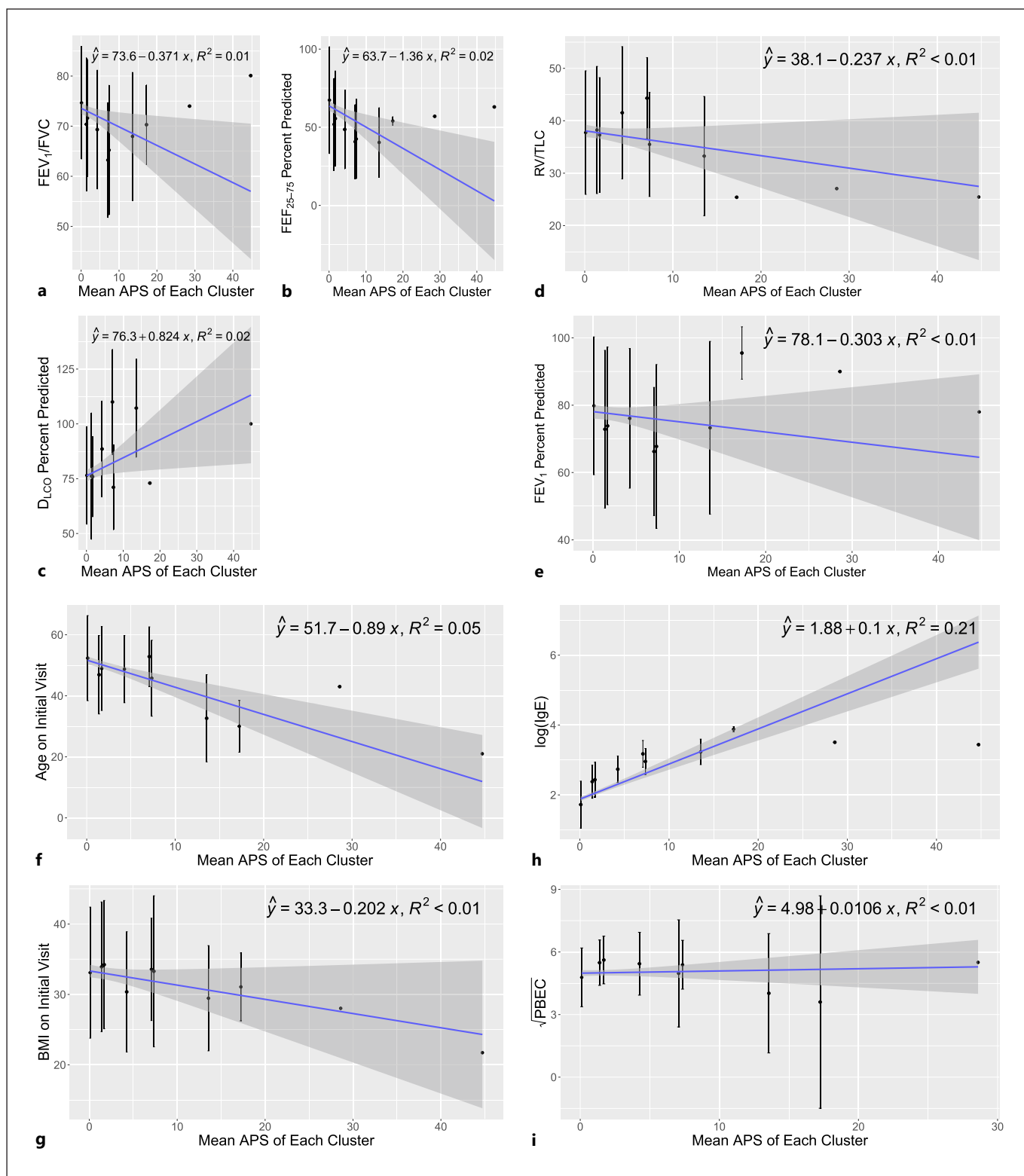


Fig. 2. Linear regression plots. **a-i** APS for each sequential cluster is plotted left to right (clusters 1 through 10), along the x-axes by their APS score and clinical variables are plotted along y-axes. APS is calculated by average sensitization level to all allergens for all

patients in the cluster. Points are mean $\pm$ SD. Overlaid is the best fit (blue line) from regression analysis and 95% confidence interval of regression line.

seasonal specific IgEs were far greater. Sensitizations to cats and dogs appeared comparable in all three superclusters and non-compartmentalized.

Significant Demographic Trends with Higher APS

Regression analysis was performed to assess the relationship between APS and common clinical markers of asthma. Findings are reported in online supplementary Tables E3 and E4 and plots are reported in Figure 2, each sequential point in the plot (left to right) represents clusters 1 through 10. A simple linear regression evaluated the relationship between age and APS cluster assignment (Fig. 2a). There was a significant negative correlation between APS and age ($p = 3.98 \times 10^{-5}$). The R^2 for age was 0.045, indicating 4.5% of the variability in the age a patient presented to a subspecialty clinic was explained by a linear regression model containing APS as a predictor. No relationship between APS and BMI was predicted by simple linear regression ($p = 0.197$, Fig. 2b). We did not find any significant differences when comparing clusters for differences in gender, race, or smoking history.

APS Correlated with More Severe Clinical Markers of Allergic Asthma

Overall, higher APS was associated with higher obstruction on PFTs (Fig. 2c–g). There was a statistically significant negative correlation between APS and FEV₁/FVC ($p = 0.01$, Fig. 2c), a statistically significant positive correlation between APS and D_{LCO} ($p = 0.02$, Fig. 2d), and a significant negative correlation between APS and FEF_{25–75} percent predicted ($p = 0.002$, Fig. 2e). The R^2 for these metrics was 0.01 and 0.02, therefore 1–2% of the variability of these metrics could be explained by a linear regression model containing APS as a predictor. There were no statistically significant correlations between APS and RV/TLC ($p = 0.14$, Fig. 2f) or FEV₁ percent predicted ($p = 0.26$, Fig. 2g).

The data for serum IgE and PBEC were highly non-linear and therefore required transformation prior to linear regression analysis. The data for serum IgE was log-transformed prior to regression analysis (not shown). There was a small percentage of patients (<5%) who had 0 to “<50/μL” PBEC recorded. Therefore, the square root was used to best achieve linearity. A significant regression equation was found between APS and the logarithm of serum IgE ($p < 2.2 \times 10^{-16}$, Fig. 2h) and between APS and the square root of PBEC ($p = 0.01$, Fig. 2i).

Linear Regression after Outlier Removal

After plotting the regression analysis, we considered the possibility that the 4 patients in clusters 8, 9, and 10 were acting as outliers and affecting the results. To assess this, we repeated simple linear regression after removing these patients. The repeat analysis led to a significant regression equation between FEV₁ percent predicted and APS ($p = 0.01$, online suppl. Table E3). The regression equations after removal of the outliers for the above variables were unchanged with respect to statistical significance.

Discussion

Our study population consisted of patients with specialist-diagnosed, moderate to severe asthma mostly from zip codes below the federal poverty line. The majority were racial minorities (Table 3). Vulnerable patients in low-resource areas have well-documented deficiencies in health outcomes [41], underscoring the need for tools to improve care that are low-cost and accessible [42]. Previous cluster analyses have relied in-part on clinical markers not available in many practitioner’s offices including exhaled nitric oxide, skin prick testing, and sputum eosinophilia versus neutrophilia [4, 5, 43]. Furthermore, no previous cluster analysis (Table 1) has used state-of-the-art quantitative allergen identification methods to examine if clusters matched for allergens, their levels of sensitization (0.1–100 IU/mL), and expressions of co-sensitizations would yield meaningful and clinically actionable insights. Atopy, when considered, was a uniform binary metric (any positive skin test vs. no skin test positive), failing to capture potential heterogeneity [44]. We turned to Dirichlet Process mixture modeling [29], a well-established machine learning approach, to impartially analyze our dataset of zone 1 Northeastern US ImmunoCAP® 25 aeroallergen profiles. In multiple studies, ImmunoCAP® assays are shown to be highly reproducible and accurate [20, 21]. We probed APS clusters and found retrospective assignments to predict several important clinical markers of asthma.

Unlike other clustering algorithms, the mixture modeling used here is a probabilistic approach and allows for the prospective assignment of new patients with novel APS patterns. Although we applied modeling to the zone 1 Northeastern US ImmunoCAP® in this pilot study, the collaborative development of mixture models at the population level for each ImmunoCAP® zone could be used to translate previously transactional sensitization reports into more clinically predictive and applicable informa-

tion. We also aimed to have our approach serve as a model for future work where APS clustering could enhance comprehensive databases and be available in any connected location. APS clustering may lead to further understanding of how specific sensitization patterns are influenced by additional asthma risk factors (e.g., genetics, microbiome). Insights into pharmacologic outcomes, especially for biologic therapies [44], may be gained if examined in relation to APS clusters. Lastly, recognition that atopy is not a single metric, APS as a biomarker may prove important as a standardized variable in future asthma cluster analyses (Table 1) and could enhance databases for understanding other respiratory conditions (e.g., asthma-COPD overlap) [45].

We used Dirichlet Process mixture models, which are an example of Bayesian non-parametric mixture models. Cluster number, size, shape, and cluster assignment were determined by a VB approach with a Dirichlet process prior, allowing these to be independent, unbiased, machine algorithm-driven and determined “automatically.” Bayesian non-parametric in this context means these models can have unbounded, growing, or infinite number of parameters (cluster number or complexity). Therefore, the Dirichlet Process has applications in many fields [46, 47] where potentially new observations represent a novel cluster rather than a prior one. With 25 aeroallergens in zone 1 ImmunoCAP® assays where specific IgEs range from 0.10 to 100 IU/mL, there are almost a googol (10^{100}) possible qualitative and quantitative sensitization and co-sensitization profiles. To account for the possibility that our datasets are incomplete, we turned to DPMMs, though this, along with clustering in four principal components, led to a large low APS cluster (cluster 1) and several small, high APS clusters including two “clusters” of $n = 1$. We acknowledge that clustering in fewer dimensions (see scree plot, online suppl. Fig. E1) led to fewer and, in certain cases, larger clusters. This approach, in favor of limiting dimensions, likely resulted in the identification of less rare and unique profiles. However, it did successfully identify seasonal versus perennial trends. Our expectation is that beyond this pilot study, incorporating multifold larger patient databases would lead to the high APS, small atopic clusters expanding.

Overall, cluster analysis for APS yielded several notable clinical findings. We show clusters with higher specific IgE presented to the asthma subspecialty clinic at significantly a younger age. Another finding was the strong positive correlation between APS and the Th2 asthma severity markers, total IgE and the PBEC. Finally, there was a significant negative correlation between the

obstructive severity markers FEV₁/FVC ratio and FEF₂₅₋₇₅ percent predicted with a higher APS score. Surprisingly, the correlation between APS score and FEV₁ percent predicted was not statistically significant, however, we found by elimination, this unexpected result was influenced by the 4 patients in clusters 8 through 10.

Hierarchical clustering indicated three major distinguishable classes of APS superclusters (Fig. 2), which may represent divergent manifestations of the poly-atopic march. The first was made up of patients who tended to have progressively increasing perennial sensitization, while the other tended to have progressively increasing seasonal sensitization. Interestingly, the hierarchical clustering algorithm segregated the one cluster 10 patient profile as representing a third supercluster, though we tend to favor the view that they belong to the seasonal supercluster (see allergen profile, Table 4).

Cluster analyses to date have been heavily weighted on Euro-American Caucasian populations and not on urban black and Hispanic (Latino) patients. However, both a strength and limitation of our study is our demographic makeup. The limited diversity, particularly the low number of Caucasians and the preponderance of middle-aged, female, Afro-Americans from high poverty zip codes, should be considered when extending our findings to the moderate to severe asthma populations at large. The inclusion of diverse populations, particularly Euro-American Caucasians, is needed in future studies of APS. We recognize other limitations. Our study population is medically vulnerable and access to medications may have influenced PFT results. Medication access also confounds the relationships between APS clusters and precise assignment to GINA steps (4 or 5) or ATS/ERS steps (4 through 6). Lastly, this study does not assess how clusters predict exacerbations, emergency department visits, or hospitalizations. Future work is needed to define these relationships.

To date, clustering approaches in asthma have not considered which allergen sensitizations are positive, the numbers of positive allergens, potential interactions of allergens affecting phenotypes, and levels of sensitization. Our work represents the first detailed attempt to define polysensitization and how it may affect and refine asthma cluster analyses. Our clustering approach may lead to intuitive clinical questions including if treatment can be predictably tailored to each cluster, if the seasonal supercluster patients are indeed at an increased risk for seasonal exacerbations, and many others. Our studies show significant heterogeneity in the poly-atopic population, which should be considered in future approaches to clus-

tering. Understanding the importance of APS will also help physicians understand the spectrum of sensitivities that drive patients' symptoms, the poly-atopic march, and possibly optimize treatment strategies.

Acknowledgments

The authors thank Dr. Phillip E. Silkoff for reviewing an early version of this manuscript.

Statement of Ethics

The need for written informed consent was waived by the Institutional Review Board at Drexel University (IRBID: 1803006182), who also reviewed and approved the study design. According to 45 CFR 46.104 (4) (iii), this study is approved as Exempt Category 4.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

Margaret Wolf Pulmonary Research Memorial Endowment Fund contributes to Dr. Schulman's salary. The endowment had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

References

- 1 Dharmage SC, Perret JL, Custovic A. Epidemiology of asthma in children and adults. *Front Pediatr*. 2019 Jun 18;7:246.
- 2 Agache I, Akdis CA. Endotypes of allergic diseases and asthma: an important step in building blocks for the future of precision medicine. *Allergol Int*. 2016;65(3):243–52.
- 3 Kaur R, Chupp G. Phenotypes and endotypes of adult asthma: moving toward precision medicine. *J Allergy Clin Immunol*. 2019 Jul 1; 144(1):1–12.
- 4 Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008 Aug 1;178(3):218–24.
- 5 Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the severe asthma research program. *Am J Respir Crit Care Med*. 2010 Feb 15;181(4):315–23.
- 6 Amelink M, de Groot JC, de Nijs SB, Lutter R, Zwinderman AH, Sterk PJ, et al. Severe adult-onset asthma: a distinct phenotype. *J Allergy Clin Immunol*. 2013 Aug 1;132(2):336–41.
- 7 Kim TB, Jang AS, Kwon HS, Park JS, Chang YS, Cho SH, et al. Identification of asthma clusters in two independent Korean adult asthma cohorts. *Eur Respir J*. 2013 Jun 1; 41(6):1308–14.
- 8 Silkoff PE, Strambu I, Laviolette M, Singh D, FitzGerald JM, Lam S, et al. Asthma characteristics and biomarkers from the airways disease endotyping for personalized therapeutics (ADEPT) longitudinal profiling study. *Respir Res*. 2015 Dec;16(1):142.
- 9 Fitzpatrick AM, Moore WC. Severe asthma phenotypes: how should they guide evaluation and treatment? *J Allergy Clin Immunol Pract*. 2017 Jul 1;5(4):901–8.
- 10 Schatz M, Hsu JWY, Zeiger RS, Chen W, Dorenbaum A, Chipps BE, et al. Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma. *J Allergy Clin Immunol*. 2014 Jun 1;133(6):1549–56.
- 11 Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, et al. Statistical cluster analysis of the British Thoracic Society Severe refractory Asthma Registry: clinical outcomes and phenotype stability. *PLoS One*. 2014 Jul 24;9(7):e102987.
- 12 Zaihra T, Walsh CJ, Ahmed S, Fugere C, Hamid QA, Olivenstein R, et al. Phenotyping of difficult asthma using longitudinal physiological and biomarker measurements reveals significant differences in stability between clusters. *BMC Pulm Med*. 2016 Dec;16(1):74.
- 13 Silkoff PE, Moore WC, Sterk PJ. Three major efforts to phenotype asthma: severe asthma research program, asthma disease endotyping for personalized therapeutics, and unbiased biomarkers for the prediction of respiratory disease outcome. *Clin Chest Med*. 2019 Mar 1;40(1):13–28.
- 14 Deng K, Zhang X, Liu Y, Zhang L, Wang G, Feng M, et al. Heterogeneity of paucigranulocytic asthma: a prospective cohort study with hierarchical cluster analysis. *J Allergy Clin Immunol Pract*. 2021 Jun 1;9(6):2344–55.

Author Contributions

Brian J. Patchett and Edward S. Schulman contributed to the conception of the work; study design; acquisition, analysis, and interpretation of the data; and drafting and revision of the paper. Bede N. Nriagu contributed to the analysis, acquisition, and interpretation of the data and drafting and revision of the paper. Granit Mavraj contributed to the analysis, acquisition, and revision of the paper. Ruchi R. Patel contributed to the analysis and interpretation of the data and revision of the paper. Christopher MacLellan contributed to the study design, analysis and interpretation of the data, and revision of the paper. Tushar Thakur contributed to the acquisition and interpretation of the data and revision of the paper. All the authors approved the final version of the paper and agreed to be accountable for all aspects of the work.

Data Availability Statement

The data that support the findings of this study are not publicly available due to containing information that could compromise the privacy of research participants but are available from the corresponding author upon reasonable request. Further inquiries can be directed to the corresponding author. A preprint version of this article is available on medRxiv: <https://doi.org/10.1101/2022.05.08.22274727>.

- 15 Lefaudeux D, De Meulder B, Loza MJ, Peffer N, Rowe A, Baribaud F, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. *J Allergy Clin Immunol*. 2017 Jun 1;139(6):1797–807.
- 16 Platts-Mills TA. The role of immunoglobulin E in allergy and asthma. *Am J Respir Crit Care Med*. 2001 Oct 15;164(8 Pt 2):S1–5.
- 17 Djukanovic R, Wilson SJ, Kraft M, Jarjour NN, Steel M, Chung KF, et al. Effects of treatment with anti-immunoglobulin E antibody omalizumab on airway inflammation in allergic asthma. *Am J Respir Crit Care Med*. 2004 Sep 15;170(6):583–93.
- 18 Miguères M, Dávila I, Frati F, Azpeitia A, Jeanpetit Y, Lhéritier-Barrand M, et al. Types of sensitization to aeroallergens: definitions, prevalences and impact on the diagnosis and treatment of allergic respiratory disease. *Clin Transl Allergy*. 2014 Dec;4(1):16.
- 19 Ciprandi G, Cirillo I. Monosensitization and polysensitization in allergic rhinitis. *Eur J Intern Med*. 2011 Dec 1;22(6):e75–9.
- 20 Szeinbach SL, Barnes JH, Sullivan TJ, Williams PB. Precision and accuracy of commercial laboratories' ability to classify positive and/or negative allergen-specific IgE results. *Ann Allergy Asthma Immunol*. 2001 Apr 1;86(4):373–81.
- 21 Leimgruber A, Mosimann B, Claeys M, Seppey M, Jaccard Y, Aubert V, et al. Clinical evaluation of a new in-vitro assay for specific IgE, the immuno CAP system. *Clin Exp Allergy*. 1991 Jan;21(1):127–31.
- 22 Schulman ES, Pohlig C. Rationale for specific allergen testing of patients with asthma in the clinical pulmonary office setting. *Chest*. 2015 Jan 1;147(1):251–8.
- 23 Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2014 Feb 1;43(2):343–73.
- 24 Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. Fontana; 2021. Available from: <https://gin-asthma.org/reports/> (Last accessed February 20, 2022).
- 25 González Burchard E, Borrell LN. Need for racial and ethnic diversity in asthma precision medicine. *N Engl J Med*. 2021 Dec 9;385(24):2297–8.
- 26 Semega J, Kollar M, Shrider EA, Creamer JF. *Income and Poverty in the United States: 2019*. U.S. Census Bureau, Current Population Reports; 2020. p. 60–270.
- 27 Data access | City Health Dashboard. Retrieved October 5, 2020. Available from: <https://www.cityhealthdashboard.com/data-access>.
- 28 Cattell RB. The scree test for the number of factors. *Multivariate Behav Res*. 1966 Apr 1;1(2):245–76.
- 29 Bishop CM, Nasrabadi NM. *Pattern recognition and machine learning*. New York: Springer; 2006 Aug 17.
- 30 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011 Nov 1;12:2825–30.
- 31 Blei DM, Jordan MI. Variational inference for Dirichlet process mixtures. *Bayesian Anal*. 2006 Mar;1(1):121–43.
- 32 Attias H. A variational bayesian framework for graphical models. *Adv Neural Inf Process Syst*. 1999;12.
- 33 Pitman J. Combinatorial stochastic processes. *Lect Notes Math*. 2002 Jul 7;1875:7–24.
- 34 Abdi H, Williams LJ. Principal component analysis. *Wires Comput Stat*. 2010 Jul;2(4):433–59.
- 35 Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package “corrplot”. *Statistician*. 2017 Oct 16;56(316):e24.
- 36 Kassambara A, Mundt F. Package ‘factoextra’. Extract and visualize the results of multivariate data analyses. 2017;76(2).
- 37 Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008 Mar 18;25:1–8.
- 38 R Core Team. *R. A language and environment for statistical computing*.
- 39 Wickham H, Chang W, Wickham MH. Package ‘ggplot2’. Create elegant data visualisations using the grammar of graphics. *Version*. 2016;2(1):1–89.
- 40 Wickham H, François R, Henry L, Müller K. dplyr: A grammar of data manipulation. *R Package Version 0.4*. 2015;3:p156.
- 41 Centers for Disease Control and Prevention (CDC) CfDcAp. Asthma prevalence and control characteristics by race/ethnicity: United States, 2002. *MMWR Morb Mortal Wkly Rep*. 2004 Feb 27;53(7):145–8.
- 42 Press VG, Pappalardo AA, Conwell WD, Pincavage AT, Prochaska MH, Arora VM. Interventions to improve outcomes for minority adults with asthma: a systematic review. *J Gen Intern Med*. 2012 Aug;27(8):1001–15.
- 43 Heaney LG, McGarvey LP. Personalised medicine for asthma and chronic obstructive pulmonary disease. *Respiration*. 2017;93(3):153–61.
- 44 Dávila I, Campo P, Cimbollek S, Almonacid Sánchez C, Quirce S, Moreira A, et al. Cluster subanalysis of patients with severe asthma who responded to omalizumab. *J Investig Allergol Clin Immunol*. 2022;32(3):213–5.
- 45 Leung C, Sin DD. Asthma-COPD overlap: what are the important questions? *Chest*. 2022;161(2):330–44.
- 46 Favaro S, Lijoi A, Mena RH, Prünster I. Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior. *J Roy Stat Soc Ser B*. 2009 Nov;71(5):993–1008.
- 47 Teh YW, Jordan MI. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*. 2010 Mar;1:158–207.