

UCSF

UC San Francisco Previously Published Works

Title

Multiview Cluster Analysis Identifies Variable Corticosteroid Response Phenotypes in Severe Asthma

Permalink

<https://escholarship.org/uc/item/60f5b5jd>

Journal

American Journal of Respiratory and Critical Care Medicine, 199(11)

ISSN

1073-449X

Authors

Wu, Wei
Bang, Seojin
Bleecker, Eugene R
et al.

Publication Date

2019-06-01

DOI

10.1164/rccm.201808-1543oc

Peer reviewed

Multiview Cluster Analysis Identifies Variable Corticosteroid Response Phenotypes in Severe Asthma

Wei Wu^{1*†}, Seojin Bang^{1*}, Eugene R. Bleecker², Mario Castro³, Loren Denlinger⁴, Serpil C. Erzurum⁵, John V. Fahy⁶, Anne M. Fitzpatrick⁷, Benjamin M. Gaston⁸, Annette T. Hastie⁹, Elliot Israel^{10,11}, Nizar N. Jarjour⁴, Bruce D. Levy^{10,11}, David T. Mauger¹², Deborah A. Meyers², Wendy C. Moore⁹, Michael Peters⁶, Brenda R. Phillips¹², Wanda Phipatanakul^{11,13}, Ronald L. Sorkness⁴, and Sally E. Wenzel^{14‡}

¹Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania; ²Department of Medicine, University of Arizona, Tucson, Arizona; ³Washington University, St. Louis, Missouri; ⁴University of Wisconsin-Madison, Madison, Wisconsin; ⁵Cleveland Clinic, Cleveland, Ohio; ⁶University of California San Francisco, San Francisco, California; ⁷Emory University, Atlanta, Georgia; ⁸School of Medicine, Case Western Reserve University, Cleveland, Ohio; ⁹School of Medicine, Wake Forest University, Winston-Salem, North Carolina; ¹⁰Harvard Medical School, Boston, Massachusetts; ¹¹Brigham and Women's Hospital, Boston, Massachusetts; ¹²Pennsylvania State University, University Park, Pennsylvania; ¹³Boston Children's Hospital, Boston, Massachusetts; and ¹⁴Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Abstract

Rationale: Corticosteroids (CSs) are the most effective asthma therapy, but responses are heterogeneous and systemic CSs lead to long-term side effects. Therefore, an improved understanding of the contributing factors in CS responses could enhance precision management. Although several factors have been associated with CS responsiveness, no integrated/cluster approach has yet been undertaken to identify differential CS responses.

Objectives: To identify asthma subphenotypes with differential responses to CS treatment using an unsupervised multiview learning approach.

Methods: Multiple-kernel *k*-means clustering was applied to 100 clinical, physiological, inflammatory, and demographic variables from 346 adult participants with asthma in the Severe Asthma Research Program with paired (before and 2–3 weeks after triamcinolone administration) sputum data. Machine-learning techniques were used to select the top baseline variables that predicted cluster assignment for a new patient.

Measurements and Main Results: Multiple-kernel clustering revealed four clusters of individuals with asthma and different CS responses. Clusters 1 and 2 consisted of young, modestly CS-responsive individuals with allergic asthma and relatively normal lung function, separated by contrasting sputum neutrophil and macrophage percentages after CS treatment. The subjects in cluster 3 had late-onset asthma and low lung function, high baseline eosinophilia, and the greatest CS responsiveness. Cluster 4 consisted primarily of young, obese females with severe airflow limitation, little eosinophilic inflammation, and the least CS responsiveness. The top 12 baseline variables were identified, and the clusters were validated using an independent Severe Asthma Research Program test set.

Conclusions: Our machine learning–based approaches provide new insights into the mechanisms of CS responsiveness in asthma, with the potential to improve disease treatment.

Keywords: asthma phenotype; corticosteroids; severe asthma; eosinophils

(Received in original form August 20, 2018; accepted in final form January 23, 2019)

*Co–first authors.

‡Co–senior authors.

Supported by grants from the NIH (P30DA035778 and R01GM114311 to W.W. [co–principal investigator (co-PI)]) and grants that were awarded by the NHLBI to the Severe Asthma Research Program Pls, Clinical Centers, and Data Coordinating Center as follows: Wake Forest University (E.R.B. and D.A.M.) and Emory University (U10 HL109164, A.M.F. subaward PI); Washington University (U10 HL109257, M.C.); University of California San Francisco (U10 HL109146, J.V.F.); Case Western Reserve University (U10 HL109250, B.M.G.); Cleveland Clinic (U10 HL109250, S.C.E. co-PI, Virginia-Cleveland Consortium); Brigham and Women's Hospital, Harvard Medical School (E.I. and B.D.L.) and Boston Children's Hospital, Harvard Medical School (U10 HL109172, W.P. subaward PI); University of Wisconsin (U10 HL109168, N.N.J.); University of Pittsburgh (U10 HL109152, S.E.W.); and Penn State University (U10 HL109086, Data Coordinating Center, D.T.M.). In addition, this program is supported through the following NIH National Center for Advancing Translational Sciences awards: UL1 TR001420 to Wake Forest University, UL1 TR000427 to the University of Wisconsin, UL1 TR001102 to Harvard University, UL1 TR000454 to Emory University.

Author Contributions: Conception and design: W.W. and S.E.W. Data analysis: W.W., S.B., and S.E.W. Data collection: E.R.B., M.C., L.D., S.C.E., J.V.F., A.M.F., B.M.G., A.T.H., E.I., N.N.J., B.D.L., D.T.M., D.A.M., W.C.M., M.P., W.P., R.L.S., and S.E.W. Manuscript writing committee: W.W., S.B., and S.E.W. Manuscript review and editing: W.W., S.B., E.R.B., M.C., L.D., S.C.E., J.V.F., A.M.F., B.M.G., A.T.H., E.I., N.N.J., B.D.L., D.T.M., D.A.M., W.C.M., M.P., B.R.P., W.P., R.L.S., and S.E.W.

Am J Respir Crit Care Med Vol 199, Iss 11, pp 1358–1367, Jun 1, 2019

Copyright © 2019 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201808-1543OC on January 25, 2019

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Corticosteroids (CSs) are the most effective therapy for asthma. However, heterogeneity in CS responsiveness and long-term side effects make it more difficult to control. Although few predictors of response have been identified, no cluster analysis has been performed to identify differential response patterns.

What This Study Adds to the

Field: Multiple-kernel clustering analysis was used to cluster 346 adult subjects with asthma and paired (before and after CS) sputum data using 100 variables, including both baseline and “changes” of dynamic variables. We identified four clusters with differential CS response patterns and baseline characteristics, among which cluster 3 was the most responsive and cluster 4 was the least responsive. We also identified 12 predictive baseline variables that predicted cluster assignment with high accuracy. These findings suggest that CS responses follow clinical, inflammatory, and physiological patterns. The 12 predictive variables we identified suggest that software can be developed to predict responses to CSs, which would help make precision medicine possible. These machine-learning approaches provide novel insights into CS response patterns that could improve asthma management.

Asthma is a heterogeneous chronic airway disorder that consists of multiple phenotypes with diverse clinical characteristics (1, 2). Corticosteroids (CSs), particularly systemic ones, are the most effective asthma controller therapy, but they have numerous harmful side effects (3–6) and the responses are heterogeneous and difficult to predict (1, 7). Previous works investigated a small number of targeted predictors of CS responses (8); however, no large-scale, multivisit analyses

have been performed using machine-learning approaches.

Cluster analysis has identified subphenotypes of asthma using large clinical data sets (9–11). However, no unsupervised learning approach has been undertaken to identify differential CS response patterns among subjects with asthma by integrating dynamic variables measured before and 2–3 weeks after CS treatment. Unfortunately, unsupervised learning approaches that are commonly used to identify asthma subphenotypes, such as *k*-means and hierarchical clustering, are insufficient for dealing with a complex mix of clinical and biological data containing both static and changing/dynamic variables. In addition, traditional clustering gives equal weights to all variables (10, 11), when, in fact, prior knowledge supports a stronger relationship between some variables and asthma and its outcomes than others (for instance, age at onset vs. family history of asthma-allergies).

Multiview learning methods have been developed to classify or cluster samples by integrating different types of data in the biological field (12, 13). Multiple-kernel *k*-means clustering (MKKC)-based methods (Figure E1 in the online supplement) (14, 15), which are classified as unsupervised multiview learning methods, have demonstrated advantages over traditional single-view clustering approaches, such as *k*-means and hierarchical clustering, in that they find clusters by using information collected from different types or sources of data. For example, they have been used to identify cancer subtypes using different omics data, including DNA copy number profiling, mRNA gene expression, and DNA methylation data (14).

To help understand differential CS responsiveness among subjects with asthma, we developed a novel multiview learning strategy that allows us to identify clusters of subjects with asthma subjects and differential patterns of response to CS by 1) incorporating different types of variables, including both baseline and change variables, into the cluster analysis; and 2) assigning variables to different views based on their clinical importance. Toward that end, we recently developed a new multiple-kernel clustering approach, called

MML-MKCC, which finds clusters by using a minimax formulation and l_2 regularization (15). This MML-MKCC approach was applied to a rigorously characterized cohort of adults with asthma from the NIH NHLBI's Severe Asthma Research Program (SARP) who were studied before and after a standardized systemic CS treatment to characterize their responses (8). Seventy static baseline variables, as well as 15 “dynamic” baseline variables and their “changes” in response to CS treatment, were included. Top relevant and nonredundant variables were then selected from the 85 baseline variables using feature selection techniques, to enable clinicians to eventually predict patient responses using a support vector machine (SVM). Our results provide new insights into CS response patterns in asthma.

Methods

Participants

All of the participants were from the SARP cohort. Severe asthma was defined according to criteria established by the European Respiratory Society (ERS) and American Thoracic Society (ATS) (16). All other subjects were considered to have nonsevere asthma. The subjects in this work were limited to 346 adults with paired (before and 2–3 wk after intramuscular triamcinolone 40 mg) sputum cell counts.

Cluster Analysis Using MKKC

Variables with $\geq 10\%$ of missing values were excluded. Others were imputed as described in Reference 11. A total of 100 variables measured from the 346 participants (also called the discovery set) were used in the cluster analysis (Table E1); 85 of these were baseline variables (70 “static” baseline variables whose values were measured only before CS [baseline] and 15 “dynamic” variables whose values were measured both before [baseline] and after CS). Fifteen “change” variables were created by subtracting the baseline values of the dynamic variables from the values of the corresponding variables after CS, and were also included in the cluster analysis.

Correspondence and requests for reprints should be addressed to Wei Wu, Ph.D., Computational Biology Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: weiwu2@cs.cmu.edu

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

MKCC. To identify clusters of subjects with asthma who showed differential responses to CS, we clustered 346 subjects using our recently developed MKCC methodology, MML-MKCC (15). We assigned variables to three different groups (also called “views” in the machine-learning literature) based on their clinical importance according to prior knowledge and our previous cluster analysis (11) (Table E1).

Assigning variables to three different views. View 1 included 27 static baseline variables, which have looser ties to asthma pathobiology, including household socioeconomic information, as well as comorbid conditions such as diabetes and depression. View 2 was composed of 53 baseline variables containing 38 “more important” static variables, including asthma clinical questionnaires, vital signs, Asthma Quality of Life Questionnaire (AQLQ), family history of allergy, IgE, and biological features such as inflammatory cell counts, which our previous study suggested were discriminatory in asthma clusters (11), as well as 15 dynamic variables, including Asthma Control Questionnaire (ACQ) scores, fractional exhaled nitric oxide (FE_{NO}), and sputum cell counts/differentials. View 3 contained the “changes” of the 15 dynamic variables (from view 2) after CS treatment, as well as five demographic variables with importance for asthma, including age of onset, age at baseline, sex, race, and body mass index (BMI) (9, 11, 17, 18).

Because the variables in views 1–3 had increasing clinical importance, constraints on the weights of the views in the cluster analysis were set so that the least weight was put on view 1, and the most was put on view 3. An optimal number of the clusters k was determined using the elbow method (Figure E2). A principal component analysis was performed to illustrate the identified clusters (15).

Stability analysis. A stability analysis was performed to determine the stability of the identified clusters when variables were assigned to different views.

Statistical Tests

To examine significant differences of the variables among the identified clusters, we performed one-way ANOVAs (19) for continuous variables, Kruskal-Wallis rank-sum tests (20) for ordinal variables, and Pearson’s chi-squared tests (21) for

nominal variables. To examine whether a variable was significantly different between any pairs of the clusters, we performed two-sample pairwise t tests for continuous variables, Wilcoxon rank-sum tests (22) for ordinal variables, and Pearson’s chi-squared tests (21) for nominal variables. P values were adjusted to control the false discovery rate using the Benjamini-Hochberg procedure (23). P values < 0.05 were considered significant.

Predicting Clustered Subjects Using the Most Informative Baseline Variables

The top predictive baseline variables were selected for prediction using a two-step feature selection machine-learning procedure (11). To classify the subjects whose cluster labels were identified by MML-MKCC, we used a multiclass SVM algorithm with a 10-fold cross-validation strategy to determine the top relevant nonredundant baseline variables. Then, SVM classification and a 10-fold cross-validation with different splits of the samples were used to evaluate how well these variables predicted cluster labels of the test samples.

Validation of the Newly Identified Clusters Using an Independent SARP Test Set

An independent test set of 182 adult SARP participants without sputum data was used to validate/replicate the clusters identified by MML-MKCC. First, a multiclass SVM classifier was trained using the 346 patients in the discovery set with the 12 predictive baseline variables, of which two were surrogate variables [blood neutrophil and eosinophil counts] to replace sputum macrophage and eosinophil percentages because the test samples had no available sputum data. Cluster labels of the participants in the test set were predicted using the trained SVM classifier. Finally, the clusters of the patients in the test data were characterized using the statistical tests described above.

Details regarding the methods used in this work are provided in the online supplement.

Results

Demographics of the Participants Used in the Cluster Analysis

A total of 346 participants (≥ 18 yr old) with asthma and paired sputum data were

analyzed. Of these 346, 204 met the ERS/ATS criteria for severe asthma (16). The demographics of the 346 participants did not differ from those of the complete cohort of 528 adult participants with asthma (Table 1).

Identification of CS Response Patterns Using a Novel Multiview Learning Strategy

One hundred clinical, physiological, inflammatory, and demographic variables were included in the cluster analysis. To identify CS response patterns, we developed a novel multiview learning strategy that allows clustering of participants using our recently developed MKCC algorithm (15), taking into account the “change” values of the dynamic variables as well as the baseline values of both the static and dynamic variables.

Clustering Results from a Multiple-Kernel k -Means Approach

Clustering 346 participants with 100 variables using our MKCC algorithm revealed four distinct asthma clusters with differing CS responses (Table 2), as the elbow method determined four clusters as the optimal cluster number identified by MML-MKCC (Figure E2). The clusters were well separated from one another, as shown by a multiple-kernel principal component analysis plots (Figure E3). The four participant clusters had distinct baseline and patterns of response to the CS treatments, as detailed in Figure 1. Summaries of the variables that differed significantly among the clusters, as measured by traditional statistical tests, can be found in Tables E2 and E3.

Cluster 1. At baseline, cluster 1 participants ($n = 81$) were relatively asymptomatic but highly allergic, with the earliest age at onset, and evenly mixed between males and females (Figures 2 and 3). They had normal lung function and the highest sputum neutrophil percentages, but low neutrophil and eosinophil numbers in blood, and the lowest medication use and urgent healthcare use (Figures 3–5, E5, and E6). Despite this being the healthiest cluster, 38% of the participants had severe asthma as defined by the ERS/ATS (Table 2).

After triamcinolone treatment, given their relatively normal baselines, the subjects in this cluster showed only small improvements in FEV_1 , FVC, or FE_{NO} (Figures 4D and 4F, and Table E2). After

Table 1. Demographics of the Adult Subjects with Asthma in the Entire SARP Cohort and the Subjects with Paired Sputum Samples Used in this Analysis

	Subjects in the SARP Cohort	Subjects for Cluster Analysis	P Value (FDR)*
Sample size	528	346	
Age	49.4 (37.1–57.9)	49.3 (37.0–57.9)	0.73
Sex, %, F/M	67/33	67/33	0.99
BMI	30.9 (26.5–36.9)	30.9 (26.5–37.0)	0.91
Black/African American, %	27	24	0.88
Age at onset	12.0 (5.0–28.0)	12 (5.0–27.0)	0.91
Baseline pre-BD FVC% predicted	83.7 (73.1–97.7)	84.1 (74.7–97.3)	0.47
Maximal FEV ₁ % predicted	84.8 (71.5–97.8)	85.5 (72.9–97.6)	0.35
Baseline pre-BD FEV ₁ /FVC% predicted	85.8 (75.8–93.5)	86 (77.1–92.9)	0.67
Severe asthma, %	64	64	1.00

Definition of abbreviations: BD = bronchodilator; BMI = body mass index; FDR = false discovery rate; SARP = Severe Asthma Research Program. Numerical data are presented as median (first–third quartiles).

*FDR-adjusted *P* value from Welch's *t* test or chi-square test among the 528 subjects in the entire SARP cohort and the 346 subjects used for the cluster analysis.

CS treatment, this cluster had the largest increase in macrophage percentages (Figure 3D).

Cluster 2. The patients in cluster 2 (*n* = 73) were also young, mostly women, and allergic, but they had a slightly older age at onset than those in cluster 1 (Figures 2 and 3). Their lung function was modestly lower and more reversible compared with the patients in cluster 1 (Figure 4), but a larger percentage of these patients were on high-dose inhaled CS (ICS) and other controllers. Perhaps because they had received more treatment, they had low baseline T2 biomarkers (second lowest

blood eosinophil counts, and lowest sputum eosinophil and neutrophil percentages), low symptoms, and a relatively good quality of life, with low urgent healthcare use (Figures 4 and 5). Thus, the participants in this cluster, 55% of whom met the ERS/ATS definition of severe asthma, generally had severe but well-controlled asthma (Table 2). They also had the lowest sputum neutrophil and highest sputum macrophage percentages (Figures 3A and 3C).

Similar to what was observed in cluster 1, there was little change in sputum eosinophil percentages or lung function

after triamcinolone treatment, probably because the subjects had relatively normal baseline values (Figures 4C–4F). However, in contrast to cluster 1, this cluster was characterized by the highest decrease in sputum macrophage percentages (and reciprocal increase in neutrophil percentages) after triamcinolone treatment (Figures 3A–3D).

Cluster 3. Patients in cluster 3 (*n* = 96) were the oldest and least allergic, with the latest age at onset (Figures 2 and 3). They were generally better educated and had the lowest lung function and low reversibility (Figures 4C, 4E, and 4F). The participants

Table 2. Demographic Data of the Subjects in the Clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P Value (FDR)*
Sample size	81	73	96	96	
Age	42.3 (32.1 to 50.5)	42.1 (33.9 to 52.3)	60.5 (54.3 to 66.0)	45.1 (33.7 to 52.2)	<0.0001
Sex, %, F/M	48/52	70/30	67/33	82/18	<0.0001
BMI	29.2 (25.8 to 32.2)	30.6 (26.5 to 35.3)	28.1 (25.6 to 33.6)	37.7 (31.6 to 44.9)	<0.0001
Black/African American, %	20	25	7	45	<0.0001
Age at onset	8.0 (3.0 to 16.0)	11.1 (5.0 to 22.0)	30.0 (14.9 to 40.1)	8.5 (3.0 to 19.0)	<0.0001
Baseline pre-BD FVC% predicted	95.9 (87.7 to 104.7)	86.9 (79.5 to 101.0)	78.2 (70.9 to 85.2)	79.8 (72.0 to 91.1)	<0.0001
Maximal FEV ₁ % predicted	91.7 (83.0 to 101.4)	89.6 (79.2 to 105.7)	75.8 (63.6 to 86.2)	84.9 (73.6 to 94.8)	<0.0001
Baseline pre-BD FEV ₁ /FVC% predicted	86.9 (80.2 to 94.0)	86.5 (77.8 to 92.4)	83.7 (73.5 to 91.0)	86.5 (76.3 to 92.4)	1.70 × 10 ⁻¹
Change in pre-BD FVC% predicted	-0.5 (-3.3 to 3.8)	1 (-2.6 to 5.3)	3.6 (-0.6 to 7.6)	1.5 (-1.5 to 5.4)	4.73 × 10 ⁻⁴
Change in maximal FEV ₁ % predicted	0.4 (-2.3 to 3.1)	0 (-2.8 to 3.5)	3.7 (-1.0 to 8.1)	-0.7 (-3.6 to 3.5)	<0.0001
Change in pre-BD FEV ₁ /FVC% predicted	1.3 (-1.7 to 3.1)	0.5 (-1.2 to 3.3)	1.4 (-1.3 to 4.7)	1.1 (-2.1 to 4.8)	3.89 × 10 ⁻¹
Severe asthma, %	38	55	77	79	<0.0001

Definition of abbreviations: BD = bronchodilator; BMI = body mass index; FDR = false discovery rate. Numerical data are presented as median (first to third quartiles).

*FDR-adjusted *P* value from the ANOVA or chi-square test among the 528 subjects of the total Severe Asthma Research Program cohort and the 346 subjects used for the cluster analysis.

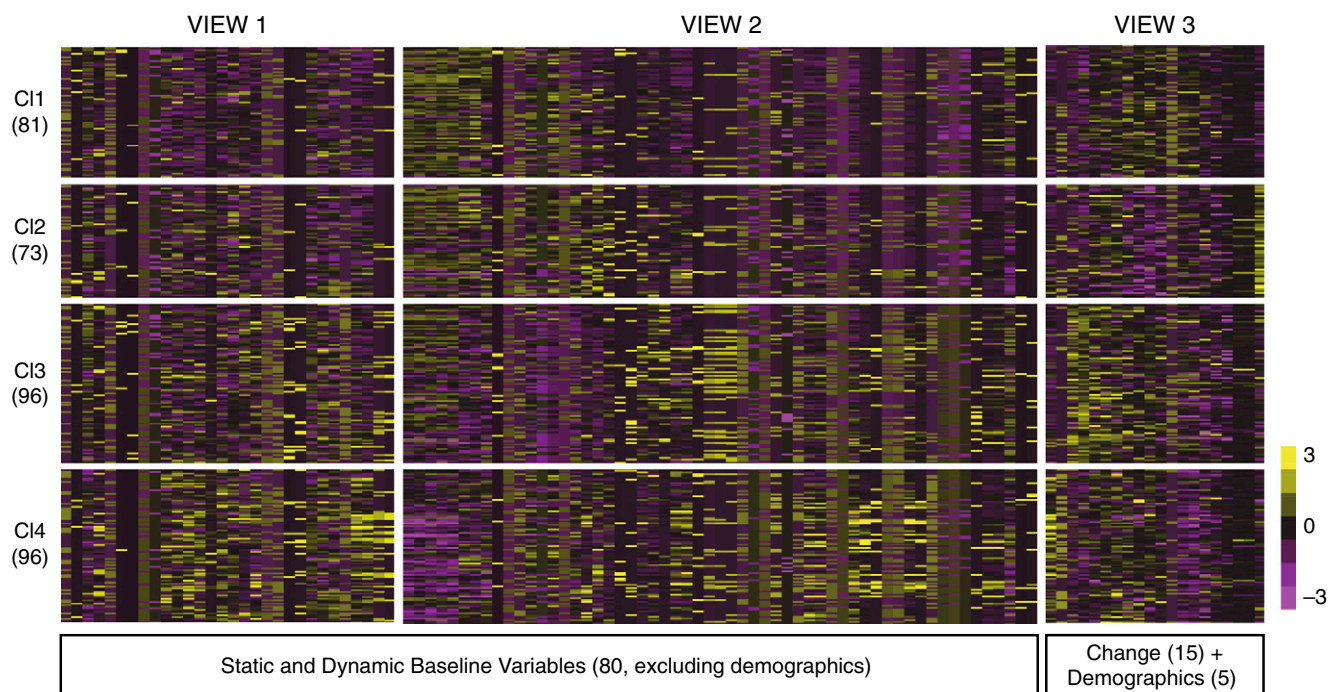


Figure 1. Heatmap of the four clusters identified among 346 patients by the multiple-kernel k -means clustering method. Rows represent the patients, and columns represent the variables in each view (for an ordered list of variables, see Table E1).

in cluster 3 were characterized by high lung and blood T2 biomarkers (highest sputum and blood eosinophil percentages/counts and FE_{NO}), the most gastroesophageal reflux disease, high blood pressure, and the highest percentage of nasal polyps (33%) and sinus disease (Figure 4A and Table E2). Although they reported low asthma symptoms, they had the second highest exacerbations and oral CS use, and 77% had severe asthma (Table 2). To support the relationship with T2 inflammation, we compared T2 gene mean (T2GM) RNA levels in a *post hoc* analysis (see Figure E7). In support of these higher eosinophil levels, the *post hoc* analysis of a subgroup with RNA from sputum showed that patients in this cluster had the highest baseline T2GM (Figure E7A).

Perhaps related to their lower lung function and high T2 biomarkers, the participants in this cluster had the biggest improvements in lung function and sputum eosinophils, accompanied by the second biggest improvement in symptoms (Figures 4B, 4D, and 4F). Additional small improvements in FEV_1 and FVC after albuterol after triamcinolone (Table E4) suggest that triamcinolone did not completely correct the airflow limitation. Despite these improvements, T2 biomarker

levels were still higher than in other clusters after the CS injection.

Cluster 4. Cluster 4 participants ($n = 96$) were primarily young, obese females with early-onset asthma, 45% of whom were black/African Americans (Figure 2). They reported the worst asthma control, with the most symptoms and severe exacerbations, and low but highly reversible lung function (Figures 5A and 5E). They had low T2 biomarkers (the lowest FE_{NO} , low sputum eosinophils, and moderate elevations in blood eosinophils) but more comorbidities, including diabetes, anxiety, depression, and sleep disorders than all other clusters (Figure 4A and Table E2). These patients reported the highest use of biologics and oral CSs, and had high blood neutrophils; 79% of the participants in this cluster had severe asthma (Table 2).

After triamcinolone treatment, these patients had only a small decrease in sputum eosinophil percentages (Figure 4B). These improvements were not accompanied by improvements in ACQ scores, FEV_1 , or FVC. In fact, after triamcinolone treatment, the bronchodilator (BD) response (percent change) in both FEV_1 and FVC decreased compared with the pretreatment baseline (Figure 5F and Table E4). This

decrease in BD response occurred even in the absence of any CS-induced improvement in pre-BD FEV_1 or FVC (Table E4). This suggests that high-dose systemic CSs may have contributed to worsening the BD response.

Prediction of the Identified Clusters Using Baseline Variables

We developed a classification strategy with SVM that enables prediction of the cluster label (and subsequent CS response) for a given patient using only the baseline variables. Using a 10-fold cross-validation, we identified the top 12 baseline variables with an INFOGAIN filter and the Markov blanket algorithm (Table E5). These top 12 variables included (in order) age, activity limitation AQLQ, age at onset, BMI, baseline pre-BD FVC% predicted, sputum macrophage percentages, number of specific IgE, FVC albuterol response, baseline pulse, total white blood cell count, black/African American racial background, and sputum eosinophil percentages.

Using these 12 baseline variables and SVM with 10-fold cross-validation, we predicted the cluster labels of the test samples (10% of the participants in the discovery set) with 81% overall accuracy,

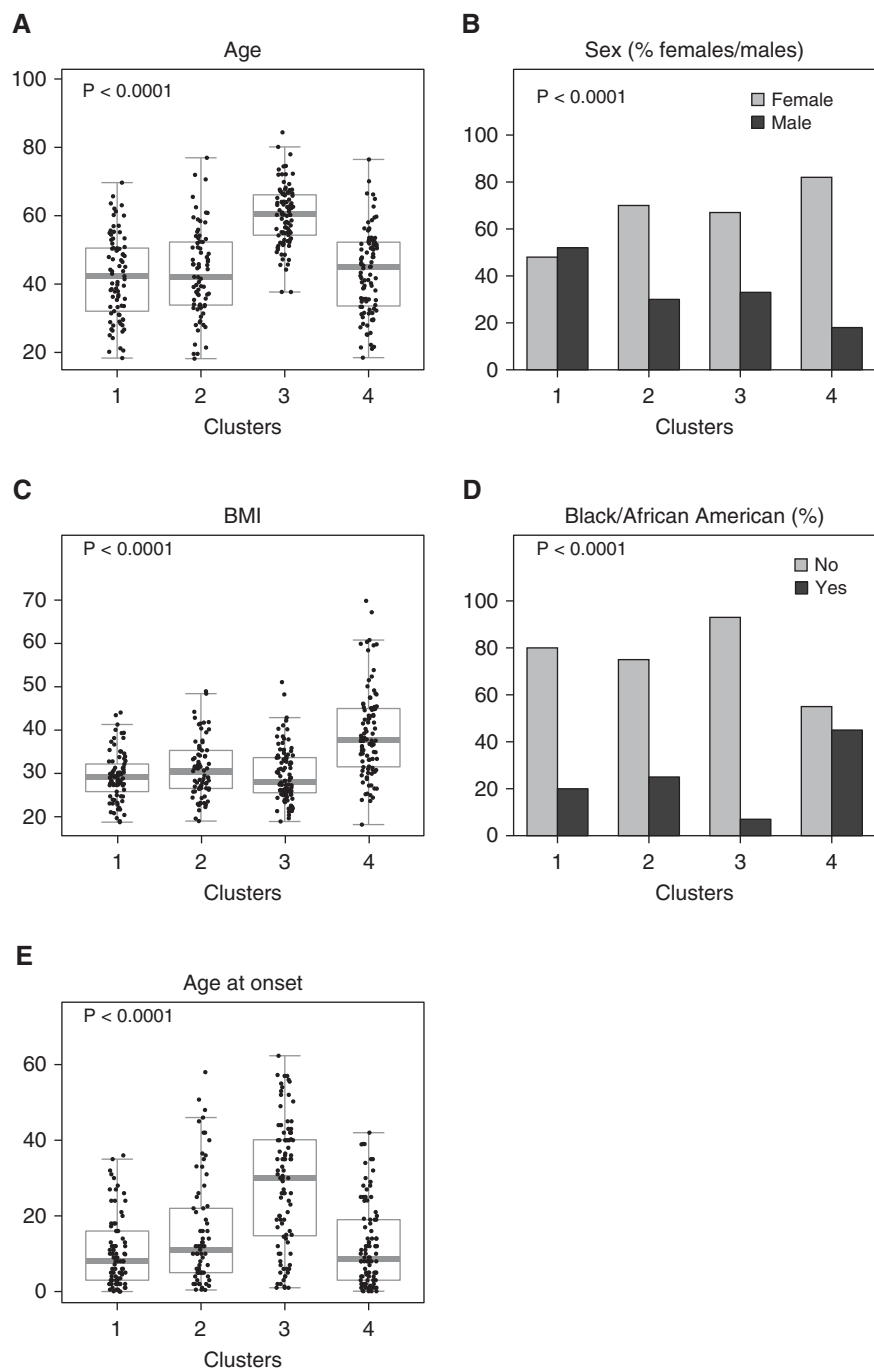


Figure 2. Five demographic variables differed by cluster. (A and E) Cluster 3 had the oldest subjects (A) with the latest age at onset (E). (B) Cluster 1 had the largest percentage of males, and cluster 4 had the lowest. (C) Cluster 4 had the highest body mass index (BMI). (D) Cluster 4 had the highest percentage of black/African Americans, and cluster 3 had the lowest percentage.

62% sensitivity, and 87% specificity (see the online supplement). For individual clusters, the respective sensitivity for clusters 1–4 was 64%, 32%, 77%, and 70%, and specificity was 87%, 89%, 90%, and 84%.

Validation of the Newly Identified Clusters Using an Independent SARP Test Set

The 12 slightly modified baseline variables were used to predict cluster labels of the 182 SARP participants in the test set with SVM

(see the online supplement). The demographics of the 182 participants did not differ from those of the participants in the discovery set (Table E6). The participants clustered similarly to those in the discovery set. Clusters 1 ($n = 52$) and 2 ($n = 35$) included subjects with mild allergic asthma and early age at onset. Patients in cluster 3 ($n = 57$) were older, with the latest age at onset, the most nasal polyps, and high blood eosinophils, and after CS treatment, their lung function increased the most. In addition, the small subsequent increases in FEV₁ and FVC after albuterol after CS treatment were also observed in this test set (Table E4).

Cluster 4 ($n = 38$) included primarily obese females with early-onset asthma, about half of whom were black/African Americans. After CS treatment, there was minimal improvement in lung function, with no incremental benefit of triamcinolone on maximal lung function compared with that of albuterol alone (Table E4).

Together, these results suggest that the top 12 baseline variables and an SVM classifier trained using patients with known cluster labels could eventually allow clinical prediction of cluster assignment and response to systemic CSs.

Discussion

To investigate differential CS responses among participants with asthma, and the various factors that contribute to them, we analyzed 70 demographic, clinical, physiological, and inflammatory variables at baseline, as well as an additional 15 variables both before and 2–3 weeks after an injection of triamcinolone. For this purpose, we used our newly developed multiple-kernel k -means approach, MML-MKCC (15), which integrates data with distinct features and allows for evaluation of changes over time or with treatment. This approach identified four distinct patient clusters with variable CS responsiveness. Using feature selection techniques, we identified the top 12 predictive baseline variables, including easily identifiable variables such as age and age at disease onset, BMI, and baseline lung function. Using SVM and the 12 variables with the two surrogate variables, mandated by the absence of sputum data, the test set clusters replicated the discovery set clusters.

Our study has similarities and differences in comparison with an earlier,

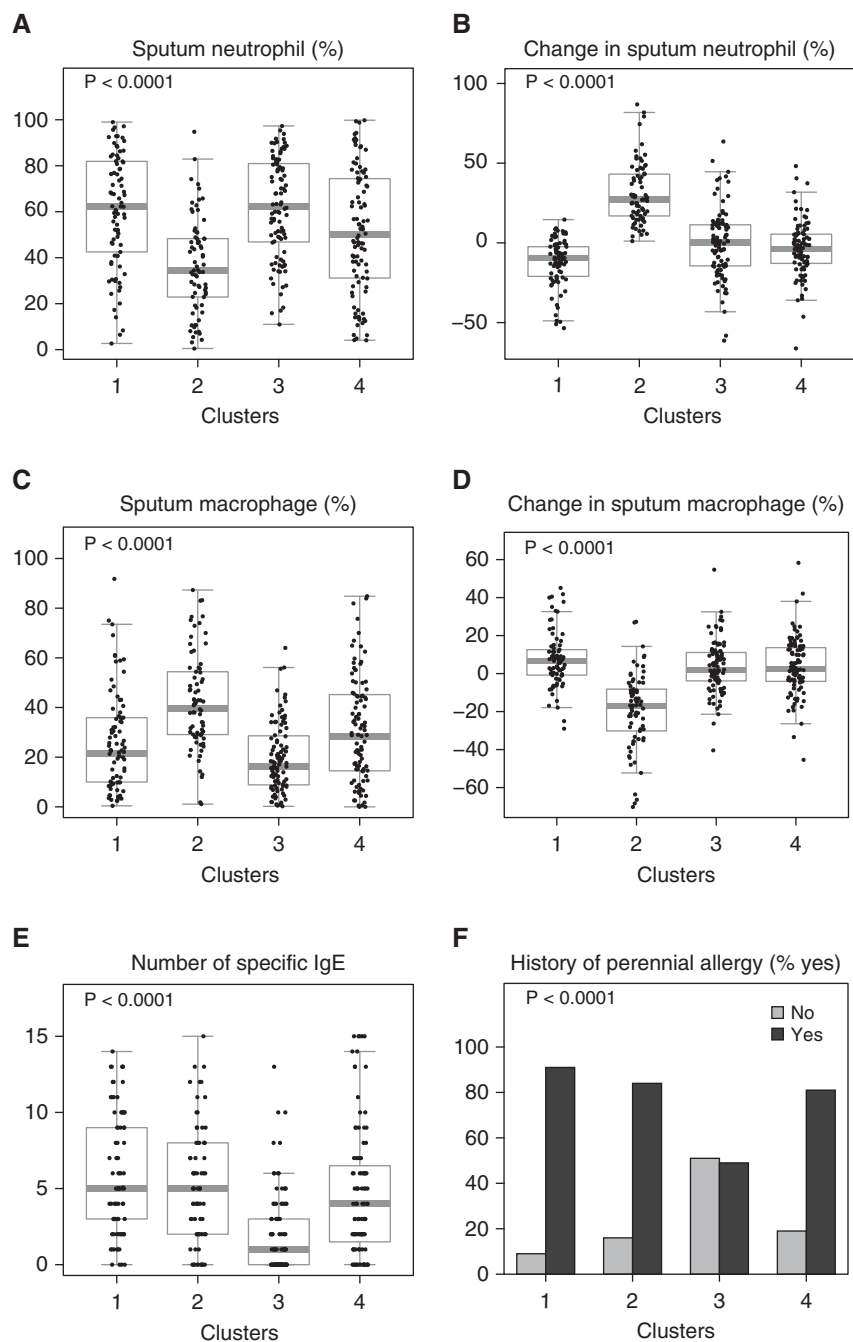


Figure 3. Distinguishing features of clusters 1 and 2. (A–D) At baseline, clusters 1 and 2 had opposite sputum neutrophil (A) and macrophage (C) percentages and reciprocal changes after triamcinolone treatment (B and D). (E and F) They also had the highest number of specific IgE (E) and the highest percentage of patients with perennial allergy (F).

more directed study of CS responses in these patients from the SARP cohort (8). In that study, responses to triamcinolone were assessed with traditional generalized linear mixed effects models followed by receiver operating characteristic curve analyses to identify the best predictors of CS response

(specifically defined as a 10% improvement in FEV₁ after triamcinolone). Using the much smaller number of variables in the models, baseline BD response and FE_{NO} were identified as predictors of response. Blood and sputum eosinophils were only slightly worse predictors, but BMI and race

were identified as poor predictors. Rather than focusing on a limited number of baseline variables, in the current study we identified subphenotypes of asthma encompassing many baseline variables and, for a large number of variables, their change over time. Thus, these results integrate both types of variables, greatly expand on the original analysis, and reveal novel new relationships.

There were two major challenges to our approach. The first was to incorporate both baseline and “change” variables in one analysis, and the second was to combine different types of baseline variables that affect asthma in a single analysis. To address these issues, we used our multiple-kernel *k*-means approach, MML-MKCC, to 1) incorporate both static and dynamic baseline variables, as well as the dynamic “change” variables, into a single cluster analysis, and 2) assign variables to different views based on their clinical importance for asthma according to general prior knowledge and our previous cluster analysis. This MML-MKCC approach identified four clusters of patients who exhibited not only distinct CS response patterns but also distinct baseline characteristics.

Of the four clusters identified, only one cluster, cluster 3 (28% of the overall population), would be widely recognized as highly CS responsive. Not surprisingly, the patients in this cluster had the highest baseline eosinophilia, markedly obstructed lung function, and the most nasal polyps. They also had the highest values of T2GM by *post hoc* analysis. Despite their significantly older age and later age at onset, they had the greatest improvements in obstruction and inflammation. Although they had small improvements in lung function after albuterol treatment, their ACQ6 scores did not improve with triamcinolone. This cluster was previously identified in our cross-sectional cluster analysis (in a different population) and is well recognized clinically (11, 24, 25). In contrast, clusters 1 and 2 were only modestly CS responsive and showed no improvement in ACQ6 scores, most likely because the subjects in these clusters had near-normal baselines. It is impossible to predict how responsive these patients might be during an acute episode of worsening, when their lung function would likely be lower. On the other hand, this analysis identified the participants in

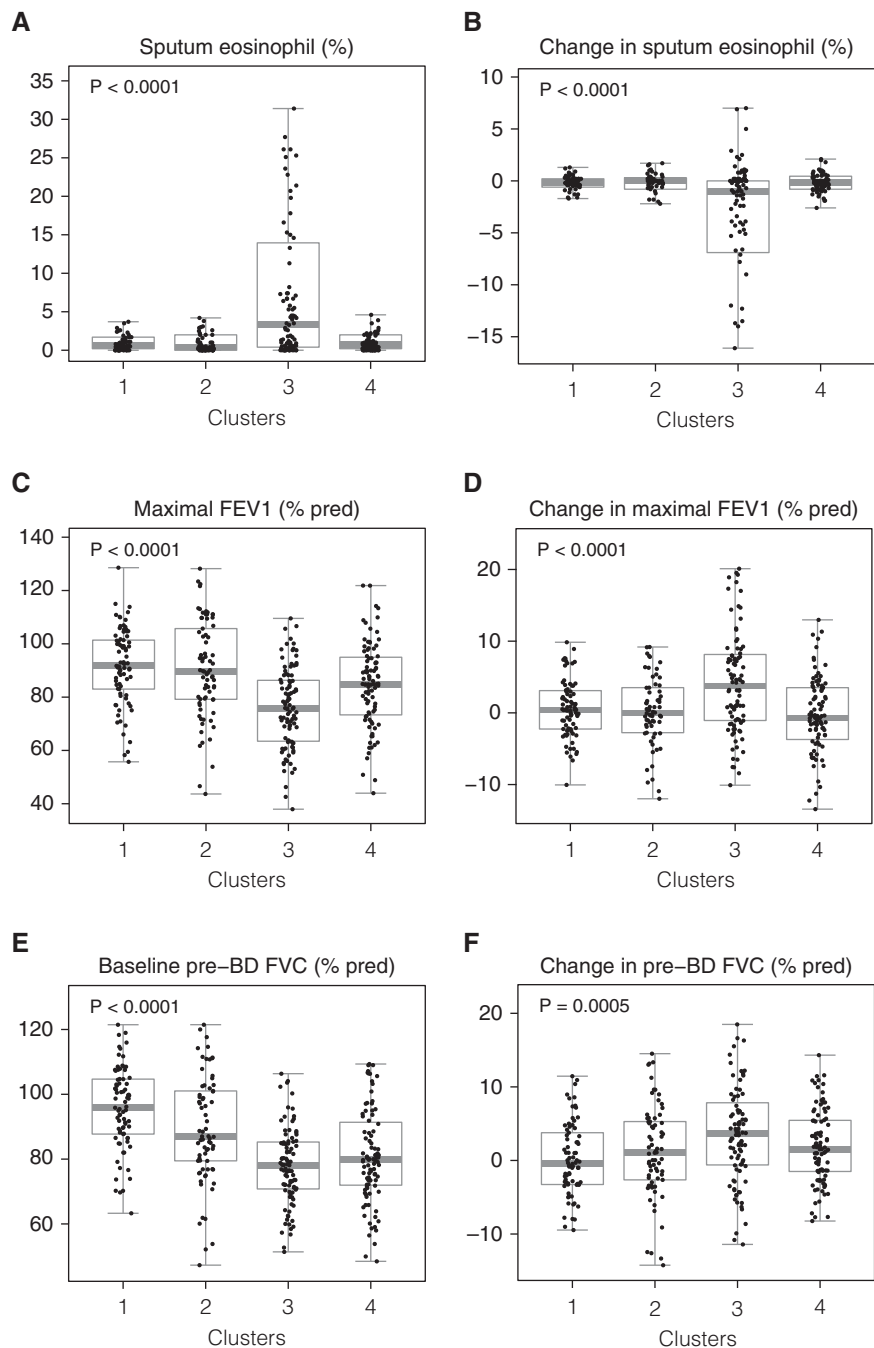


Figure 4. Distinguishing features of cluster 3. (A, C, and E) At baseline, cluster 3 had the highest sputum eosinophil percentages (A) and lowest lung function (C and E). (B, D, and F) After triamcinolone treatment, cluster 3 had the greatest decrease in sputum eosinophil percentages (B) and the largest improvement in lung function (D and F). The box plots in this figure are shown without outliers. See Figure E5 for box plots showing all data points for the same variables. BD = bronchodilator.

cluster 4 as the least CS responsive, even though they had rather severe baseline airflow limitations. After CS treatment, they were still worse than the subjects in the other clusters, with almost no change

in lung function, and even a small decrease in maximal (after albuterol) lung function, particularly in the change in FVC. Although the explanation for this is uncertain, this finding was consistent in

cluster 4 for both the discovery and test sets. Thus, the CS responses in cluster 4 are likely complex. In these patients, CS could detrimentally “stiffen the lungs/airways” to decrease BD responses. Although it is not typically thought of in relation to asthma, CSs can increase the stiffness of the extracellular matrix, particularly in the eye in relation to CS-induced glaucoma (26–28). A similar effect could occur in the airway matrix, which could decrease BD responses in susceptible patients. Although further confirmation is needed, the implications of this finding could be important.

The most striking differences between clusters 1 and 2 were the nearly opposite and marked differences in macrophage and neutrophil percentages in sputum at baseline and their reciprocal changes after CS treatment. Cluster 1, with low starting sputum macrophage percentages, had a large increase after CS treatment, and cluster 2, with low neutrophil percentages, had the largest macrophage increase after CS treatment. In each case, and not surprisingly, as neutrophils and macrophages make up the vast majority of sputum cell types, there were decreases in the “other” cell type. In the INFOGAIN approach, these changes in sputum neutrophil percentages and macrophage percentages had the highest and third highest INFOGAIN values, suggesting that these two variables are highly discriminatory in cluster separation. Despite these marked changes, clinically and physiologically perceivable changes did not occur in response to CSs. Thus, the actual biological importance associated with this “yin-yang” response is unknown. Longer-term studies of chronic CS dosing are necessary to better determine the relevance (or mechanisms) of these differences.

These results also have implications for treatment in line with the recent guidelines (29, 30). Clusters 1 and 2 can likely be well controlled with low to high doses of ICS. Cluster 3 may require higher doses of ICS and even systemic CSs, but is likely an excellent group in which to consider the use of targeted T2 biologic therapy. Finally, cluster 4, the poorly CS responsive group, may require novel approaches, perhaps with lesser benefit (and even detriment) from escalating doses of CSs.

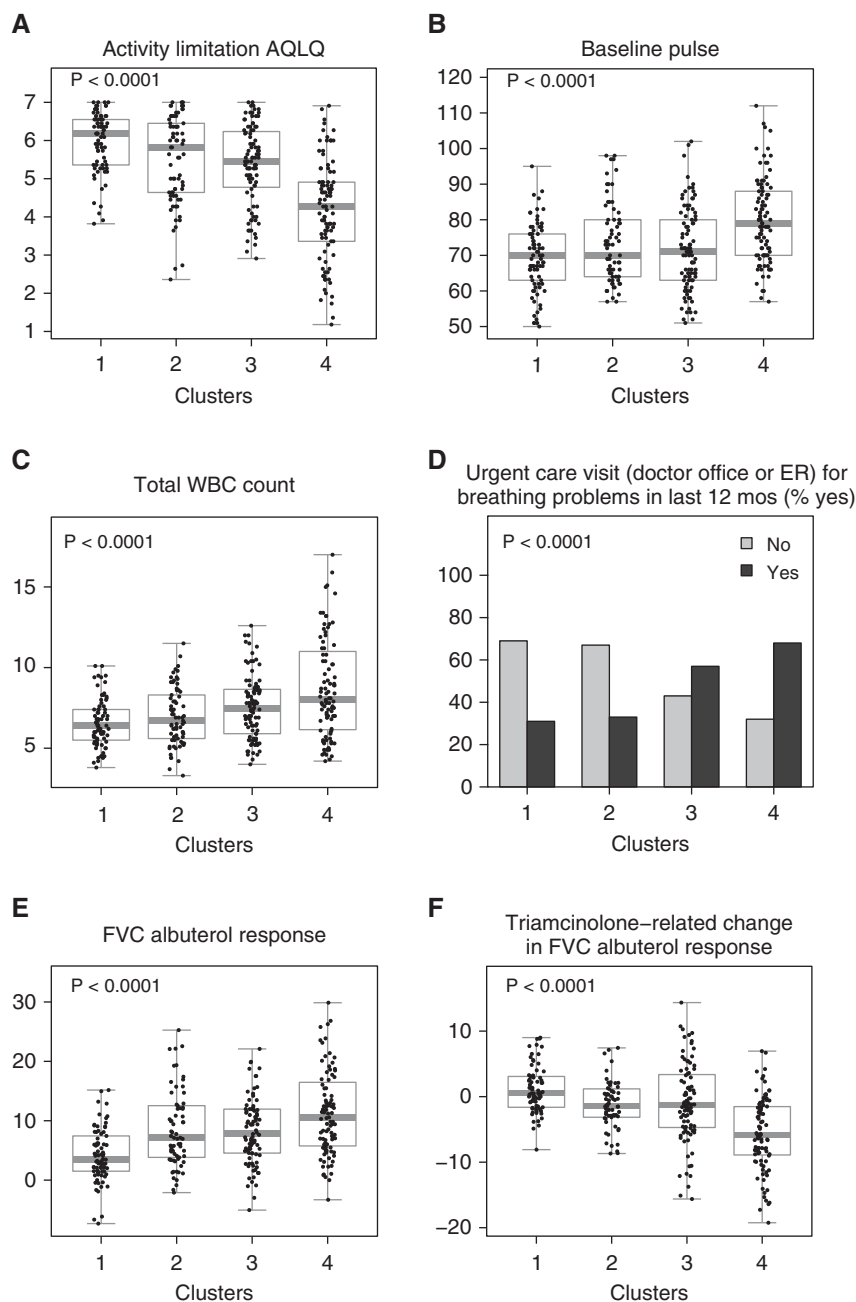


Figure 5. Distinguishing features of cluster 4. (A–E) At baseline, cluster 4 had the worst activity limitation Asthma Quality of Life Questionnaire (AQLQ) (A); highest pulse (B), total white blood cell (WBC) count (C), and healthcare use (D); and largest increase in FVC after albuterol treatment (E). (F) After triamcinolone treatment, the increase in FVC after albuterol treatment decreased the most among cluster 4 patients. The box plots in this figure are shown without outliers. See Figure E6 for box plots showing all data points for the same variables. ER = emergency room.

With the concept that a tool could be developed to help physicians determine which cluster a new patient belongs to, we identified the top 12 predictive baseline variables using feature selection techniques. With the use of SVM and a cross-validation

strategy, these variables predicted subjects in all clusters in the discovery set with high accuracy. Given the very distinct baseline characteristics of cluster 3, it is not surprising that the highest prediction sensitivity and specificity were achieved for cluster 3.

The identified clusters were validated/replicated using an independent SARP test set without sputum data. Using the 12 predictive baseline variables with the two surrogate variables (blood neutrophil and eosinophil counts) for the unavailable sputum variables, the cluster labels of the patients in the test set were replicated with SVM and showed characteristics remarkably similar to those in the discovery set, confirming our results. The development of publicly available software would enable clinicians to input these variables, manually or (more likely) in an automated format direct from the electronic medical record, to help them determine the likelihood of a response to systemic CSs.

Our study was limited to a single dose of systemic CSs (without placebo) and a single point in time. As shown previously (31, 32), physiological markers can be highly dynamic and change over time; hence, a robust longitudinal study design is required to assess the stability of these clusters. These patterns cannot be extrapolated to longer-term effects of chronic CS use, or the use of higher doses.

In summary, we identified four asthma clusters with differential CS responses using a multiview learning approach. These findings give insight into clinical, biological, and physiological determinants of CS response patterns that could be used mechanistically to better link molecular responses to clinical responses. The identification of small numbers of highly predictive nonredundant variables (using feature selection techniques) suggests that software could be developed to predict responses not only to CSs but also to expensive biologic therapies, which would help to improve the application of precision medicine. These machine-learning approaches are providing new insights into CS responsiveness in asthma and could ultimately lead to improved asthma management. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank all of the coordinators and laboratory technicians from the SARP network who made this study possible, and most importantly, the study participants themselves.

References

- Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006;368:804–813.
- Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012;18:716–725.
- Buchman AL. Side effects of corticosteroid therapy. *J Clin Gastroenterol* 2001;33:289–294.
- Dahl R. Systemic side effects of inhaled corticosteroids in patients with asthma. *Respir Med* 2006;100:1307–1317.
- Djukanović R, Wilson JW, Britten KM, Wilson SJ, Walls AF, Roche WR, et al. Effect of an inhaled corticosteroid on airway inflammation and symptoms in asthma. *Am Rev Respir Dis* 1992;145:669–674.
- Lipworth BJ. Systemic adverse effects of inhaled corticosteroid therapy: a systematic review and meta-analysis. *Arch Intern Med* 1999;159:941–955.
- Moore WC, Peters SP. Severe asthma: an overview. *J Allergy Clin Immunol* 2006;117:487–494; quiz 495.
- Phipatanakul W, Mauger DT, Sorkness RL, Gaffin JM, Holguin F, Woodruff PG, et al.; Severe Asthma Research Program. Effects of age and disease severity on systemic corticosteroid responses in asthma. *Am J Respir Crit Care Med* 2017;195:1439–1448.
- 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;178:218–224.
- Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al.; National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010;181:315–323.
- Wu W, Bleecker E, Moore W, Busse WW, Castro M, Chung KF, et al. Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data. *J Allergy Clin Immunol* 2014;133:1280–1288.
- Lanckriet GR, De Bie T, Cristianini N, Jordan MI, Noble WS. A statistical framework for genomic data fusion. *Bioinformatics* 2004;20:2626–2635.
- Noble WS. Support vector machine applications in computational biology. In: Schoelkopf B, Tsuda K, Vert J-P, editors. *Kernel methods in computational biology*. Cambridge, MA: MIT Press; 2004. pp. 71–92.
- Gönen M, Margolin AA. Localized data fusion for kernel k-means clustering with application to cancer biology. *Adv Neural Inf Process Syst* 2014;2:1305–1313.
- Bang S, Wu W. Multiple kernel k-means clustering using min-max optimization with l_2 regularization [preprint]. arXiv; 2018 [accessed 2018 Aug 20]. Available from: <https://arxiv.org/abs/1803.02458>.
- 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* 2013;43:343–373.
- Moore WC, Bleecker ER, Curran-Everett D, Erzurum SC, Ameredes BT, Bacharier L, et al.; National Heart, Lung, Blood Institute's Severe Asthma Research Program. Characterization of the severe asthma phenotype by the National Heart, Lung, and Blood Institute's Severe Asthma Research Program. *J Allergy Clin Immunol* 2007;119:405–413.
- Vortmann M, Eisner MD. BMI and health status among adults with asthma. *Obesity (Silver Spring)* 2008;16:146–152.
- Agresti A, Kateri M. Categorical data analysis. In: Lovric M, editor. *International encyclopedia of statistical science*. Berlin: Springer; 2011. pp. 206–208.
- Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods*. Hoboken, NJ: John Wiley & Sons; 2013.
- Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond Edinb Dublin Philos Mag J Sci* 1900;50:157–175.
- Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull* 1945;1:80–83.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.
- Miranda C, Busacker A, Balzar S, Trudeau J, Wenzel SE. Distinguishing severe asthma phenotypes: role of age at onset and eosinophilic inflammation. *J Allergy Clin Immunol* 2004;113:101–108.
- 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;132:336–341.
- Shan SW, Do CW, Lam TC, Kong RPW, Li KK, Chun KM, et al. New insight of common regulatory pathways in human trabecular meshwork cells in response to dexamethasone and prednisolone using an integrated quantitative proteomics: SWATH and MRM-HR mass spectrometry. *J Proteome Res* 2017;16:3753–3765.
- Raghunathan VK, Morgan JT, Park SA, Weber D, Phinney BS, Murphy CJ, et al. Dexamethasone stiffens trabecular meshwork, trabecular meshwork cells, and matrix. *Invest Ophthalmol Vis Sci* 2015;56:4447–4459.
- Kasetti RB, Maddineni P, Patel PD, Searby C, Sheffield VC, Zode GS. Transforming growth factor β 2 (TGF β 2) signaling plays a key role in glucocorticoid-induced ocular hypertension. *J Biol Chem* 2018;293:9854–9868.
- Global Initiative for Asthma. Global strategy for asthma management and prevention. [updated 2018; accessed 2018 Aug 20]. Available from: <https://ginasthma.org>.
- National Heart, Lung, and Blood Institute. Expert panel report 3 (EPR3): guidelines for the diagnosis and management of asthma. 2007 [accessed 2018 Aug 20]. Available from: <http://www.nhlbi.nih.gov/guidelines/asthma/asthdln.htm>.
- Frey U, Brodbeck T, Majumdar A, Taylor DR, Town GI, Silverman M, et al. Risk of severe asthma episodes predicted from fluctuation analysis of airway function. *Nature* 2005;438:667–670.
- Thamrin C, Stern G, Strippoli MP, Kuehni CE, Suki B, Taylor DR, et al. Fluctuation analysis of lung function as a predictor of long-term response to beta2-agonists. *Eur Respir J* 2009;33:486–493.