[ Original Research ]

CHEST

# Phenotypic Clusters Predict Outcomes in a Longitudinal Interstitial Lung Disease Cohort

Ayodeji Adegunsoye, MD; Justin M. Oldham, MD; Jonathan H. Chung, MD; Steven M. Montner, MD; Cathryn Lee, MD; Leah J. Witt, MD; Danielle Stahlbaum, MD; Rene S. Bermea, MD; Lena W. Chen, BS; Scully Hsu, BS; Aliya N. Husain, MD; Imre Noth, MD; Rekha Vij, MD; Mary E. Strek, MD; and Matthew Churpek, MD, MPH, PhD

**BACKGROUND:** The current interstitial lung disease (ILD) classification has overlapping clinical presentations and outcomes. Cluster analysis modeling is a valuable tool in identifying distinct clinical phenotypes in heterogeneous diseases. However, this approach has yet to be implemented in ILD.

**METHODS:** Using cluster analysis, novel ILD phenotypes were identified among subjects from a longitudinal ILD cohort, and outcomes were stratified according to phenotypic clusters compared with subgroups according to current American Thoracic Society/European Respiratory Society ILD classification criteria.

**RESULTS:** Among subjects with complete data for baseline variables (N = 770), four clusters were identified. Cluster 1 (ie, younger Caucasian obese female subjects) had the highest baseline FVC and diffusion capacity of the lung for carbon monoxide (DLCO). Cluster 2 (ie, younger African-American female subjects with elevated antinuclear antibody titers) had the lowest baseline FVC. Cluster 3 (ie, elderly Caucasian male smokers with coexistent emphysema) had intermediate FVC and DLCO. Cluster 4 (ie, elderly Caucasian male smokers with severe honeycombing) had the lowest baseline DLCO. Compared with classification according to ILD subgroup, stratification according to phenotypic clusters was associated with significant differences in monthly FVC decline (Cluster 4, –0.30% vs Cluster 2, 0.01%; $P < .0001$). Stratification by using clusters also independently predicted progression-free survival ($P < .001$) and transplant-free survival ($P < .001$).

**CONCLUSIONS:** Among adults with diverse chronic ILDs, cluster analysis using baseline characteristics identified four distinct clinical phenotypes that might better predict meaningful clinical outcomes than current ILD diagnostic criteria. CHEST 2017; ■(■):■-■

**KEY WORDS:** cluster; interstitial lung disease; mortality; phenotype; pulmonary fibrosis

**ABBREVIATIONS:** ANA = antinuclear antibody; ATS = American Thoracic Society; CHP = chronic hypersensitivity pneumonitis; CTD = connective tissue disease; CTD-ILD = connective tissue disease-associated interstitial lung disease; DLCO = diffusing capacity of the lungs for carbon monoxide; ERS = European Respiratory Society; GAP-ILD = gender, age, physiology-interstitial lung disease; HR = hazard ratio; HRCT = high-resolution CT; ILD = interstitial lung disease; IPAF = interstitial pneumonia with autoimmune features; IPF = idiopathic pulmonary fibrosis; NSIP = nonspecific interstitial pneumonia; PA = pulmonary artery; PAM = partitioning around medoids; PFT = pulmonary function test; PFS = progression-free survival; TFS = transplant-free survival; SLB = surgical lung biopsy

**AFFILIATIONS:** From the Section of Pulmonary & Critical Care (Drs Adegunsoye, Lee, Witt, Noth, Vij, Strek, and Churpek and Ms Chen), Department of Medicine, University of Chicago, Chicago, IL; Division of Pulmonary, Critical Care & Sleep Medicine (Dr Oldham), Department of Medicine, University of California at Davis, Davis, CA; and the Department of Radiology (Drs Chung and Montner), Department of Medicine (Drs Stahlbaum and Bermea and Ms Hsu), and the Department of Pathology (Dr Husain), University of Chicago, Chicago, IL.

Drs Strek and Churpek contributed equally.

**FUNDING/SUPPORT:** This investigation was supported by a National Institutes of Health T32 training grant (Grant T32-HL007605).





FLA 5.5.0 DTD ■ CHEST1355_proof ■ 17 October 2017 ■ 9:57 am ■ EO: CHEST-17-1344

Interstitial lung diseases (ILDs) are a heterogeneous group of pulmonary disorders characterized by architectural distortion and lung function impairment.[1] The approach to ILD has evolved over time, attempting to improve diagnostic precision while decreasing the need for invasive procedures.[2] Based on observed disease behavior, the most recent 2013 update of the American Thoracic Society(ATS)/European Respiratory Society (ERS) guidelines classifies ILDs into diagnostic subgroups that often overlap in their presentation and prognosis, making the clinical application of this recommended diagnostic algorithm challenging.[1,2]

A confident diagnosis is often limited by the inability of patients to undergo surgical lung biopsy or the presence of discordant radiologic and histopathologic patterns.[3,4] Such realities preclude the ability to diagnose patients with a specific ILD, leaving them without a clear prognosis or treatment options. In fact, some of these individuals are deemed to have "unclassifiable" ILD.[1,2] In addition, a significant subset of patients with ILDs exhibits serologic and clinical features suggestive of an underlying autoimmune process but do not meet defined criteria for a connective tissue disease (CTD). Various terminologies with subtly different criteria have been used to describe this subset of patients, including undifferentiated CTD-associated ILD (CTD-ILD), lung-dominant CTD, and autoimmune-featured ILD.[5-8] Recent ATS/ERS guidelines designate these patients as having interstitial pneumonia with autoimmune features (IPAF).[9] These diagnostic difficulties, along with tremendous variability in disease course within and between different ILDs, limit the utility of the current classification in stratifying patients into clinically meaningful subgroups with uniform outcomes over time.[10-13]

Statistical cluster analysis techniques have proven valuable in identifying homogeneous clusters of patients with shared clinical characteristics within pulmonary diseases such as COPD, asthma, and bronchiectasis.[14-16] Because ILDs may benefit from a similar approach, we conducted an innovative cluster analysis in a large longitudinal cohort of patients with chronic ILDs to identify unique ILD phenotypes based on clinical characteristics, serologic data, lung function, and radiographic features. We hypothesized that application of a cluster analysis approach would identify more homogeneous ILD phenotypes than the current ILD classification system with regard to clinically meaningful outcomes, and it could provide a foundation for improved understanding of ILD pathogenesis, disease progression, and optimizing approach to management.

## Methods

### Study Design and Patient Selection

Subjects in the present analysis are from the University of Chicago ILD Registry, a longitudinal ILD cohort in which data are collected prospectively. The University of Chicago Institutional Review Board approved this investigation (institutional review board protocols #14163-A; #16-1062), and all patients signed informed consent forms.

Patients followed up at our institution between 2006 and 2015 with a multidisciplinary diagnosis of chronic ILD according to ATS/ERS criteria[2,9,17-19] were screened. Subjects with idiopathic pulmonary fibrosis (IPF), IPAF, CTD-ILD, chronic hypersensitivity pneumonitis (CHP), and unclassifiable idiopathic interstitial pneumonias were identified and eligible for study inclusion. Multidisciplinary diagnosis of ILD at our institution is performed in a rigorous fashion in conjunction with rheumatologists, dedicated chest radiologists, and a thoracic pathologist. Patients with CTD-ILD were required to have a multidisciplinary diagnosis of CTD-ILD for inclusion in the study. The vast majority of the CHP cohort (> 90%) had either consistent histopathologic findings on lung biopsy or an identifiable environmental antigen. However, because multidisciplinary diagnosis remains the current gold standard for a diagnosis of CHP, the minimal criterion for a diagnosis of CHP was a multidisciplinary review of the clinical, pulmonary function test (PFT), radiographic, and pathologic characteristics that were most consistent with a diagnosis of CHP, after exclusion of all other possible etiologies. Patients with IPF, and those with unclassifiable idiopathic interstitial pneumonia, were required to meet 2011 ATS/ERS criteria for inclusion in the study. The minimal criteria for classification as IPAF were that patients must have an interstitial pneumonia (according to high-resolution CT [HRCT] scan or surgical lung biopsy) with exclusion of alternative etiologies, incomplete features of a defined CTD, and at least one feature from at least two IPAF domains (clinical, radiographic, and morphologic) as proposed by the initial IPAF research statement.[9]

### Data Collection

The electronic medical record was retrospectively reviewed to extract pertinent data, and 24 baseline variables were identified from each patient's initial clinic visit with substantial clinical relevance for inclusion in the cluster analysis model based on previous literature. These variables were as follows: demographic information (age, race/ethnicity, and sex), patient-reported historical information (tobacco use and other environmental exposure [organic or inorganic]), comorbid disease conditions (gastroesophageal reflux and hypothyroidism), physical examination findings [BMI, $SpO_2$:$FiO_2$ ratio, clubbing, and crackles), laboratory studies (antinuclear antibody [ANA] titer, positive rheumatoid factor [> 2× the upper

**CORRESPONDENCE TO:** Deji Adegunsoye, MD, Section of Pulmonary & Critical Care, Department of Medicine, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637; e-mail: deji@uchicago.edu



**DOI:** https://doi.org/10.1016/j.chest.2017.09.026





FLA 5.5.0 DTD ■ CHEST1355_proof ■ 17 October 2017 ■ 9:57 am ■ EO: CHEST-17-1344