

Scientific Letter

Utility of Z Scores for Defining Progressive Pulmonary Fibrosis in Interstitial Lung Disease

To the Editor,

Interstitial lung diseases (ILDs) comprise a heterogeneous group of disorders characterized by inflammation and/or fibrosis of the alveolar interstitium. A subset of patients develops progressive pulmonary fibrosis (PPF), defined by the presence of at least two of the following criteria: (1) worsening respiratory symptoms, (2) physiologic evidence of disease progression, or (3) radiologic evidence of disease progression. Physiologic progression is defined as a decline in forced vital capacity (FVC) of $\geq 5\%$ predicted or a decrease in diffusing capacity of the lung for carbon monoxide (DLCO) of $\geq 10\%$ predicted within 1 year or over the follow-up period [1].

According to ERS/ATS (2022) recommendations for interpreting pulmonary function tests, Global Lung Function Initiative (GLI) reference equations should be used for spirometry, lung volumes, and DLCO. These clinical practice guidelines emphasize the use of z-scores rather than percent predicted (PP) values to classify the severity of lung function impairment [2]. This shift reflects evidence that PP values miscategorize disease severity. For example, in a cohort of more than 11,000 patients with chronic obstructive pulmonary disease (COPD), Miller et al. reported that PP values miscategorized severity in approximately 24% of patients, while 10% of individuals with normal lung function were incorrectly categorized as being impaired [3].

In ILD population, applying the GLI equations leads to 4.9% higher DLCO PP (58.5% vs 63.4%; $p < 0.01$) and 8.8% lower FVC PP (87.7% vs 78.9%; $p < 0.01$), decreasing the number of patients that met the clinical trial criteria for antifibrotic treatment [4]. Boros et al. reported that the z-score-based system recategorized lung function severity in FVC. A total of 24.6% of patients were reallocated to less severe categories, whereas 28.1% were reclassified into more severe categories based on TLCO. When predicting mortality risk, z-score-based thresholds showed a stronger association with mortality than the traditional PP system [5]. In this context, we evaluated the relationship between z-score changes and PP changes in FVC and DLCO using GLI Caucasian reference equations in patients with ILD, and examined the applicability of z-scores for defining physiologic progression in PPF.

We conducted a retrospective and observational study across two centers in Latin America (the National Institute of Respiratory Diseases, México City, and the Colombian Pneumonological Foundation, Bogotá). Patients aged ≥ 18 years with baseline and 1-year follow-up (± 3 months) pulmonary function testing and chest computed tomography were eligible for inclusion, provided they had a multidisciplinary team-confirmed diagnosis of ILD due to connective tissue disease (CTD), hypersensitivity pneumonitis (HP), or idiopathic pulmonary fibrosis (IPF).

Absolute volumes of FEV1, FVC, and DLCO were collected using the GLI calculator (Version 2.0), available online [6], from which the z-score and PP values were obtained. Demographic data, including age, height, sex, ethnicity, tomographic patterns, diagnosis, and treatment, were also collected.

We analyzed the categorical variables using percentages and frequencies, and the continuous variables using means and standard deviations. The Mann–Whitney *U* test, chi-square test, and Fisher's exact test were used to compare the FPP and non-FPP groups. A simple linear regression analysis was conducted to assess the association between the dependent variable, z-score, and the independent variable, PP. Because Q–Q plots and Shapiro–Wilk test showed non-normal distributions and heteroscedastic residuals, a Yeo–Johnson transformation was applied for both variables prior to lineal modeling. As heteroscedasticity persisted (Breusch–Pagan test and Shapiro–Wilk test for residuals, both $p < 0.01$), the model was fitted using SE HC3 (heteroscedasticity-consistent standard errors, version HC3) which provide bias-reduced and leverage-adjusted variance estimates that are particularly robust in the presence of heteroscedasticity and influential observations. Robust *p*-values were therefore reported. For clinical interpretation, estimates were back-transformed to the original PP and z-score scales. Agreement between classifications was assessed using Cohen's kappa coefficient and 2×2 contingency tables; sensitivity and specificity were also calculated. A *p*-value < 0.05 was considered statistically significant. Statistical analysis was performed with the IBM SPSS Statistics Version 26 software package and RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA.

We evaluated a total of 254 patients, 135 (53%) from México and 119 (47%) from Colombia. Predominantly, the population included women, 161 (64%), with a mean age of 63 years. Most patients were diagnosed with CTD (137; 54%), HP (60, 24%), and IPF (57; 22%).

In the PPF group, there was a higher proportion of patients with IPF (30% vs 19%) and HP (30% vs 21%), as well as a significantly greater use of antifibrotic therapy, particularly nintedanib (29% vs 8%).

Regarding tomographic patterns, both groups showed a similar distribution of UIP (31% vs 33%); however, fibrotic patterns such as f-NSIP (24% vs 18%), f-HP (27% vs 8%), and CPFE (8% vs 3%) were more common in the PPF group. In contrast, non-fibrotic patterns – including nf-HP, c-NSIP, and OP – were more frequent in the non-PPF group.

Regarding lung function, only DLCO differed significantly between the two groups at baseline. At the 1-year follow-up, patients with PPF showed lower FVC, FEV and DLCO values, which is consistent with a more fibrotic and progressive behavior. The baseline and 1-year follow-up lung function data for one year are shown in Table 1.

<https://doi.org/10.1016/j.arbres.2026.01.021>

0300-2896/© 2026 The Authors. Published by Elsevier España, S.L.U. on behalf of SEPAR. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Characteristics of the study population.

Variable	n = 254	PPF n = 74	Non-PPF n = 180	p
México	135 (53)	27 (36)	108 (60)	<0.01
Colombia	119 (47)	47 (63)	72 (40)	
Female	161 (64)	41 (55)	120 (66)	0.1
Age	63 ± 12	65 ± 12	63 ± 12	0.1
Diagnosis				
Due to CTD	137 (54)	30 (40)	107 (60)	0.02
HP	60 (24)	22 (30)	38 (21)	
IPF	57 (22)	22 (30)	35 (19)	
Treatment				
PDN	99 (39)	33 (45)	65 (36)	0.2
MMF	16 (6)	25 (33)	88 (48)	0.03
MTX	13 (5)	3 (4)	9 (5)	1
Nintedanib	38 (15)	22 (29)	16 (8)	<0.001
Pirfenidone	40 (16)	13 (17)	27 (15)	0.7
Tomographic patten				
UIP	84 (33)	23 (31)	60 (33)	<0.001#
f-NSIP	50 (20)	18 (24)	33 (18)	
f-HP	36 (14)	20 (27)	16 (8)	
nf-HP	26 (10)	3 (4)	23 (12)	
c-NSIP	17 (7)	2 (2)	15 (8)	
OP	19 (8)		19 (10)	
CPFE	12 (5)	6 (8)	6 (3)	
LIP	4 (2)	1 (1)	3 (1)	
Lung function test				
<i>Baseline</i>				
FVC, l	2.4 ± 0.8	2.2 ± 0.7	2.4 ± 0.9	0.08
FVC, pp	79 ± 22	76 ± 23	80 ± 22	0.1
FVC, z-score	-1.4 ± 1.6	-1.7 ± 1.7	-1.3 ± 1.5	0.1
FEV1, l	1.9 ± 0.6	1.8 ± 1.5	1.9 ± 1.4	0.2
FEV1, pp	80 ± 22	78 ± 22	81 ± 21	0.4
FEV1, z-score	-1.3 ± 1.4	-1.4 ± 1.2	-1.2 ± 1.5	0.5
DLCO, mL/min/mmHg	14.3 ± 6.2	12 ± 5.1	15 ± 6.5	<0.01
DLCO, pp	74 ± 31	64 ± 28	77 ± 30	<0.01
DLCO, z-score	-2.1 ± 2.4	-2.9 ± 2.3	-1.8 ± 2.4	<0.01
<i>1-Year follow-up</i>				
FVC, l	2.2 ± 0.8	2.0 ± 0.7	2.3 ± 0.8	<0.01
FVC, pp	76 ± 22	68 ± 21	78 ± 21	<0.01
FVC, z-score	-1.6 ± 1.6	-2.1 ± 1.6	-1.4 ± 1.5	<0.01
FEV1, l	1.8 ± 0.6	1.6 ± 1.5	1.8 ± 1.3	0.04
FEV1, pp	77 ± 22	72 ± 22	78 ± 20	0.04
FEV1, z-score	-1.5 ± 1.4	-1.7 ± 1.5	-1.3 ± 1.3	0.06
DLCO, mL/min/mmHg	12.9 ± 6	9 ± 4	14 ± 6	<0.001
DLCO, pp	67 ± 30	51 ± 24	73 ± 30	<0.001
DLCO, z-score	-2.6 ± 2.5	-4 ± 2.3	-2 ± 2	<0.001

PP: progressive pulmonary fibrosis; PP: predicted percent; CTD: connective tissue disease; HP: hypersensitivity pneumonitis; IPF: idiopathic pulmonary fibrosis; PDN: prednisone; MMF: mycophenolate mofetil; MTX: methotrexate; UIP: usual interstitial pneumonia; f-NSIP: fibrotic nonspecific pneumonia; f-HP: fibrotic hypersensitivity pneumonitis; nf-HP: non fibrotic hypersensitivity pneumonitis; c-NSIP: cellular nonspecific pneumonia; OP: organized pneumonia; CPFE: combined pulmonary emphysema fibrosis; LIP: lymphocytic interstitial pneumonia; FVC: forced vital capacity, FEV1: forced expiratory volume in 1 second, DLCO: diffusing capacity of the lungs for carbon monoxide. All values are expressed as frequencies, percentages, means, and standard deviations (SD).

Fisher's exact test.

The cross-sectional relationship between z-scores and PP values was found to deviate from linearity at the lower end of the pulmonary function patients (Fig. 1A and B), particularly for DLCO. As illustrated in Fig. 1B observations with markedly reduced PP DLCO exhibit greater dispersion and a steeper curvature in the z-score scale vs values in the mid-to-high physiological range. This pattern reflects the inherent nonlinearity of the reference equations, the compression of z-scores near the lower tails of the distribution, and the increased measurement variability known to occur in individuals with more severe impairment. Consequently, small absolute changes in PP DLCO at low baseline values may correspond to disproportionately large shifts in z-scores, emphasizing the importance of interpreting both metrics jointly rather than interchangeably in this segment of the distribution.

A 5% drop predicted in FVC corresponded, on average, to a decline of approximately 0.37 units in the FVC z-score. This

equivalence was supported by a high level of agreement in classifying patients with functional decline (Cohen's kappa = 0.89, $p < 0.001$), with sensitivity and specificity of 0.99 and 0.94, respectively (Fig. 1C). These findings indicate that z-scores can be used to define functional progression in a manner consistent with widely used PP thresholds.

In the regression analysis, each 1-unit increase in FVC PP (Yeo-Johnson transformed) was associated with an expected increase of 0.992 units in the FVC z-score (Yeo-Johnson transformed) (95%CI, 0.968-1.017; robust SE, 0.012; $p < 0.001$). Although residual heteroskedasticity was present, we estimated the model using HC3 heteroscedasticity-consistent standard errors, which provide unbiased inference under these conditions. Thus, the statistical significance and confidence intervals reported are valid even in the presence of persistent heteroskedasticity. Back-transformed estimates are reported to facilitate clinical interpretation.

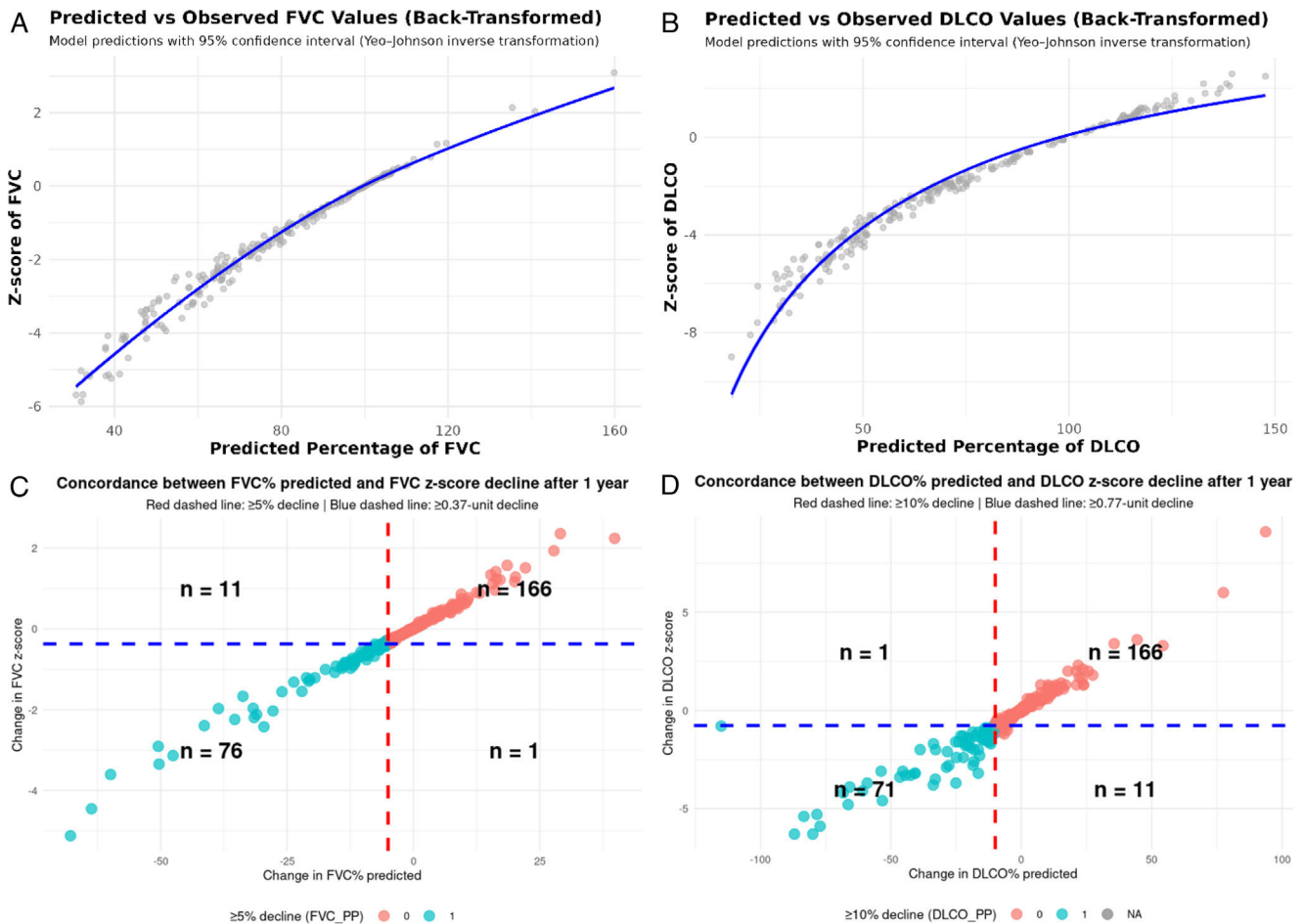


Fig. 1. (A and B) Scatter plot of predicted vs observed values (Yeo-Johnson inverse transformation) predicted percent vs z-score for FVC (A) and DLCO (B). (C and D) Concordance between PP and z-score decline after 1 year. Each dot represents a patient. The red dashed line indicates a $\geq 5\%$ drop in FVC% predicted, and the blue dashed line represents a ≥ 0.37 -unit decline in FVC z-score (C). FVC: forced vital capacity; DLCO: diffusing capacity of the lungs for carbon monoxide; PP: predicted percent.

For DLCO, a 10% drop predicted in DLCO corresponded, on average, to a decline of approximately 0.77 units in DLCO z-score. Agreement in classifying patients who experienced decline was high (Cohen's kappa 0.88, $p < 0.001$), with a sensitivity of 0.99 and a specificity of 0.94. The regression model show that every unit of change in DLCO pp (Yeo-Johnson transformed) corresponds to 0.984 units in DLCO z-score (Yeo-Johnson transformed) (95%CI, 0.95–1.017; Robust SE, 0.017; $p < 0.001$).

To our knowledge, this is the first study to introduce the use of z-scores instead of PP to identify physiological criteria of PPF using GLI reference equations and assess the agreement between them.

Z-score values of 0.37 for FVC and 0.77 for DLCO identifies approximately a similar subset of patients with meaningful lung function deterioration over time and would take place in the physiological criteria of PPF instead of 5% predicted in FVC and 10% predicted in DLCO.

Caution is advised when adopting this reference equation, as Brazzale et al. demonstrate that DLCO can result in an altered interpretation depending on the equation used, compared to other equations such as Miller, NHANES-III, or Crapo [7]. In ILD patients, the use of GLI reference equations has been limited; Li et al. concluded that applying these reference equations in patients with ILD leads to higher DLCO pp values, and fewer patients met the criteria for antifibrotic agents [8]. The impact extends not only to treatment but also to the inclusion of trials, where FVC and DLCO are considered, as using the GLI reference equation could change the eligibility status of these patients [4]. Recently, Boros et al. found that a 1-unit decline in the FVC z-score was associated with a 10.3%

increase in the risk of death. In comparison, a one-unit decline in DLCO z-score was associated with an over 30% increase in mortality risk, highlighting the importance of z-score as a mortality predictor [5].

One of this study limitations is that our entire population was Hispanic, which limits the generalizability of the results. Although the GLI equation is multiethnic and does not fully represent the Hispanic population, it did include data from healthy adult individuals from 6 Latin American countries that together represent the mixed-race population characteristic of this region [9], as well as information from the Mexican-American population [10]; indeed, a global multiethnic study is crucial to validate the GLI equations references in ILD patients.

Physiological measures of lung function are central to identifying disease progression and determining when to initiate antifibrotic therapy. Our findings support the potential use of z-scores to assess functional progression in ILD. However, larger and prospective studies are needed to confirm the prognostic value of z-score based thresholds and to determine how best to integrate them into clinical decision making and treatment criteria.

CRedit authorship contribution statement

The authors confirm their contributions to the manuscript as follows: study conception and design were performed by Buendía-Roldán I and Alarcón-Dionet A; data collection by Villafuerte R and Rincón-Álvarez E; analysis and interpretation of the results

by González-García M, Alarcón-Dionet A, and Selman M; and manuscript drafting by Alarcón-Dionet A, Buendía-Roldán I, and Selman M.

All authors reviewed the results and approved the final version of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

This material has not been produced with the help of any artificial intelligence software or tool.

Funding

None declared.

Conflicts of interest

None declared.

References

- [1] Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2022;205(9):e18–47. <http://dx.doi.org/10.1164/rccm.202202-0399ST>.
- [2] Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022;60(1). <http://dx.doi.org/10.1183/13993003.01499-2021>, 2101499.
- [3] Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011;139(1):52–9. <http://dx.doi.org/10.1378/chest.10-0189> [Epub 3 June 2010; erratum in: *Chest*. 2011 Mar;139(3):733. PMID: 20522571].
- [4] Li A, Teoh A, Troy L, Glaspole I, Wilsher ML, de Boer S, et al. Implications of the 2022 lung function update and GLI global reference equations among patients with interstitial lung disease. *Thorax* 2024;79(11):1024–32. <http://dx.doi.org/10.1136/thorax-2024-221813>.
- [5] Boros PW, Martusewicz-Boros MM, Lewandowska KB. Assessment of lung function and severity grading in interstitial lung diseases (% predicted versus z-scores) and association with survival: a retrospective cohort study of 6,808 patients. *PLoS Med* 2025;22(5). <http://dx.doi.org/10.1371/journal.pmed.1004619>, e1004619.
- [6] Hall GL, Filipow N, Ruppel G, Okitika T, Thompson B, Kirkby J, et al. Contributing GLI Network members. Official ERS technical standard: Global Lung Function Initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J* 2021;57:2000289. <http://dx.doi.org/10.1183/13993003.00289-2020>.
- [7] Brazzale DJ, Seccombe LM, Welsh L, Lanteri CJ, Farah CS, Ruehland WR. Effects of adopting the Global Lung Function Initiative 2017 reference equations on the interpretation of carbon monoxide transfer factor. *Eur Respir J* 2020;55(5). <http://dx.doi.org/10.1183/13993003.01905-2019>, 1901905.
- [8] Wapenaar M, Miedema JR, Lammering CJ, Mertens FW, Wijsenbeek MS. The impact of the new Global Lung Function Initiative TLCO reference values on trial inclusion for patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2019;53(2). <http://dx.doi.org/10.1183/13993003.01895-2018>, 1801895.
- [9] Pérez-Padilla R, Valdivia G, Muiño A, López MV, Márquez MN, Montes de Oca M, et al. Valores de referencia espirométricos en cinco grandes ciudades latinoamericanas para sujetos de 40 años o más. *Arch Bronconeumol* 2006;42:317–25. [http://dx.doi.org/10.1016/s1579-2129\(06\)60540-5](http://dx.doi.org/10.1016/s1579-2129(06)60540-5).
- [10] Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999;159:179–87.

Aimé Alarcón-Dionet^a, Mauricio González-García^b,
Ricardo Villafuerte^a, Emily Rincón-Alvarez^b,
Moises Selman^a, Ivette Buendía-Roldán^{a,*}

^a Instituto Nacional de Enfermedades Respiratorias, Dr. Ismael Cosío Villegas, Mexico City, Mexico

^b Fundación Neumológica Colombiana, Bogotá, Colombia

*Corresponding author.

E-mail address: ivettebu@yahoo.com.mx (I. Buendía-Roldán).