NCT
Neumología y Cirugía de Tórax

Check for updates

LETTER TO THE EDITOR

Respiratory function tests: interpretation or classification?

Pruebas de función respiratoria: ¿interpretación o clasificación?

Luis Torre-Bouscoulet

Dirección Médica, Instituto para el Desarrollo e Innovación en Fisiología Respiratoria, Mexico City, Mexico

Current strategies for interpreting pulmonary function tests (PFTs) are, in reality, guidelines for classifying them[1]. The verbs *to classify* and *to interpret* are not interchangeable. Interpretation implies assigning meaning to a result. The interpretation of PFT results corresponds to the clinician, who, in light of the medical history, physical examination, and other diagnostic studies, gives meaning to functional findings. Interpretation implies integration of information. In contrast, classifying the results of functional studies is a simpler task, whose objective is to standardize the way we refer to physiological patterns. Standardized classification requires adherence to operational definitions, which inherently have limitations. Despite the clear differences between classification and interpretation, the terms are frequently used interchangeably, and this misuse underlies many of the clinical controversies surrounding PFT results.

Functional patterns derived from PFTs only increase or decrease the pretest probability of a specific diagnosis. Classification algorithms are based on mathematical models. These models may be deterministic or probabilistic. An example of the former includes classical physical laws, in which all factors determining an outcome are known. For example, velocity is determined by distance and time. By contrast, in clinical medicine, models are generally probabilistic. That is, they allow estimation of the probability that a phenomenon is present. Probabilistic models increase or decrease certainty regarding the existence of a phenomenon. For example, if a subject has a forced vital capacity (FVC) of 0.8 L, there is a high probability (greater certainty) of fibrosing lung disease; however, other causes (eg, muscle weakness, air trapping) may explain the finding. Probability is therefore a measure of certainty and ranges from 0 to 1. The predictive performance of mathematical models used in respiratory physiology—even when incorporating multiple parameters—typically ranges between 0.4 and 0.6. The remaining difference from unity reflects uncertainty.

In clinical physiology, the comparator is usually a multiple regression equation (accounting for age, height, and sex) used to estimate expected lung function in a healthy individual. The data used to construct these equations are derived from individuals considered healthy. However, what defines “healthy”? The stricter the criteria used to define health, the greater the selection bias. Functional values derived from exceptionally healthy individuals may be inappropriately high when used as comparators, thereby increasing false-positive rates[2]. This represents only one source of variability in reference equations. Currently, global reference equations constructed from data from thousands of healthy individuals across multiple countries are widely accepted[3]. Implementation of these equations has altered the prevalence of functional abnormalities and clinical diagnoses, with a significant increase in the frequency of respiratory disease, particularly among individuals of African ancestry[3-4].

Even assuming that reference equations accurately represent healthy individuals in a population, another issue arises: the definition of normality. Normality has

Correspondence:
Luis Torre-Bouscoulet
E-mail: luistorreb@gmail.com

Date of reception: 31-05-2025
Date of acceptance: 19-06-2025
DOI: 10.24875/NCTE.M25000027

Available online: 12-05-2026
Neumol Cir Torax (Eng). 2025;84(2):160-162
www.revistanct.org.mx

3082-0006 / © 2025 Sociedad Mexicana de Neumología y Cirugía de Tórax. Published by Permanyer. This is an open access article under the license. CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

been defined statistically; however, what is frequent is not necessarily synonymous with healthy. For example, based on frequency alone, one might argue that in Mexico a body mass index (BMI) between 28 and 32 is "normal," but although frequent, it is not healthy. "Normal" should not be conflated with "healthy."

To define normality, we rely on data distribution. A standard normal distribution has a mean of 0 and a standard deviation (SD) of 1, with mean, median, and mode identical. This curve has been used as a reference for defining normality[5]. Returning to FVC: if an individual has an FVC of 4.9 L and the reference mean is 5 L, there is a high probability that the value is "normal," since it deviates minimally from the mean. How close must a measurement be to the reference mean to be considered normal? The distance between the measured value and the reference mean may be expressed in original units (eg, liters) or standardized as SD units. The number of SDs represents a uniform and comparable measure of deviation from the group mean. SD units are synonymous with the Z score[6]. The Z score standardizes data dispersion[6].

In a standard normal distribution, 90% of observations lie between –1.64 and +1.64 Z scores, 95% between –1.96 and +1.96, and 99% between –2.58 and +2.58. By convention, the lower limit of normal (LLN) corresponds to the 5th percentile (Z = –1.64). Notably, 5% of healthy individuals will fall below this threshold. Thus, defining normality using the LLN excludes 5% of healthy individuals and represents an imperfect cutoff.

To illustrate cutoff limitations, consider the FEV_1/FVC ratio. This spirometric parameter defines obstruction. Most pulmonologists accept that FEV_1/FVC < LLN indicates obstruction; consequently, values above LLN are considered nonobstructive. However, when less stringent definitions are used (eg, FEV_1/FVC < 0.7), functional abnormalities are detected even among individuals classified as normal by the stricter LLN criterion. These abnormalities include reduced diffusing capacity for carbon monoxide (DLCO), impaired ventilatory efficiency during exercise, and lower oxygen consumption[7].

Even within the area under the curve considered normal, important differences exist. In terms of mortality, it is not equivalent to be classified as normal with an FEV_1 at –1.5 Z score vs an FEV_1 at +0.5 Z score. Both values fall within the normal range; however, the former is associated with higher mortality than the latter. This mortality gradient among individuals considered healthy was recently demonstrated in a US study that included 2 cohorts followed for 20 years[8]. These findings underscore that nature does not recognize arbitrary cutoff points. Lung function is a continuous variable, and categorizing individuals into discrete functional intervals is therefore somewhat artificial and potentially misleading. A recent British study including more than 300,000 participants demonstrated – consistent with the observations of John Hutchinson in 1846 – that lung function measured in its original units (liters of FVC and FEV_1), without covariate adjustment, predicts 12-year mortality[9-10].

Given the many limitations of the current definition of normality – some of which have been discussed above – it is reasonable to ask how we might achieve a more meaningful and efficient interpretation of PFT results. Concluding that a given range of FEV_1 or FVC values is normal or abnormal, and further categorizing them as mild, moderate, or severe, clearly oversimplifies functional findings.

One alternative to this limitation is to perform repeated measurements of lung function, thereby describing the temporal course of functional parameters, also referred to as functional trajectories. Numerous prenatal and postnatal factors influence lung development during childhood and, consequently, the level of lung function achieved in adulthood[11]. Lung function attained between 20 and 25 years of age ultimately reflects both genetic and environmental influences. The clinical and epidemiologic challenge is to identify and modify risk factors to alter the trajectory of lung function so that individuals reach optimal pulmonary capacity. Otherwise, individuals may enter adulthood with reduced lung function and subsequently experience progressive decline associated with aging, environmental exposures, or chronic disease. Thus, it may be advisable to move beyond purely cross-sectional analysis of lung function toward a longitudinal approach, in which the individual serves as their own comparator. Several methods exist to evaluate longitudinal changes in both children and adults[1]. The FEV_1Q method is among the most robust approaches for this purpose in adults[12]. FEV_1Q is calculated by dividing the measured FEV_1 (in liters) by 0.5 in men and 0.4 in women; these denominators correspond to the first percentile of FEV_1 in a healthy population for each sex. As the ratio (FEV_1Q) approaches 1, respiratory function is increasingly impaired. Under normal conditions, a decline of one unit in this ratio occurs over approximately 18 years. If such a decline occurs over a shorter period – eg, 10 years – this suggests accelerated loss of lung function. For nonspirometric functional parameters, however, longitudinal data remain limited and represent an important area for future research. Most functional trajectories have been derived from

cross-sectional data; therefore, additional long-term cohort studies are needed to reduce sources of variability.

Another alternative in the interpretation of PFTs-beyond cross-sectional and longitudinal analysis-is systems medicine and artificial intelligence (AI)[13]. Biomedical research has evolved substantially. Historically, medicine was reactive, focusing on addressing emerging health problems through clinical descriptions, case series, or relatively simple clinical trials. Subsequently, research efforts shifted toward phenotype identification, particularly through omics-based approaches. Multiomics, which examines interactions among omic layers, has improved understanding of systems biology, where the interactions among components are more important than isolated elements. This paradigm has advanced the development and implementation of precision medicine[14]. In this context, PFTs can be integrated as functional aggregates into large-scale data networks, enabling the construction of complex biological models capable of predicting therapeutic response and individual prognosis with minimal error. Under such circumstances, pulmonary function assessment would acquire greater clinical meaning. Until then, the interpretation of PFTs will remain – as Dr Alberto Neder has noted – a study with n = 1[15].

By way of corollary, several conclusions may be drawn. First, cross-sectional PFT results inherently generate classification errors depending on the cutoff used. Second, statistical normality does not necessarily equate to clinical health. Third, individualized analysis of longitudinal functional trajectories appears to be a more efficient strategy for interpreting PFTs. Fourth, incorporation of lung function into complex AI-based models may improve the identification of health priorities.

Funding

This study received no specific funding from public, commercial, or not-for-profit agencies.

Conflicts of interest

The author declared no conflicts of interest whatsoever.

Ethical considerations

Protection of humans and animals. No experiments involving humans or animals were performed.

Confidentiality and informed consent. The study does not involve personal patient data and did not require ethics approval. SAGER guidelines do not apply.

Declaration on the use of artificial intelligence (AI). The authors declare that no generative artificial intelligence was used in the writing or creation of the content of this manuscript.

References

1. 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):2101499. https://doi.org/10.1183/13993003.01499-2021.
2. Eisen EA, Wegman DH, Louis TA, Smith TJ, Peters JM. Healthy worker effect in a longitudinal study of one-second forced expiratory volume (FEV1) and chronic exposure to granite dust. Int J Epidemiol. 1995;24(6):1154-61. https://doi.org/10.1093/ije/24.6.1154.
3. Bhakta NR, Bime C, Kaminsky DA, McCormack MC, Thakur N, Stanojevic S, et al. Race and ethnicity in pulmonary function test interpretation: an Official American Thoracic Society Statement. Am J Respir Crit Care Med. 2023;207(8):978-95. https://doi.org/10.1164/rccm.202302-0310st.
4. Moffett AT, Bowerman C, Stanojevic S, Eneanya ND, Halpern SD, Weissman GE, et al. Global, race-neutral reference equations and pulmonary function test interpretation. JAMA Netw Open. 2023;6(6):e2316174. https://doi.org/10.1001/jamanetworkopen.2023.16174.
5. The Standard Norma Distribution [Internet]. Scibbr [consulted May 26 2025]. Available from: https://www.scribbr.com/statistics/standard-normal-distribution/.
6. Standardizing a normal distribution [Internet]. Scibbr [consulted May 26 2025]. Available from: https://www.scribbr.com/statistics/standard-normal-distribution/.
7. Neder JA, Milne KM, Berton DC, de-Torres JP, Jensen D, Tan WC, et al.; CRRN (Canadian Respiratory Research Network) and the CanCOLD (Canadian Cohort of Obstructive Lung Disease) Collaborative Research Group. Exercise tolerance according to the definition of airflow obstruction in smokers [letter]. Am J Respir Crit Care Med. 2020;202(5):760-2. https://doi.org/10.1164/rccm.202002-0298le.
8. Cannon MF, Goldfarb DG, Zeig-Owens RA, Hall ChB, Choi J, Cohen HW, et al. Normal lung function and mortality in World Trade Center Responders and National Health and Nutrition Examination Survey III Participants. Am J Respir Crit Care Med. 2024;209(10):1229-37. https://doi.org/10.1164/rccm.202309-1654oc.
9. Zhou L, Yang H, Zhang Y, Wang Y, Zhou X, Liu T, et al. Predictive value of lung function measures for cardiovascular risk: a large prospective cohort study. Thorax. 2024;79(3):250-8. https://doi.org/10.1136/thorax-2023-220703.
10. Kouri A, Dandurand RJ, Usmani OS, Chow CW. Exploring the 175-year history of spirometry and the vital lessons it can teach us today. Eur Respir Rev. 2021;30(162):210081. https://doi.org/10.1183/16000617.0081-2021.
11. Agusti A, Faner R. Lung function trajectories in health and disease. Lancet Respir Med. 2019;7(4):358-64. https://doi.org/10.1016/s2213-2600(18)30529-0.
12. Balasubramanian A, Wise RA, Stanojevic S, Miller MR, McCormack MC. FEV_1Q: a race-neutral approach to assessing lung function. Eur Respir J. 2024;63(4):2301622. https://doi.org/10.1183/13993003.01622-2023
13. Torre-Bouscoulet L. Medicina respiratoria de sistemas. Neumol Cir Torax. 2023;82(1):36-7. https://dx.doi.org/10.35366/114227.
14. Torre-Bouscoulet L. Los retos de la medicina personalizada. Neumol Cir Torax. 2015;74(4):238-9.
15. Neder JA. The new ERS/ATS standards on lung function test interpretation: some extant limitations. Eur Respir J. 2022;60(2):2200252. https://doi.org/10.1183/13993003.00252-2022.