ERS/ATS Global Lung Function Initiative normal values and classifying severity based on z-scores instead of per cent predicted

Rachel Ong-Salvador[1], Pierantonio Laveneziana [2,3] and Franciscus de Jongh[4,5,6]

[1]Pulmonary Function Testing, Department of Pulmonary Medicine, Amsterdam University Medical Centres, Amsterdam, The Netherlands. [2]AP-HP, Groupe Hospitalier Universitaire APHP-Sorbonne Université, Hôpitaux Pitié-Salpêtrière, et Tenon, Service des Explorations Fonctionnelles de la Respiration, de l'Exercice et de la Dyspnée (Département R3S), Paris, France. [3]Sorbonne Université, INSERM, UMRS1158 Neurophysiologie Respiratoire Expérimentale et Clinique, Paris, France. [4]Pulmonary Function Testing, Department of Pulmonary Medicine, Medisch Spectrum Twente, Enschede, The Netherlands. [5]Faculty of Mechanical Engineering, Department of Engineering Fluid Dynamics, University of Twente, Enschede, The Netherlands. [6]Department of Neonatology, Amsterdam University Medical Centres, Amsterdam, The Netherlands.

Corresponding author: Rachel Ong-Salvador (r.ong@amsterdamumc.nl)

Check for updates

Shareable abstract (@ERSpublications)
Continuous effort is being made to standardise the interpretation and reporting of pulmonary function test (PFT) results. Misinterpreting PFT results could lead to misclassification of disease class and severity, which would negatively impact patient care. https://bit.ly/4gVyMGr

Cite this article as: Ong-Salvador R, Laveneziana P, de Jongh F. ERS/ATS Global Lung Function Initiative normal values and classifying severity based on z-scores instead of per cent predicted. *Breathe* 2024; 20: 230227 [DOI: 10.1183/20734735.0227-2023].

Received: 16 May 2024
Accepted: 1 Oct 2024

Abstract

Among the most discussed issues in the field of pulmonary function testing are reference equations and standardised methods of interpreting pulmonary function test (PFT) results. The Global Lung Function Initiative (GLI) task force gathered information from multiple countries with diverse ethnicities to create a multi-ethnic reference formula suitable for a broad age range. Their first output was the GLI-2012 spirometry reference equations. Together with the GLI-2012 reference equations, the GLI document recommended the use of z-scores to standardise the interpretation of PFT results. Z-scores were introduced as an alternative to per cent predicted and fixed thresholds in interpreting PFTs. After more than a decade, there is still debate on which method of interpretation is better.

Educational aims

- To provide an overview of the strengths and limitations of the GLI-2012 reference equations.
- To re-evaluate the clinical use of fixed ratios and cut-off values in classifying pulmonary impairment.
- To discuss the controversies surrounding the GLI reference values and the use of z-scores in interpreting PFT results.

Introduction

Pulmonary function tests (PFTs) are an important diagnostic tool in pulmonary medicine, aiding clinicians in diagnosing pulmonary problems and assessing airway obstruction [1]. Spirometry, the most basic PFT, is simple yet reliable in measuring airway obstruction and providing prognostic information. Spirometry has been in use for centuries in pulmonary medicine and clinical research [2, 3]. The days of measuring air from a person seated in a wine barrel (Robert Menzies, 1799) and sucking water from an upside-down graded bell jar (Edward Kentish, 1844) have long since passed in the field of pulmonary function testing [3]. However, past knowledge gained through basic instruments remains crucial today, such as the correlation between age, height, forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV_1). Therefore, the focus in recent years has shifted towards addressing issues like the lack of consensus in reference equations and standardising PFT interpretations [4].

To standardise PFT interpretation, a universally applicable reference equation is required, regardless of age or ethnicity [5]. The Global Lung Function Initiative (GLI) 2012 reference equation, derived from huge global data sets, offers robust, relevant and reliable reference values for PFT interpretation. It covers a wide age range and is based on multi-ethnic data [1]. Additionally, the GLI task force suggests interpreting PFTs using z-scores, which could be the first step toward global PFT standardisation [1, 6, 7]. Because the concept of z-scores is more difficult to understand for both clinicians and patients, it is still not widely used.

Here we address the following topics: 1) the importance of using a reliable reference equation when evaluating spirometry results; 2) the strengths and limitations of the GLI reference equations and the challenges that go with them; 3) the classification of airway obstruction; and 4) the practical implications of the GLI proposed changes in pulmonary function testing.

Reference equations in pulmonary function testing

Historically, spirometry indices have shown variations across different sexes, ages, heights and ethnicities [8, 9]. The standard procedure for evaluating PFT results involves comparing the measured value to the reference value for the patient based on their age, height, sex and ethnicity. Due to the different biological factors that are included in the PFT reference equations, we could conclude that the quality of the data gathered from the reference population is an important factor in choosing the best reference equation for interpreting PFT results.

There have been initiatives to improve spirometry measurement accuracy, precision and quality, as well as patient experience [4]. However, even if a test result is of the highest quality, it may only be clinically meaningful if evaluated in conjunction with a robust, trustworthy and relevant reference value [6]. There are numerous reference equations for PFTs accessible; however, there has been no consensus on which equation to use due to variances in approach and data quality [6]. Furthermore, many reference equation sets for children and adults did not match once a youngster became an adult, at around the age of 18 years. The GLI task force was established to develop international spirometry reference equations to address the lack of consensus [1].

GLI-2012 reference equations

The most frequently asked question is "What makes the GLI-2012 reference equations better than the other equations that the clinics have been using for years?" In this section, we discuss the strengths and limitations of the GLI-2012 reference equations.

Strength in numbers

One of the strengths of the GLI-2012 reference values lies in the large sample size, which provides a better representation of different ethnicities and age groups [1]. The GLI-2012 multi-ethnic reference equation was based on quality data sets collected from 72 centres in 33 countries worldwide, and these data went through a selection process following strict quality control criteria [1]. It offers the most reliable and representative reference equation sets for the 3–95-years old age range.

Prediction model

The GLI-2012 prediction models were derived from healthy individuals aged 3–95 years from different ethnic groups [5]. These prediction models were developed using the LMS method (L for lambda or location, which is an index of skewness; M for mu or mean value; and S for sigma or scatter, which is the coefficient of variation). This mathematical method allows for continuous prediction equations, adjusts for differential dispersion with age or height, and takes into consideration distribution asymmetry [10]. Furthermore, the GLI-2012 prediction models use splines to provide a smooth fit across the full age range and a better curve fit during the transition from adolescence to adulthood [1]. Using the same method, lower limits of normal for spirometric indices were created.

All-age equation

The lack of a proper equation that covers a wide age range necessitates the patching together of different reference equations to be able to cover all ages. Stitching together different reference equations would create arbitrary breakpoints that increase the potential for misinterpretation in young children and during the transition from paediatrics to adulthood [11–13].

In childhood, the FVC outgrows FEV_1, leading to falls in FEV_1/FVC, which is reversed during adolescence [12]. This trend can be seen in figure 1e and f as a small "bump" in FEV_1/FVC values around

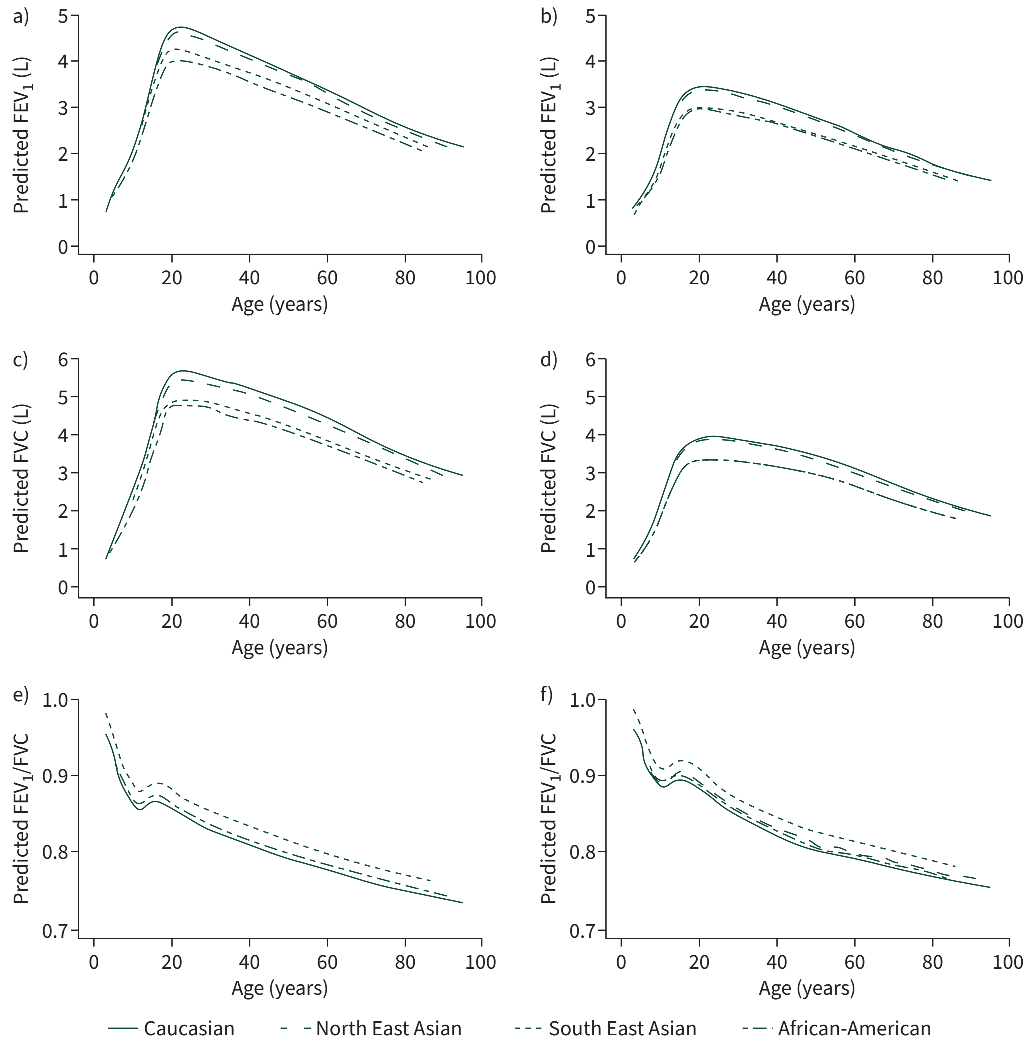

FIGURE 1 Predicted values for a, b) forced expiratory volume in 1 s (FEV_1), c, d) forced vital capacity (FVC) and e, f) FEV_1/FVC by sex and ethnic group. a, c, e) Males and b, d, f) females. Graphs were generated using the mean height for age in Caucasians to illustrate proportional differences between ethnic groups of the same height and age; in practice, differences in height for age further affect predicted values. Reproduced from [1] with permission.

adolescence [1, 13, 14]. We are now able to illustrate this pattern because of the statistical techniques used in formulating the GLI-2012 reference equations. This was not possible with the statistical techniques used in the past [1].

Unlike during childhood and adolescence, lung ageing is expected in adulthood. In the past, due to the lack of appropriate reference values for individuals older than 65 years, values extrapolated from existing data were used. With life expectancy increasing, the number of elderly patients attending our clinics also increases. It is then critical to be able to detect airway anomalies early on and avoid misdiagnosis. The wide age representation of the GLI-2012 reference equation is a big step towards addressing this problem [1, 15]. However, data from persons aged >75 years are limited (figure 2) and needs to be increased to improve the diagnostic accuracy of lung function in the elderly. More data from healthy adults aged >80 years are still needed to develop an updated age-appropriate reference equation, otherwise this population will remain underrepresented [5]. Another problem with data collected from individuals older than 80 years of age is survival bias. Admittedly, better representation of this age range from various ethnicities would continue to be challenging to gather.

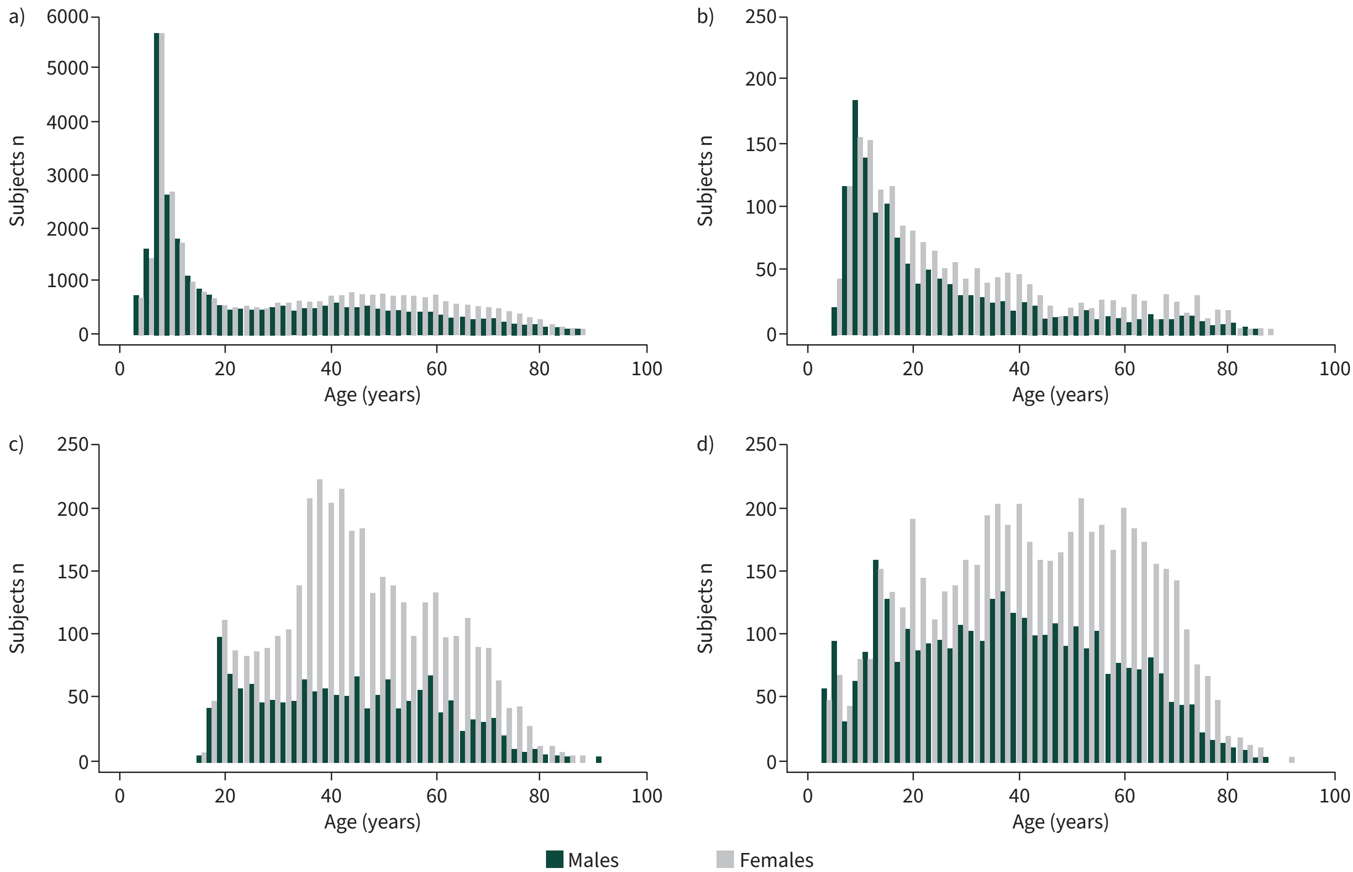

FIGURE 2 Age distribution in the final sample (n=74 187) of a) Caucasians, b) African Americans, c) Northeast Asians and d) Southeast Asians. Reproduced from [1] with permission.

Race/ethnicity

An important part of standardising PFT interpretation is to have a reference equation that can be used globally. The multi-ethnic GLI-2012 reference equation included 33 countries from five continents, thereby including several ethnic groups. It is recognised that this is far from complete. The GLI "other" equation was developed to accommodate patients of mixed descent [1, 5].

The use of ethnicity in the prediction model has received widespread criticism. Some contend that ethnicity is a social construct and should not be used to interpret PFT results. Ethnicity may not accurately represent other factors, such as socioeconomic factors, which might affect PFT results. Using race-specific equations could increase the risk of underdiagnosing respiratory diseases, while ignoring race could widen the normal range and cause pathophysiologically reduced lung function to go undiagnosed in some individuals [16].

How did the concept of racial and ethnic distinctions in PFTs come about? Since the 1800s, there has been a difference in lung function results based on race and ethnicity. Previous investigations have revealed that Caucasians have a slightly larger lung capacity than other ethnicities [17–19]. In 1925, a PFT reference manual for clinicians defining racial differences was produced [20].

Alternatives include a race-neutral equation [21] or FEV_1 quotient (FEV_1Q) [22]. Using the GLI-2012 and the US National Health and Nutrition Examination Survey (NHANES) III datasets, Bowerman *et al.* [21] developed a race-neutral equation, the GLI global 2022, which was similar to the GLI-2012 for FEV_1, FVC and FEV_1/FVC. This serves as a benchmark to predict a range of spirometry values that are expected for a given height, age and sex. The GLI global 2022 equation showed good agreement with the GLI "other" equation [21]. Balasubramanian *et al.* [22] suggested the use of FEV_1Q as an alternative to current practice. The FEV_1Q could be used to predict mortality and had better discriminative performance compared to currently recommended interpretation strategies [22].

Birth sex

It is a well-known fact that sex is one of the important predictors of lung function. The difference in lung growth and development between males and females begins in the prenatal stage. Female lungs mature quicker but have fewer bronchioles than boys at birth [23]. For biological purposes, the PFT reference values are calculated based on birth sex [24]. On the other hand, gender identification is a person's perception of their gender. Gender identity is not always in line with the biological markers an individual is born with. However, given recent scientific advancements and the sensitive nature of this subject, the patient's gender identification is an important aspect that must be recognised. No prediction model for PFT incorporates non-binary reference ranges [25]. The electronic patient health record should allow for gender identification, saving the patient the burden of explaining themselves at each appointment. A legal gender change could be specified using a separate information field.

In one study on the transgender population, the use of male-predicted values for FEV_1 and FVC for a female-sized body resulted in pseudo-restriction, and the reverse scenario may mask true restriction. However, the gender assignment did not affect the FEV_1/FVC [25]. Another study showed an overestimation of lung function in a trans woman [26]. Further studies are needed to fully address the need for proper gender classification or the lack thereof.

Classifying airway obstruction

It is standard practice in PFTs to compare the obtained spirometry result to the chosen reference value, in this case, GLI-2012. Traditionally, the comparison was done using a percentage predicted and a fixed cut-off value of 0.7 for FEV_1/FVC. The 2005 European Respiratory Society (ERS) statement [5] recommended using the lower limit of normal, and in 2019 [4] the use of standard residuals was introduced. In this section, we discuss these methods of classifying airway obstruction.

Percentage predicted

Spirometry has been around for a century. It is not surprising to see that certain practices have not changed over time. An example is interpreting PFT results based on the percentage predicted. The percentage predicted is calculated by taking the absolute values of FEV_1 and FVC in litres and dividing by the reference predicted value multiplied by 100 [6]. This practice began due to convenience and was later perceived as a good "rule of thumb". However, this is unfounded and lacks scientific support [7, 9, 27]. Although questioned in the past [9, 28], the practice continued because the concept is easy to understand for patients and clinicians.

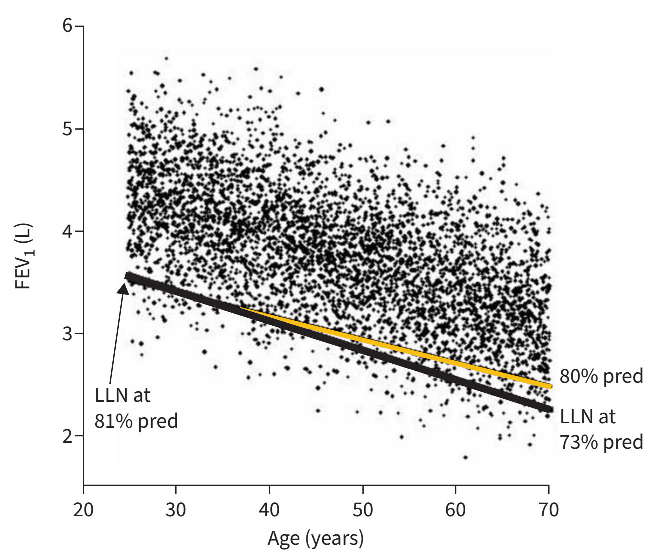

FIGURE 3 Plot of idealised population data for forced expiratory volume in 1 s (FEV_1): 80% predicted and lower limit of normal (LLN). FEV_1 is plotted against age for men of height 1.77 m showing a yellow line representing 80% of the predicted value and a black line showing the true LLN, that is, the fifth centile value 25–70 years [9]. Reproduced from [29] with permission.

When we use the percentage of predicted, we assume that the deviation of one value can be compared with the deviation from another [9]. However, the natural scatter of lung function data within the healthy, normal population is not proportional to the mean value (figure 3) [6]. If 80% of predicted values are used as the lower limit of normal, 20% of subjects would be misclassified for the likelihood of disease [8]. Using per cent prediction leads to age- and height-related bias and should be discontinued [6].

Fixed cut-off values

Using cut-off values for FEV_1/FVC is another practice in need of attention [1, 7]. The major problem with using a fixed threshold is that it does not account for the natural variability in spirometry results of healthy adults. Identifying a normal PFT result should not be based on a fixed cut-off value. This is especially true when classifying the severity of airway obstruction in COPD patients. It is not reasonable to classify a patient with a disease simply because the PFT result falls below the stated threshold while other test results remain within the normal range [30]. Because FEV_1/FVC declines with age in both sexes, using a fixed ratio is scientifically incorrect and would result in overdiagnosis of obstruction in the elderly (especially men) and underdiagnosis in younger people (especially women) [31]. The GLI document illustrated that using 80% as the cut-off for an abnormal result led to a high false-positive rate of being classified as having airway obstruction, especially in patients over 50 years of age (figure 4). This is because the normal range is usually wider than the traditionally used 80–120% [7]. This is especially true for the elderly.

However, a recent publication comparing FEV_1/FVC to FEV_1 % predicted to classify the severity of airflow limitation in terms of exertional breathlessness demonstrated that using the observed FEV_1/FVC ratio, rather than normative reference equations, may be more appealing for countries without validated or frequently updated spirometry equations. They also emphasised that the crude FEV_1/FVC is more stable than the FEV_1 % predicted across different populations and ethnic groups [32].

Z-scores

Assuming all data gathered from typical participants follows a normal distribution, the data points would form a bell-shaped curve (figure 5). The mean would be found at the centre of the bell and would have a z-score of zero [1]. This is considered to be the predicted value. The z-score is the number of standard deviations (SD) in which a parameter deviates from the centre of the bell. As demonstrated in figure 5, a z-score of −1.64 is the lower limit (fifth percentile), and +1.64 is the upper limit of normal (95th percentile). This would mean that 90% of the data (z-scores of −1.64 to +1.64) is statistically considered to be normal. However, when applying the z-scores to individuals with no symptoms and no known risk factors, −1.96 z-scores should be used [7].

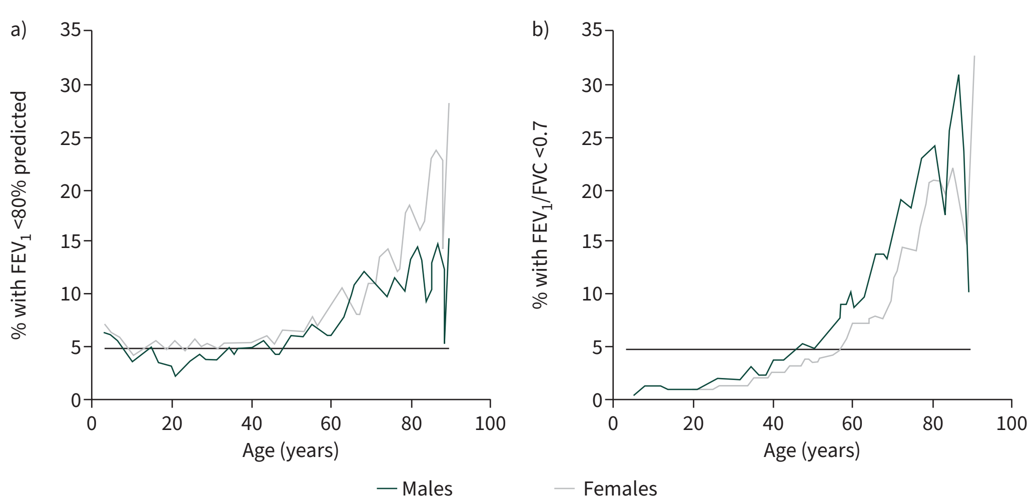

FIGURE 4 The percentage of healthy subjects in whom a) forced expiratory volume in 1 s (FEV_1) is <80% predicted or b) the FEV_1/forced vital capacity (FVC) ratio is <0.7 [1]. As can be seen, the use of 80% as a fixed threshold for FEV_1 leads to a high percentage of false positives in the elderly, while the use of <0.7 as the threshold for abnormal FEV_1/FVC will lead to underdiagnosis of airway obstruction in the young and overestimation in the elderly. Reproduced from [7] with permission.

https://doi.org/10.1183/20734735.0227-2023

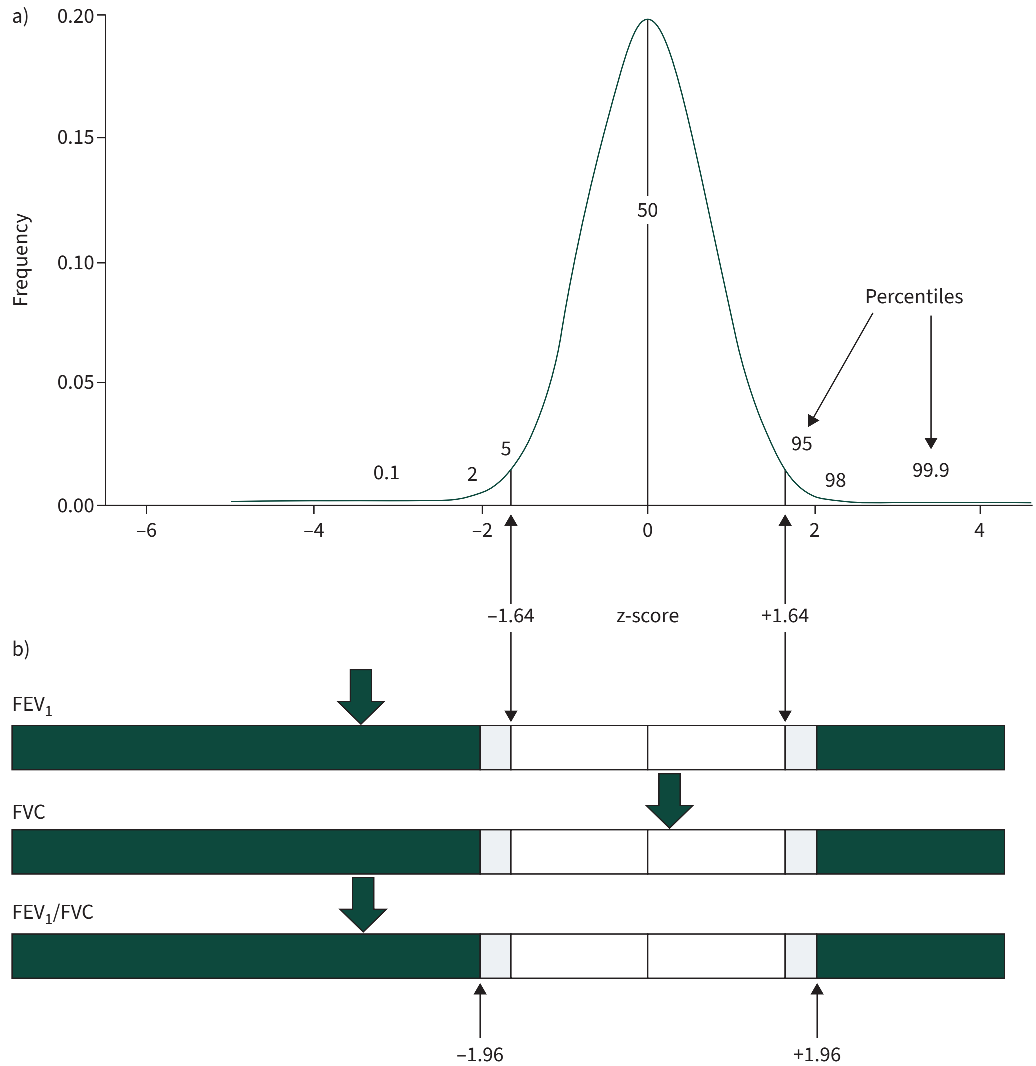

FIGURE 5 a) Illustration of how z-scores relate to percentiles. b) Pictogram demonstrating the normal range. The arrows indicate how far a value is from the normal range. The 50th percentile is equivalent to 100% predicted [1, 6]. Light shading indicates the mean of −1.64 z-score (LLN 5%) and +1.64 z-score (ULN 5%); darker shading indicates the mean of −1.96 z-score (LLN 2.5%) and +1.96 z-score (ULN 2.5%). LLN: lower limit of normal; ULN: upper limit of normal. Reproduced and modified from [1] with permission.

It has been a long-standing goal to standardise PFT interpretation. Using the z-score is the most statistically correct approach to standardise the interpretation of PFT results [1, 6, 12]. This method compares the measured PFT result and gives the probability of a healthy person having an abnormal test result regardless of age, height, gender and ethnic group. This provides a bias-free interpretation of serial measurements within a person during growth [6]. The z-scores indicate how far a measured value is with regard to the predicted value. However, it does not indicate how close a measured value is to the minimum acceptable quality of life or with respect to the mortality score. These limitations in classifying the severity of obstructive lung disease and defining the prognosis are why the z-scores should never be used in isolation [6]. The 2021 ERS/American Thoracic Society (ATS) technical standard recommended that the interpretation of PFT results should focus on airflow, lung volume and gas transfer measurements to recognise patterns of altered physiology; it should not be used to diagnose a specific pathologic condition [33].

Controversies in PFTs

Why is the percentage predicted still used in PFT results after implementing GLI?

The use of percentage predicted is being discouraged, yet it is still widely used in the clinic. This is because using the percentage predicted values is not wrong as long as there is no fixed cut-off value used to define abnormalities. Arguably, it is still much easier for patients to grasp the idea of per cent prediction rather than the concept of z-scores. Most clinicians and patients have been trained or conditioned to look at their percentage predicted PFT values. As previously mentioned, as long as there is no threshold value used, there is nothing wrong with using the percentage predicted in conjunction with the GLI reference equation. Reporting both the percentage predicted and z-scores would aid in adapting to the z-score in interpreting PFT results.

Bronchodilator response

Eliminating bias in measurement outcomes is crucial for monitoring and evaluating therapy [7]. The response to a bronchodilator is measured with spirometry before and after administration of the chosen bronchodilator. The recent guidelines recommend reporting the response as an increase relative to the predicted value to eliminate bias based on height and sex [1, 33–35]. There are conflicting publications following this new recommendation. Reanalysing previously measured data using the 2005 recommendation, some reported a decrease in significant bronchodilator response (BDR+), while others reported up to a 32% increase in BDR+ using the 2021 recommendation [34, 35]. This shift in classification could lead to a change in the prescription of medication and could possibly add an emotional burden to patients. Future research is needed to understand the impact of bronchodilator protocols (*e.g.* delivered dose) on results and on the ability of an acute response to bronchodilators to predict future clinical status other than survival.

Patient education

A practical concern when adapting the GLI reference equation is the possible shift in diagnosis when changing predicted values [11]. Clinicians could inform the patient about the updated reference equations and interpretation guidelines, which could cause a possible shift in severity classification of their lung function. Furthermore, previous PFT results should be reanalysed using GLI reference equations to prevent unexplained changes in the trend reporting of PFT parameters. This is vital for the longitudinal assessment of PFT results.

What is next?

The various misconceptions regarding GLI reference equations have been addressed [7]. Still, there are some clinicians and PFT laboratories who have not implemented the GLI reference equations in their practice. It is an important step to adopt z-scores in clinical guidelines such as Global Initiative for Asthma and Global Initiative for Chronic Obstructive Lung Disease. Future studies should be performed on evaluating patients using z-score with the 2022 interpretation guideline to investigate if this new method would indeed lead to a better diagnosis and/or treatment.

Key points

- There is continuous effort being made to standardise the interpretation and reporting of PFT results.
- Misinterpreting PFT results could lead to misclassification of disease class and severity, which would negatively impact patient care.

Self-evaluation questions

1. Is the following statement true or false?
   Interpreting PFT results using the z-score is based on statistics. This means that when data collected from normal subjects are normally distributed, it would follow a bell curve, with the top of the bell representing the 100%. The z-score gives the number of standard deviations (SD) a measured value is from the centre of the bell curve. PFT is mostly concerned about the lower limit of normal, meaning we look at a z-score of −1.64. So, a healthy volunteer with no symptoms and no history of smoking and a PFT result of −1.96 z-score would have a lung function that is outside the normal range.
2. One of the major concerns in adapting the GLI reference equations was that not all spirometry parameters have a reference equation. Which parameter is this?
   a) FVC
   b) FEV_1
   c) FEV_1/FVC
   d) PEF
3. What are the possible consequences that could arise when using a fixed cut-off value for the FEV_1/FVC ratio?
   a) Airway obstruction in young adults would be underdiagnosed.
   b) Higher number of false positive results for airway obstruction in elderly subjects.
   c) The fixed threshold of 0.7 for FEV_1/FVC is sometimes not attained before 50 years of age in men and later in women.
   d) All of the above.

4. Which PFTs have GLI reference equations based only on Caucasian population?
   a) Spirometry
   b) Lung volumes
   c) Transfer factor of the lung for carbon monoxide

Conflict of interest: R. Ong-Salvador has nothing to disclose. P. Laveneziana reports payment or honoraria for lectures, presentations, manuscript writing or educational events from CHIESI, AstraZeneca and GSK, outside the submitted work. F. de Jongh reports payment or honoraria for lectures, presentations, manuscript writing or educational events from European Respiratory Society; support for attending meetings from European Respiratory Society; and a leadership role with the European Respiratory Society, all disclosures made outside the submitted work.

References

1 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.

2 Kouri A, Dandurand RJ, Usmani OS, *et al.* Exploring the 175-year history of spirometry and the vital lessons it can teach us today. *Eur Respir Rev* 2021; 30: 210081.

3 Gibson GJ. Spirometry: then and now. *Breathe* 2005; 1: 206–216.

4 Graham BL, Steenbruggen I, Miller MR, *et al.* Standardization of spirometry 2019 update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med* 2019; 200: e70–e88.

5 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.

6 Cooper BG, Stocks J, Hall GL, *et al.* The Global Lung Function Initiative (GLI) Network: bringing the world's respiratory reference values together. *Breathe* 2017; 13: e56–e64.

7 Stanojevic S, Quanjer P, Miller MR, *et al.* The Global Lung Function Initiative: dispelling some myths of lung function test interpretation. *Breathe* 2013; 9: 462–474.

8 Hutchinson J. On the capacity of the lungs, and on the respiratory function, with a view to establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Trans* 1846; 29: 137–252.

9 Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265–267.

10 Cole TJ, Stanojevic S, Stocks J, *et al.* Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. *Stat Med* 2009; 28: 880–898.

11 Quanjer PH, Brazzale DJ, Boros PW, *et al.* Implications of adopting the Global Lungs Initiative 2012 all-age reference equations for spirometry. *Eur Respir J* 2013; 42: 1046–1054.

12 Stocks J, Kirkby J, Lum S. How to avoid misinterpreting lung function tests in children: a few practical tips. *Paediatr Respir Rev* 2014; 15: 170–180.

13 Kirkby J, Aurora P, Spencer H, *et al.* Stitching and switching: the impact of discontinuous lung function reference equations. *Eur Respir J* 2012; 39: 1256–1257.

14 Quanjer PH, Stanojevic S, Stocks J, *et al.* Changes in the FEV_1/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010; 36: 1391–1399.

15 Quanjer PH, Stocks J, Cole TJ, *et al.* Influence of secular trends and sample size on reference equations for lung function tests. *Eur Respir J* 2011; 37: 658–664.

16 Bhakta NR, Kaminsky DA, Bime C, *et al.* Addressing race in pulmonary function testing by aligning intent and evidence with practice and perception. *Chest* 2022; 161: 288–297.

17 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.

18 Braun L, Wolfgang M, Dickersin K. Defining race/ethnicity and explaining difference in research studies on lung function. *Eur Respir J* 2013; 41: 1362–1370.

19 Braun L. Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics. Minneapolis, University of Minnesota Press, 2014.

20 Myers JA. Vital Capacity of the Lungs: A Handbook for Clinicians and Others Interested in the Examination of the Heart and Lungs both in Health and Disease. Baltimore, Williams & Wilkins, 1925.

21 Bowerman C, Bhakta NR, Brazzale D, *et al.* A race-neutral approach to the interpretation of lung function measurements. *Am J Respir Crit Care Med* 2023; 207: 768–774.

22 Balasubramanian A, Wise RA, Stanojevic S, *et al.* FEV_1Q: a race-neutral approach to assessing lung function. *Eur Respir J* 2024; 63: 2301622.

23 LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe* 2018; 14: 131–140.

24 Foer D, Rubins D, Almazan A, *et al.* Gender reference use in spirometry for transgender patients. *Ann Am Thorac Soc* 2021; 18: 537–540.

25 Al-Hadidi N, Baldwin JL. Spirometry considerations in trangender patients. *J Allergy Clin Immunol* 2017; 139: AB198.

26 Bates DV, Christie RV, eds. Respiratory Function in Disease. Philadelphia, WB Saunders, 1964.
27 Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966; 21: 445–449.
28 Miller MR. What defines abnormal lung function? *Thorax* 2007; 62: 1107.
29 Sylvester KP, Clayton N, Cliff I, *et al*. ARTP statement on pulmonary function testing 2020. *BMJ Open Respir Res* 2020; 7: e000575.
30 Swanney MP, Ruppel G, Enright PL, *et al.* Using the lower limit of normal for the FEV_1/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008; 63: 1046–1051.
31 Ward H, Cooper BG, Miller MR. Improved criterion for assessing lung function reversibility. *Chest* 2015; 148: 877–886.
32 Backman H, Vanfleteren LEGW, Mannino DM, *et al.* Severity of airflow obstruction based on FEV_1/FVC *versus* FEV_1 percent predicted in the general US population. *Am J Respir Crit Care Med* 2024; 210: 1308–1316.
33 Stanojevic S, Kaminsky DA, Miller MR, *et al.* ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022; 60: 2101499.
34 Li Y, Lin J, Wang Z, *et al.* Bronchodilator responsiveness defined by the 2005 and 2021 ERS/ATS criteria in patients with asthma as well as chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2022; 17: 2623–2633.
35 Bhatt SP, Fortis S, Bodduluri S. New guidelines for bronchodilator responsiveness in COPD: a test in search of a use. *Am J Respir Crit Care Med* 2022; 206: 1042–1044.

Suggested answers

1. False. A PFT is referred only when indicated. So, when a healthy volunteer with no symptoms and no history of smoking performs a test, the z-score that should be used is −1.96 and not −1.64 [1].
2. d.
3. d.
4. b and c.