

## Recommendations for a Standardized Pulmonary Function Report An Official American Thoracic Society Technical Statement

Bruce H. Culver, Brian L. Graham, Allan L. Coates, Jack Wanger, Cristine E. Berry, Patricia K. Clarke, Teal S. Hallstrand, John L. Hankinson, David A. Kaminsky, Neil R. MacIntyre, Meredith C. McCormack, Margaret Rosenfeld, Sanja Stanojevic, and Daniel J. Weiner; on behalf of the ATS Committee on Proficiency Standards for Pulmonary Function Laboratories

THIS OFFICIAL TECHNICAL STATEMENT OF THE AMERICAN THORACIC SOCIETY WAS APPROVED OCTOBER 2017

**Background:** The American Thoracic Society committee on Proficiency Standards for Pulmonary Function Laboratories has recognized the need for a standardized reporting format for pulmonary function tests. Although prior documents have offered guidance on the reporting of test data, there is considerable variability in how these results are presented to end users, leading to potential confusion and miscommunication.

**Methods:** A project task force, consisting of the committee as a whole, was approved to develop a new Technical Standard on reporting pulmonary function test results. Three working groups addressed the presentation format, the reference data supporting interpretation of results, and a system for grading quality of test efforts. Each group reviewed relevant literature and wrote drafts that were merged into the final document.

**Results:** This document presents a reporting format in test-specific units for spirometry, lung volumes, and diffusing capacity that can be assembled into a report appropriate for a laboratory's practice. Recommended reference sources are updated with data for spirometry and diffusing capacity published since prior documents. A grading system is presented to encourage uniformity in the important function of test quality assessment.

**Conclusions:** The committee believes that wide adoption of these formats and their underlying principles by equipment manufacturers and pulmonary function laboratories can improve the interpretation, communication, and understanding of test results.

**Keywords:** pulmonary function testing; reporting spirometry; reference equations; pulmonary function quality grading

<b>Contents</b>		
Overview	<b>Diffusing Capacity (Transfer Factor)</b>	<b>Reference Source Recommendations</b>
Conclusions	Comments and Interpretation	Grading the Quality of Pulmonary Function Tests
Introduction	Selecting and Reporting Reference Values	Spirometry
Methods	General Considerations	Lung Volumes
Report Format for Spirometry and Other Lung Function Tests	Current Spirometry Reference Values	Diffusing Capacity (Transfer Factor)
General Considerations	Using Reference Data in Interpretation of Results	The Quality Reviewer
Spirometry		Conclusions
Tests of Lung Volume		

Supported by a project grant from the American Thoracic Society.

Correspondence and requests for reprints should be addressed to Bruce H. Culver, M.D., Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington School of Medicine, Campus Box 356522, Seattle, WA 98195-6522. E-mail: bculver@uw.edu

This article has an online supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org).

Am J Respir Crit Care Med Vol 196, Iss 11, pp 1463–1472, Dec 1, 2017

Copyright © 2017 by the American Thoracic Society

DOI: 10.1164/rccm.201710-1981ST

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## Overview

The American Thoracic Society Committee on Proficiency Standards for Pulmonary Function Laboratories (ATS PFT Committee) has been concerned about the wide variability in pulmonary function test (PFT) reports among laboratories and has discussed the need for a more standardized format, to include information to assist accurate interpretation and to enhance the communication of results to end users. ATS support was granted to develop a technical standard to address this need and also to update reference sources and to propose a standardized quality grading system.

## Conclusions

- A uniform format for the presentation of PFT results in reports to users and in the medical record can reduce potential miscommunication or misunderstanding.
  - Only information with validated clinical application should be included.
  - The normal limit(s) of each test parameter should be displayed.
  - Consistent with other laboratory values, the measured value should be shown before reference values, ranges, or normal limits.
  - Report and/or display of the displacement of the result from a predicted value in standard deviation units ( $z$ -score) can help in understanding abnormality.
- For spirometry, many parameters can be calculated but most do not add clinical utility and should not be routinely reported.
  - Only FVC, FEV<sub>1</sub>, and FEV<sub>1</sub>/FVC need be routinely reported.
  - Measurement of slow VC and calculation of FEV<sub>1</sub>/VC are a useful adjunct in patients with suspected airflow obstruction.
  - Reporting FEV<sub>1</sub>/FVC (or FEV<sub>1</sub>/VC) as a decimal fraction, and not reporting it as a percentage of the predicted value for this ratio, will help to minimize miscommunication.
- Lung volumes
  - The nitrogen washout plot for multibreath tests and the tracings for plethysmograph tests can be shown graphically to aid quality assessment.
- For diffusing capacity the report is consistent with the 2017 European Respiratory Society (ERS)/ATS Technical Standard for this test.
  - Barometric pressure should be measured and reported and the measured value corrected to the standard pressure of 760 mm Hg.
- Newer collated reference equations for spirometry and diffusing capacity have been developed since prior ATS documents and warrant wide implementation.
  - The Global Lung Function Initiative (GLI)-2012 multiethnic spirometry reference values are recommended for use in North America and elsewhere for the ethnic groups represented. Their smooth continuity throughout growth is advantageous for laboratories testing children or adolescents.
  - The National Health and Nutrition Examination Survey (NHANES) III reference values (recommended for North America in 2005 ATS/ERS documents) remain appropriate where maintaining continuity is important.
  - Regardless of the reference source or lower limit of normal (LLN) chosen, interpreters should be aware of uncertainty when interpreting values near any dichotomous boundary.
  - For lung volumes and diffusing capacity, no prior ATS recommendation has been made because of the wide divergence of available reference values. A large compilation of international data has been completed for the diffusing capacity of the lung for carbon monoxide (DL<sub>CO</sub>) and is underway for lung volumes. The resulting reference equations should be widely adopted when published.
- Pulmonary function tests that fail to meet optimal standards may still provide useful information. A grading system for test quality can allow for this use, while providing an indication of the uncertainty imposed, and is most helpful if widely standardized.
  - For spirometry, FVC and FEV<sub>1</sub> are graded separately on an A–F scale either manually or by software. There is evidence that grades A–C are clinically useful, whereas grades D and E may have limited value, and grade F should not be used. The same scale,

with different criteria values, is used for children.

- For diffusing capacity a similar grading scale is presented on the basis of 2017 ERS/ATS standards.

## Introduction

The range of reporting formats currently in use is wide; commercial PFT systems offer differing reports, and some clinical laboratories customize their own. Differently arranged reports can lead to confusion or errors and make comparisons of data from different laboratories unnecessarily difficult. PFT equipment manufacturers have expressed a desire for, and a willingness to implement, a standardized form once it has been established. Newer reference data for spirometry and diffusing capacity have become available since the publication of prior guidelines, and a standardized system for grading the quality of lung function tests would be desirable.

## Methods

For several years the ATS PFT Committee has been discussing and sharing ideas for improvement in the reporting of PFT results. A project task force, consisting of the committee as a whole, was approved to develop this new technical standard. The committee included adult and pediatric pulmonologists and physiologists and respiratory therapists with extensive PFT experience. Three working groups addressed the presentation format, the reference data supporting interpretation of results, and a system for grading quality of test efforts. Each group reviewed relevant literature and wrote drafts that were merged into the final document. As there is rather limited literature to support the necessary choices, these were made by consensus; all members approved the final document.

## Report Format for Spirometry and Other Lung Function Tests

### General Considerations

The following recommendations and rationale are based on developing a format that will be intuitive, will include only information with validated clinical application, will be based on the use of

the LLN, and will be consistent with prior recommendations for PFT interpretation and reporting (1–4). Some recommendations are necessarily arbitrary (e.g., the order of rows or columns) but reflect a consensus of current and prior committee members and an informal survey of others (5). Although individual preferences vary, there is wide agreement that the benefit of uniformity outweighs these.

The report format recommended is presented in test-specific units that can be assembled into a report appropriate for a laboratory's practice or even an individual test session. It is designed so that for simple testing it can be printed, along with interpretive comments, on a single page as a report to a referring physician or for inclusion in the medical record. Of necessity, this contains limited information and is not intended as the only resource for the interpreter, who should have the option of displaying all individual maneuvers from a given PFT session, increasingly done on digital systems. Standardized electronic formats for the saving of all PFT data, including each individual maneuver, are being recommended (6). This will allow reviewers the flexibility to see additional detail or to reanalyze previous PFTs or apply new reference values as they become available. A standardized methodology to incorporate PFT data into electronic medical records is needed, but is beyond the scope of this report. See Appendix EA in the online supplement for a suggested list of test results to save to the electronic medical record.

In designing the standardized report, the committee recognized that aspects of data presentation can affect decision-making (7). The use of boldface or colored fonts to highlight measured values below the LLN can draw attention to these, but imposes a binary decision on a continuous variable. The number of variables reported can also have an impact because including a large number of outcomes in the report increases the statistical likelihood of one falling below an arbitrary LLN, with the risk of a false positive result (8).

All reports must begin with unambiguous patient identification, including patient name, medical record number, sex, and date of birth; the latter can be compared with previous records as a check for possible identification errors, as

well as for calculating patient age (year to one decimal place for children and adolescents, e.g., age 6.3 yr) (9, 10). Other essential information is height (to the nearest centimeter) and weight, ethnicity, and date of the test. Other useful information includes smoking history, reason for the test, and referring physician's name. Additional information may include oxygen saturation and barometric pressure.

The display will vary with the testing done, but the suggested order is spirometry, slow vital capacity, and/or lung volume measurement, and diffusing capacity of the lung for carbon monoxide ( $D_{LCO}$ ). Other tests could be added such as forced oscillometry, maximal respiratory system pressure, levels of expired nitric oxide, or other tests, but the philosophy should be similar, that is, reference source, normal limits, graphs that convey quality information, and exclusion of information without clinical value.

The recommended order of the columns in tabular data is the actual value, the LLN, the  $z$ -score (optional), and the percent predicted value. The predicted value itself is unnecessary, as it does not aid in the interpretation of abnormality. The  $z$ -score of a result is the number of standard deviations it lies away from the mean or, for regression equations, the number of standardized residuals away from the predicted value. Linear graphical displays visualize this in relationship to the normal range and assist in assessing the significance of abnormal values (11, 12). (If newly introduced to the reports, adding a brief explanation may be helpful.) The reference source from which the LLN and percent predicted value are derived must be listed, and whether or not these are adjusted or specific for race/ethnicity must also be stated in technician comments.

### Spirometry

As shown in Figure 1, numerical values are given only for the  $FEV_1$ , the FVC, and the  $FEV_1/FVC$  ratio; the latter should be reported as a decimal fraction and the space for percent predicted value left blank to minimize miscommunications. When appropriate, an additional row can be added for  $FEV_1/(slow) VC$  (1, 2). Forced expiratory time (FET) is reported to aid quality assessment. If bronchodilators are given, the LLN column need not be repeated; the absolute and percent change should be given only for  $FEV_1$  and FVC. Other numerical

values such as the forced inspiratory flow at 75% of FVC ( $FEF_{75\%}$ ) and  $FEF_{25-75\%}$  have not demonstrated added value for identifying obstruction in adults or children, and therefore are not recommended for routine use (13, 14). The flow–volume curve and the volume–time curve are displayed, from which the peak flow and FET can be seen. These graphs must have sufficient resolution to evaluate the quality of the data. For the volume–time curve, the volume scale should be at least 10 mm/L, the time scale at least 20 mm/s, and 1 second prior to the start of expiration should be displayed (2). On the flow–volume plot, the flow display should be at least 5 mm/L/s, and the ratio of flow to volume should be 2 L/s to 1 L. The scales of the graphs may be adjusted to maximize the image within the available space on the report form, especially for tests on small children. The linear analog scales, where the values for  $FEV_1$ , FVC, and their ratio are plotted as  $z$ -scores relative to the predicted value ( $z = 0$ ), give an intuitive sense of severity (12). Because there is always some uncertainty about the application of any prediction to an individual and about the exact LLN, a large star rather than a discrete point is used on the scale to suggest that caution is indicated when interpreting values close to the LLN.

For slow vital capacity, the graph shows baseline tidal breathing to assess whether inspiration occurred from a stable end-expiratory volume (3). The largest vital capacity is reported along with the inspiratory capacity and, when appropriate, the  $FEV_1/VC$ .

### Tests of Lung Volume

Values derived by body plethysmography or gas dilution are displayed with the same column order (Figure 2). We show a full complement of volume parameters listed in a physiologically rational order; however, some laboratories may choose not to report all. With a multibreath nitrogen ( $N_2$ ) washout the graph of the fall in  $N_2$  concentration gives an indication of any leaks present (3). For helium dilution functional residual capacity (no graph displayed), equilibration is considered to be complete when the change in helium concentration is less than 0.02% for 30 seconds. The histogram displays the actual volume increments beside the predicted volumes as an indication of severity, and  $z$ -scores are shown here in a vertical format.

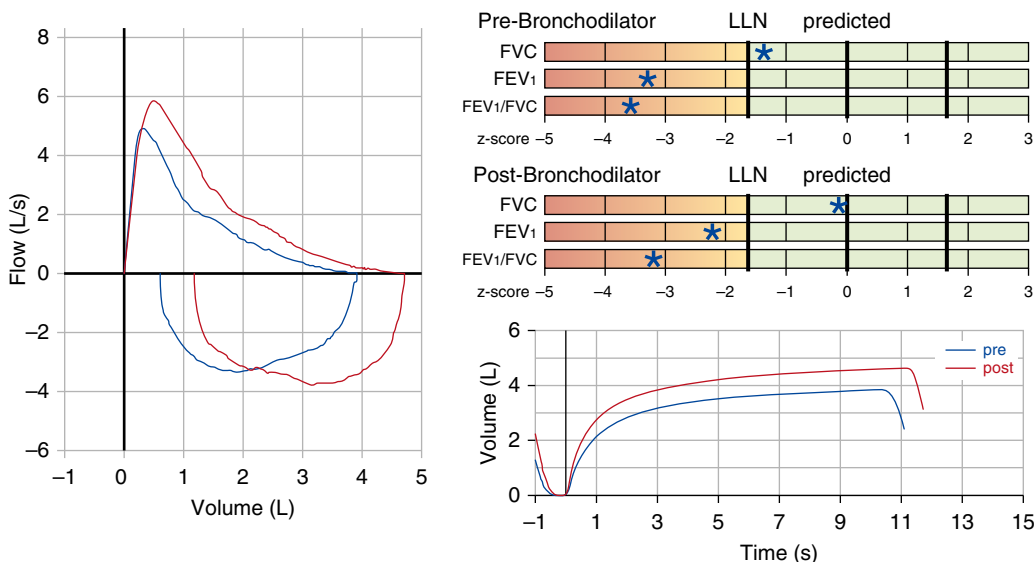
Sample Pulmonary  
Function Laboratory  
  
Anytown, Anywhere  
555-345-6789  
pftests@lunglab.com

Name:	XXXXXXXXXX, YYYYYYYY	Referred by:	Dr. G. Practitioner
ID:	111 222 333	Date of test:	2017-Feb-20 14:30
Sex:	Male	Reason:	Short of breath
Birth date:	1965-Aug-04	SpO2 at rest:	99%
Age:	51	Height:	69 in; 175 cm
Ethnicity:	Caucasian	Weight:	202 lb; 91.8 kg
Smoking:	Ex-smoker	BMI:	30.0 kg/m <sup>2</sup>

**SPIROMETRY**

	Pre-Bronchodilator				Post-Bronchodilator				
	Best	LLN	z-score	%Pred	Best	z-score	%Pred	Change	%Chng
FVC (L)	3.90	3.70	-1.34	82%	4.70	-0.09	99%	600 mL	20%
FEV1 (L)	2.02	2.91	-3.78	54%	2.61	-2.21	70%	590 mL	29%
FEV1/FVC	0.52	0.68	-3.54		0.55	-3.35			
FET (s)	10.3				11.2				

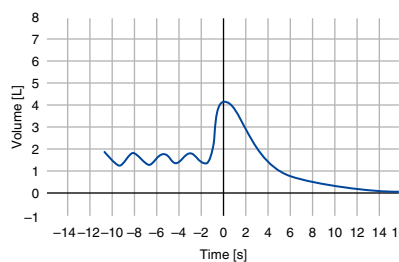
Reference values: GLI 2012 Test quality: Pre: FEV1 - A, FVC - A; Post: FEV1 - A, FVC - B



**SLOW VITAL CAPACITY (Pre-Bronchodilator)**

	Best	LLN	ULN	z-score	%Pred
VC (L)	4.17	4.0	6.0	-1.61	83%
IC (L)	2.76	2.4	3.9		87%
FEV1/VC	0.48	0.68		-3.86	

Reference values: VC - Gutierrez 2004; FEV1/VC - GLI 2012



**TECHNICIAN COMMENTS:** No medications in past 24 hr. 400 mcg albuterol given for reversibility testing.

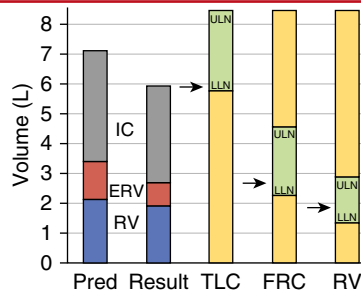
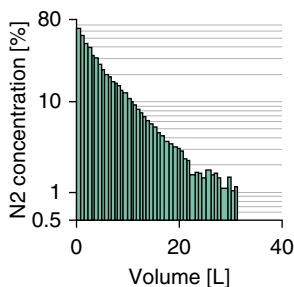
Moderately severe, partially reversible airflow obstruction.

Dr. P. Pulmonologist      2017-Feb-24

**Figure 1.** Example of a single-page report for pre- and postbronchodilator spirometry testing. The linear graphic is divided in units of 1 SD, with the LLN shown at a z-score of -1.64. This simplified report is suitable for the medical record or referring physician, but the test interpreter should have access to the data and curves of all acceptable spirometry efforts. FET = forced expiratory time; GLI = Global Lung Function Initiative; IC = inspiratory capacity; LLN = lower limit of normal; SpO<sub>2</sub> = oxygen saturation as measured by pulse oximetry; ULN = upper limit of normal; VC = vital capacity.

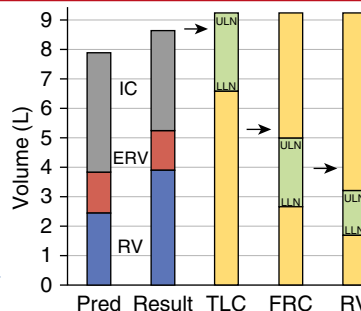
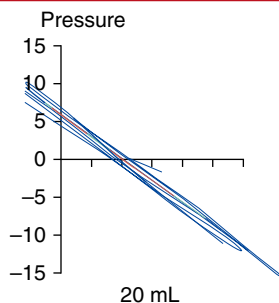
**MULTI-BREATH NITROGEN WASHOUT (Post-Bronchodilator)**

	Result	LLN	ULN	z-score	%Pred
TLC mb (L)	<b>5.95</b>	5.8	8.5	-1.45	83%
VC (L)	<b>4.05</b>	4.0	6.0	-1.61	81%
ICmb (L)	<b>3.27</b>				88%
FRC mb (L)	<b>2.68</b>	2.2	4.6	-1.03	78%
ERV (L)	<b>0.78</b>				61%
RV mb (L)	<b>1.90</b>	1.4	2.9	-0.49	89%
RV/TLC (%)	<b>32%</b>				
Reference values: Gutierrez 2004; Test quality: QA met					



**PLETHYSMOGRAPHY (Post-Bronchodilator)**

	Result	LLN	ULN	z-score	%Pred
TLC pl (L)	<b>8.68</b>	6.6	9.3	0.90	109%
VC (L)	<b>4.79</b>	4.5	6.5	-1.15	87%
IC pl (L)	<b>3.42</b>	2.7	4.1		100%
FRC pl (L)	<b>5.26</b>	2.7	5.0	2.00	137%
ERV (L)	<b>1.29</b>	1.3	2.1		76%
RV pl (L)	<b>3.89</b>	1.7	3.2	3.07	160%
RV/TLC	<b>45%</b>	27%	45%		
Vtg (L)	<b>5.92</b>	3.1	5.5		138%
Reference values: Gutierrez 2004; Test quality: QA met					



**Figure 2.** Examples of the recommended reporting format for lung volume testing in one subject by multi-breath nitrogen (N<sub>2</sub>) washout and in another subject by plethysmography. The N<sub>2</sub> washout is plotted on a log scale, resulting in a nearly linear profile. On this plethysmography tracing the box pressure has been converted to volume to show the thoracic excursions at a scale of 20 ml per division. The bar graphs on the right show the predicted and observed values of RV, FRC, and TLC, and the arrows show these results in relation to their normal range on vertical linear scales. The graphs depict RV in blue, ERV in orange, IC in gray, and normal range in green. (See Figures E1 and E2 in the online supplement for examples of consolidated reports for full pulmonary function tests.) ERV = expiratory reserve volume; IC = inspiratory capacity; LLN = lower limit of normal; mb = multi-breath; pl = plethysmography; QA = quality assurance; RV = residual volume; TLC = total lung capacity; ULN = upper limit of normal; VC = vital capacity; Vtg = thoracic gas volume.

When diffusion capacity is measured, a comparison of total lung capacity measured by both techniques can be a useful quality control measure or an indication of maldistribution.

**Diffusing Capacity (Transfer Factor)**

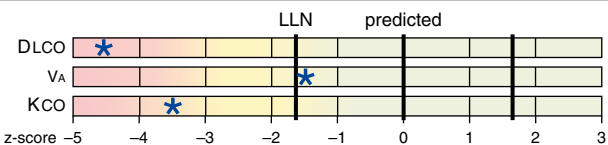
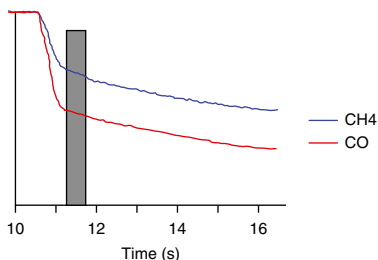
The display (Figure 3) gives the relevant values, the LLN, and the percent predicted

value along with the reference source, a quality assurance indication, and the conditions of the test, in this case post-bronchodilator. The barometric pressure should be given, as well as stating whether the values were corrected to standard barometric pressure (particularly important for laboratories at altitude) (6). Reporting the carbon monoxide transfer coefficient

(K<sub>CO</sub>) is optional, but the term DL/V<sub>A</sub> (the ratio of diffusing capacity to alveolar volume) should be avoided as it is commonly misunderstood. If measured, the hemoglobin should be shown as well as the adjusted predicted values for both DL<sub>CO</sub> and K<sub>CO</sub>. The display shows the washout of both carbon monoxide and the tracer gas and the sample volume. The sample volume

**DIFFUSING CAPACITY (Post-Bronchodilator)**

	Result	LLN	z-score	%Pred
DLCO (mL/min/mmHg)	<b>13.4</b>			
DLCO (at standard P <sub>B</sub> )	<b>13.0</b>	23.4	-4.55	42%
DLCO (pred adj Hb 13.8 g/dL)	<b>13.0</b>			53%
V <sub>A</sub> (L)	<b>5.83</b>	5.75	-1.55	82%
TLCsb (L)	<b>6.01</b>			
V <sub>I</sub> /V <sub>C</sub> (%)	<b>85%</b>			
K <sub>CO</sub> (mL/min/mmHg/L)	<b>2.23</b>	3.25	-3.47	52%
Reference values: GLI 2017; Test quality: one grade A test; P <sub>B</sub> : 721 mmHg				



**Figure 3.** Example of the recommended reporting format for the single-breath diffusing capacity test. The 2017 Technical Standard for DL<sub>CO</sub> (6) requires that the CO and tracer gas concentrations be graphed versus exhaled volume, rather than versus time as shown here. When hemoglobin is measured, it should be shown on the report with a note indicating whether the predicted value has been adjusted for this. CH<sub>4</sub> = methane (tracer gas); DL<sub>CO</sub> = diffusing capacity of the lung for carbon monoxide; GLI = Global Lung Function Initiative; K<sub>CO</sub> = carbon monoxide transfer coefficient; LLN = lower limit of normal; P<sub>B</sub> = barometric pressure; pred adj Hb = hemoglobin adjusted for predicted value; sb = single breath; TLC = total lung capacity; V<sub>I</sub>/V<sub>C</sub> = inspired volume/vital capacity.

is “virtual” and derived from rapid-acting analyzers and can be adjusted for the size of the patient after the maneuver. However, the display must show the sample volume that was used in the calculation (4). Again, there is a linear graph where the result is plotted as z-scores away from the predicted value.

### Comments and Interpretation

There is a place for technician comments on the test session, any quality issues, and other relevant information that may aid in interpretation. The accompanying figures show how the format units may be combined into one-page reports for both spirometry (Figure 1) and more complete test sessions (Figures E1 and E2). The one-page form has space for only a brief interpretive summary, and therefore laboratories preferring a more detailed interpretation and/or less crowded components may opt for a two-page report. Test tracings are shown for spirometry and other tests as well, but some laboratories may choose not to include all of these for the end user as long as the interpreter has full access to them. Although comparison with prior values is important to interpretation, a format for this is not addressed in this document.

## Selecting and Reporting Reference Values

### General Considerations

Interpretation of PFTs requires comparison with reference values because lung function depends on body dimensions and physiological changes throughout growth and aging. Reference equations use such factors as height, age, sex, and race/ethnicity to predict the average lung function as well as the range of expected values, with the goal of distinguishing the effects of disease from normal variability among healthy individuals.

PFT laboratories must select appropriate reference values for the patients being tested. These should be generated from high-quality data collected from a large sample of healthy asymptomatic individuals who have never smoked or been affected by other respiratory illness or significant exposures. Not all published reference sources meet these criteria; therefore careful consideration of the advantages and disadvantages of available reference values is necessary.

Lung function reports should identify the source of reference values, because the same measured values may be interpreted differently based on the reference source used. Furthermore, manufacturers need to be transparent when reference values have been combined from various sources. In the event that a laboratory changes its selected reference equations, this should be noted on the report, and percent predicted values for prior lung function data should be recalculated, if possible. It is preferred that there be no discontinuity between pediatric and adult equations in the reference values selected, and extrapolation of values beyond the age range of the equations should not be done during growth and will increase uncertainty in the elderly (9, 15). Any such extrapolation must be noted in technician comments.

### Current Spirometry Reference Values

In 2005, the 1999 NHANES III spirometry reference equations (16) were recommended for use in North America (1). This study provided values for whites, African Americans, and Mexican Americans living in the United States. The age span was 8–80 years in two sets of equations with a break at age 18–20 years; a separate recommendation was made for children under age 8. As there was uncertainty whether these equations were a good fit for various European populations, no recommendation was made.

Subsequently, the GLI group was formed with ERS sponsorship and with the participation of the ATS PFT Committee. The goal was to merge available data sets, including the NHANES III data, to develop more broadly applicable reference equations. This effort resulted in new spirometry reference equations using data collected from more than 74,000 individuals, ages 3–95 years, from 26 countries (12). The GLI established reference values for whites, African Americans, North East Asians, and South East Asians. The equations for the white population were shown to be applicable not only in the United States and Europe but in other parts of the world, including Hispanic regions, and for Hispanic Americans. These findings confirm a reanalysis of the NHANES III data, which showed no need for separate reference equations for Hispanic and non-Hispanic whites (17). For individuals not represented by these four groups, or who are of mixed ethnic

origin, a composite equation is provided. (See Appendix EB for guidance on the choice of equations.) The GLI data found the FEV<sub>1</sub>/FVC ratio to be generally independent of ethnic group, and thus its LLN is a useful indication of airflow limitation even when ethnicity is uncertain.

More recently, the Canadian Health Measures Survey has published population-based spirometry reference equations for whites, using a format similar to NHANES III (18). Average adjustment factors are given for several indigenous and immigrant groups. Of note, they found that individuals of Chinese ancestry living in Canada had values intermediate between white values and those predicted by GLI equations from data collected in China.

Since the GLI-2012 publication, these white reference values have been compared with those of NHANES III. In large clinical populations from Australia and Poland, the values predicted by GLI-2012 and NHANES III were similar and rates of airflow limitation (FEV<sub>1</sub>/FVC < LLN) were similar in both men (GLI, 34.5 and NHANES III, 33.3%) and women (GLI, 27.9 and NHANES III, 25.4%) (19). Similar findings have been demonstrated in additional clinical populations from Australia (20), the United States (21), and in children and adolescents (22). The Canadian predicted values compared somewhat more closely with GLI-2012 than NHANES III values, but differences among the three were minor and not likely to be clinically important (18).

In a simulation of NHANES III and GLI-2012 predicted values across a broad range of age and height, the FEV<sub>1</sub> prediction differences were within the recommended repeatability criterion of  $\pm 150$  ml across a wide range of heights and ages (21). There were larger differences at the extremes of height in older individuals, where more uncertainty would be expected due to relatively few subjects of advanced age in the GLI-2012 data and extrapolation of the NHANES III data beyond the age of its subjects.

### Using Reference Data in Interpretation of Results

Both the ATS and ERS recommend the use of the LLN, or the upper limit where appropriate (e.g., lung volumes), to delineate between health and suspected disease. These are set at the fifth percentile (equivalent to a z-score of  $-1.645$ ) so that 95% of a healthy population falls within the normal range

and the lowest 5% would be false positives. However, clinical PFTs are typically done when disease is suspected, increasing the pretest probability of an abnormal result so that the false positive rate is much lower in this setting. The LLN does not necessarily need to be the fifth percentile but, with adequate outcome data, could be adjusted higher when the pretest probability is high or lower for population screening (23, 24). More important than the applicability of a particular LLN is recognition of the uncertainty that lies near any dichotomous boundary and where caution is required, especially when results are limited to a single test occasion.

The respiratory community is familiar with using the percent predicted value to describe lung function results; however, this value should not be used to define abnormality. The true LLN is age- and/or height-dependent and therefore will occur at varying percent values in different individuals. The fixed values commonly used (e.g., 80% predicted for FVC, 0.70 for FEV<sub>1</sub>/FVC) are estimates based on middle-aged adults, and therefore erroneous clinical decisions based on these fixed cutoffs are more likely to occur in children and in older or shorter adults. Using fixed cutoffs also introduces a sex bias into clinical assessments (25). Interpretation of individual results relative to the range of values expected can be more appropriately incorporated into PFT reports using the recommended linear analog scale.

#### Reference Source Recommendations

For spirometry, the GLI-2012 reference values are recommended for use in North America for the ethnic groups represented, as well as in Europe, Australia–New Zealand (26), and other areas with represented populations. For laboratories wishing to maintain continuity, the NHANES III equations also remain recommended for whites (including Hispanics) and African Americans. Use of GLI-2012 is recommended for clinical research studies to facilitate comparisons with international studies (27, 28). The GLI-2012 equations are also preferred for laboratories testing children or adolescents because they permit tracking during this time of rapid lung growth and development without discontinuities due to switching reference sources. The Canadian reference values also provide a useful resource (18). Whatever reference source is used, interpretations

must be based on a parameter-specific lower limit determined from the distribution of the reference data.

For DL<sub>CO</sub>, no prior ATS recommendation has been made because of the wide divergence of available reference values. With ATS and ERS sponsorship a GLI group has assembled data from more than 12,000 individuals in 14 countries to develop new (white-only) reference equations from age 5 to 85 years. Publication of these is expected in 2017 (29) and their rapid adoption is recommended.

For plethysmographic or dilutional lung volumes, no recommendation can be made for reference values at this time. A new international project to address this need is underway. (Values from a Canadian study [30] are used as examples in the accompanying figures.)

## Grading the Quality of Pulmonary Function Tests

### Spirometry

Whereas considerable attention has been given to guidelines for the procedures to conduct spirometry (2), guidelines to assess the quality of the testing are still needed. The purposes of quality review are to provide feedback to the technicians and, in the clinical setting, to indicate any limitations to the interpretation of the results. In clinical research, quality review helps determine whether a subject can be included in a trial and whether data at any time point can be used in the analysis.

Various quality-grading systems have been reported (31–38) and others have been provided by spirometer manufacturers, but users would benefit from standardized methodology. The system recommended for adults and children is shown in Tables 1 and 2. It is modified from a system that has been used in research and epidemiological studies (16, 18, 38). For younger children, 2–6 years of age, the criteria are modified on the basis of the 2007 ATS/ERS recommendations for spirometry testing in preschool children (39). These systems can be used manually, or as part of spirometry software, assigning a grade (A through F) separately for the quality of FVC and FEV<sub>1</sub>. In general, tests with a grade of A, B, or C are usable; tests with grade D are suspect; tests with grade E might be used by the interpreter only to show values “within the normal range”

or “at least as high as,” without demonstrated repeatability; and tests with grade F should not be used.

The grading system consists of acceptability and repeatability components. An ideal test session conforms to prior ATS/ERS recommendations (2, 39) with at least three acceptable maneuvers and repeatable FVC and FEV<sub>1</sub> values. These criteria were intended to guide technicians to achieve the best possible results, and the goal should be to exceed them because many technicians can achieve better quality tests. However, their strict application may also lead to the exclusion of useful results. A grading system allows the user to evaluate the likelihood that spirometry results are representative of true values in the face of test performance that is not ideal. Failure to achieve optimal tests may be due to underlying disease; thus bias may be introduced into clinical research studies by eliminating subjects with more severe disease. The reviewer must consider whether or not the subject's effort in a maneuver was maximal, and/or whether lack of repeatability could be due to lung disease, using all available information including technician comments. When a test session with a poor-quality grade is used as a baseline for pre/post-bronchodilator or longitudinal comparisons, an apparent improvement in values may be the result of better effort or technique.

While strict application of the grading criteria can be done by computer software, the reviewer's role is to apply judgment by reviewing the individual curves, which may change the scoring and allow interpretation. For example, if a maneuver in the session is unacceptable only because of excessive back extrapolation volume, it can still be used to confirm repeatability of FVC. Determining whether the subject has exhaled completely is difficult, and one report has suggested that many subjects are unnecessarily excluded by the 2005 ATS/ERS end-of-test criteria (38). The FET is used to determine whether the subject has tried to exhale long enough, and the end of the volume–time curve is assessed to determine whether expiratory flow has ceased, defined as a volume change less than 0.025 L in 1 second. Often the FET is less than 6 seconds, because the software has stopped data collection once this low flow criterion is met, or an artifact during exhalation may be falsely perceived as a

**Table 1.** Quality Categories for FVC or FEV<sub>1</sub> in Adults and Children

Grade	Criteria for Adults and Older Children and for Children Aged 2–6 Years
A	≥3 acceptable tests with repeatability within 0.150 L for age 2–6, 0.100 L, or 10% of highest value, whichever is greater
B	≥2 acceptable tests with repeatability within 0.150 L for age 2–6, 0.100 L, or 10% of highest value, whichever is greater
C	≥2 acceptable tests with repeatability within 0.200 L for age 2–6, 0.150 L, or 10% of highest value, whichever is greater
D	≥2 acceptable tests with repeatability within 0.250 L for age 2–6, 0.200 L, or 10% of highest value, whichever is greater
E	One acceptable test
F	No acceptable tests

FVC or FEV<sub>1</sub> are each graded separately. The quality categories for FVC and FEV<sub>1</sub> are the same, but the definition of an acceptable curve differs in that FEV<sub>1</sub> acceptability does not consider anything after the first second, whereas FVC does (see Table 2). The adult quality criteria extend to children age 7 or greater. Information on adults and children age 7 or greater is based on Reference 2. Information on children aged 2 to 6 is based on Reference 39.

plateau, thus stopping data collection and underestimating FET and FVC. Subjects should be verbally encouraged to continue the expiratory effort at the end of the maneuver to obtain optimal results (2). The technician and/or reviewer must make a determination as to whether or not end-of-test criteria were met and whether the data are useful. Some subjects, especially children and adolescents, cannot exhale for the required 3 or 6 seconds. If these subjects have a 1-second plateau, and the reviewer judges that these maneuvers represent a maximum FVC, the grade should be adjusted higher. Subjects with airflow obstruction may never reach a plateau even at the suggested 15-second limit and may have nonrepeatable FVC values only because of a difference in FET. Similarly, subjects with restrictive lung disease

may reach an early plateau and may not be able to maintain a 6-second effort. The results may still be considered acceptable in such cases and an appropriate comment by the reviewer should be made.

FEV<sub>1</sub> is graded separately because even test efforts that are clearly unacceptable for FVC may contain a valid measurement of FEV<sub>1</sub> (or FEV<sub>0.75</sub>, the forced expiratory volume exhaled in the first 0.75 s of the FVC maneuver, and a recommended measure in preschool children) (2, 39). Anything that occurs later in the flow–time tracing (e.g., early termination, cough artifacts) does not affect the FEV<sub>1</sub> or FEV<sub>0.75</sub> value, and thus these may be used even with end-of-test errors.

Despite the attention paid to expiratory parameters, the most common reason for

low FVC, FEV<sub>1</sub>, and PEF values is an incomplete inhalation. Achievement of maximal inhalation is best assessed by measures of repeatability and also by the consistency of the shape of the flow–volume or volume–time curve (40).

FVC maneuvers that have lower PEF values compared with others in a session may produce higher FEV<sub>1</sub> values due to the negative effort dependence of flow. The technician should coach forceful initiation of the FVC maneuver (i.e., blast), to achieve the fastest/highest PEF (41). Rounded peaks on the flow–volume curve may reflect submaximal blasts that can increase variability in both FVC and FEV<sub>1</sub>.

**Lung Volumes**

Quality review of lung volume measurement is more challenging as a variety of methods are available, including body plethysmography, nitrogen washout, helium dilution, and radiographic imaging. We are not aware of any quality-grading systems that have been validated for the measurement of absolute lung volumes. Until a validated system is available, we recommend adherence to the 2005 ATS/ERS recommendations for the measurement of lung volumes (3). If the acceptability and/or repeatability criteria are not met but these data are reported, a comment should be included to caution users of the test results.

**Diffusing Capacity (Transfer Factor)**

The DL<sub>CO</sub> (or TL<sub>CO</sub>) test is complex, involving a number of technical factors, and variability can be high. In the absence of quality-grading systems that have been

**Table 2.** Spirometry Test Acceptability Requirements for Adults and for Children

1. A good start of exhalation with extrapolated volume < 5% of FVC or 0.150 L, whichever is greater (For age 2–6, extrapolated volume < 12.5% of FVC or 0.080L)
2. Free from artifacts
3. No cough during first second of exhalation (for FEV<sub>1</sub>)
4. No glottis closure or abrupt termination (for FVC)
5. No early termination or cutoff (for FVC). Timed expiratory volumes can be reported in maneuvers with early termination, but FVC should be reported only with qualification. (For age 2–6, if cessation of effort occurs at greater than 10% of peak flow, then the maneuver should be classified as showing premature termination; although timed expiratory volumes can be reported in maneuvers with early termination, FVC should not)
6. Maximal effort provided throughout the maneuver
7. No obstructed mouthpiece

FVC and FEV<sub>1</sub> are each considered separately for acceptability. FEV<sub>1</sub> acceptability does not consider anything after the first second, whereas FVC does. The adult acceptability criteria extend to children age 7 or greater. Information on adults and children age 7 or greater is based on Reference 2. Information on children aged 2 to 6 is based on Reference 39.



**Table 3.** Quality Grading for DL<sub>CO</sub> Maneuvers

Grade	V <sub>i</sub> /V <sub>C</sub>	Breathhold Time	Sample Collection
A	≥90%*	8–12 s	≤4 s
B	≥85%	8–12 s	≤4 s
C	≥80%	8–12 s	≤5 s
D	≥80%	<8 or >12 s	≤5 s
F	Any test not meeting Grade A, B, C, or D.		

*Definition of abbreviations:* DL<sub>CO</sub> = diffusing capacity of the lung for carbon monoxide; VC = largest vital capacity in the same test session; V<sub>i</sub> = inspired volume.

\*Or V<sub>i</sub>/V<sub>C</sub> ≥ 85% and V<sub>A</sub> within 0.200 L or 5% of V<sub>A</sub> from another acceptable maneuver.

validated using the new recommendations (6), a grading scheme based on maneuver acceptability is proposed (Table 3). The average DL<sub>CO</sub> value from at least two grade A DL<sub>CO</sub> maneuvers that are repeatable within 2 ml/min/mm Hg, or 0.67 mmol/min/kPa, should be reported. If only one grade A maneuver is acquired, the DL<sub>CO</sub> value from that maneuver is reported. If only maneuvers with grades B to D are available, DL<sub>CO</sub> values from these might still have clinical utility; therefore the average of the two best graded of these maneuvers should be reported, but this must be noted to caution the interpreter of the test results. Maneuvers with grade F are not usable.

**The Quality Reviewer**

In the clinical pulmonary function laboratory, technicians, supervisors/managers, or computer software can assign quality grades, whereas in the clinical research setting, an independent quality reviewer is commonly used. This reviewer should be an expert in pulmonary function testing and have extensive experience, both in direct testing and in monitoring testing performed by others. If more than one reviewer is used, comparisons across reviewers should be done to ensure consistency, for example, with a blinded sample of good and bad test sessions.

**Conclusions**

The ATS PFT Committee believes that wide adoption of the formats presented above and their underlying principles by equipment manufacturers and pulmonary function laboratories can improve the interpretation, communication, and understanding of test results. Limiting the number of parameters reported and showing the LLN next to the measured value should improve interpretive accuracy, particularly for those less experienced. Consistency in the order of important data and reserving the word *percent* to percent predicted value should reduce errors. Showing the measured values relative to the normal distribution in a simple linear graphic (with or without reporting a z-score) can enhance understanding of the result. Newer reference values for spirometry expand the applicable age range and ethnicities and eliminate troublesome jumps between equations during growth. New international (albeit only white) reference values for diffusing capacity will help to resolve a long-standing problem. Quality review of PFTs needs to move beyond “did, or did not, meet ATS standards,” but a grading system is most helpful if the grades have a common meaning; therefore adoption of a uniform system is desirable. ■

This official technical statement was prepared by an *ad hoc* subcommittee of the ATS Committee on Proficiency Standards for Pulmonary Function Laboratories.

**Members of the subcommittee are as follows:**

- BRUCE H. CULVER, M.D. (*Chair*)
- ALLAN L. COATES, M.D. (*Co-Chair*)
- CRISTINE E. BERRY, M.D., M.H.S.
- PATRICIA K. CLARKE, R.R.T., R.P.F.T.
- BRIAN L. GRAHAM, Ph.D.
- TEAL S. HALLSTRAND, M.D., M.P.H.
- JOHN L. HANKINSON, Ph.D.
- DAVID A. KAMINSKY, M.D.
- NEIL R. MACINTYRE, M.D.
- MEREDITH C. McCORMACK, M.H.S., M.D.

- MARGARET ROSENFELD, M.D., M.P.H.
- SANJA STANOJEVIC, Ph.D.
- JACK WANGER, M.Sc., R.R.T., R.P.F.T.
- DANIEL J. WEINER, M.D.

**Author Disclosures:** D.A.K. served as a speaker for MGC Diagnostics. N.M. served as a consultant for Alana, Breathe Technologies, InspiRx Pharmaceuticals, and Ventec; and as a speaker for Medtronic.

M.M. received research support from Boehringer Ingelheim Pharmaceuticals; and received royalties for authorship from UpToDate. J.W. served as a speaker and consultant for MGC Diagnostics; as a consultant for Vitalograph; and on an advisory committee for Methapharm. B.H.C., C.E.B., P.K.C., A.L.C., B.L.G., T.S.H., J.H., M.R., S.S., and D.J.W. reported no relationships with relevant commercial interests.

**References**

1. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26: 948–968.
2. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, *et al*.; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 2005;26: 319–338.
3. Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, Casaburi R, Crapo R, Enright P, van der Grinten CP, *et al*. Standardisation of the measurement of lung volumes. *Eur Respir J* 2005;26:511–522.
4. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, *et al*. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 2005;26: 720–735.
5. Culver BH. How should the lower limit of the normal range be defined? *Respir Care* 2012;57:136–145, discussion 143–145.
6. Graham BL, Brusasco V, Burgos F, Cooper BG, Jensen R, Kendrick A, MacIntyre NR, Thompson BR, Wanger J. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur Respir J* 2017; 49:1600016.
7. Hildon Z, Allwood D, Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *Int J Qual Health Care* 2012;24:55–64.

8. Vedal S, Crapo RO. False positive rates of multiple pulmonary function tests in healthy subjects. *Bull Eur Physiopathol Respir* 1983;19:263–266.
9. Quanjer PH, Hall GL, Stanojevic S, Cole TJ, Stocks J; Global Lungs Initiative. Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J* 2012;40:190–197.
10. Coates AL, Graham BL, McFadden RG, McParland C, Moosa D, Provencher S, Road J; Canadian Thoracic Society. Spirometry in primary care. *Can Respir J* 2013;20:13–21.
11. Levy ML, Quanjer PH, Booker R, Cooper BG, Holmes S, Small IR. Diagnostic spirometry in primary care: proposed standards for general practice compliant with American Thoracic Society and European Respiratory Society recommendations. *Prim Care Respir J* 2009;18:130–147.
12. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM, Zheng J, et al.; ERS Global Lung Function Initiative. 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.
13. Quanjer PH, Weiner DJ, Pretto JJ, Brazzale DJ, Boros PW. Measurement of FEF<sub>25–75%</sub> and FEF<sub>75%</sub> does not contribute to clinical decision making. *Eur Respir J* 2014;43:1051–1058.
14. Lukic KZ, Coates AL. Does the FEF<sub>25–75</sub> or the FEF<sub>75</sub> have any value in assessing lung disease in children with cystic fibrosis or asthma? *Pediatr Pulmonol* 2015;50:863–868.
15. Subbarao P, Lebecque P, Corey M, Coates AL. Comparison of spirometric reference values. *Pediatr Pulmonol* 2004;37:515–522.
16. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
17. Kiefer EM, Hankinson JL, Barr RG. Similar relation of age and height to lung function among whites, African Americans, and Hispanics. *Am J Epidemiol* 2011;173:376–387.
18. Coates AL, Wong SL, Tremblay C, Hankinson JL. Reference equations for spirometry in the Canadian population. *Ann Am Thorac Soc* 2016;13:833–841.
19. Quanjer PH, Brazzale DJ, Boros PW, Pretto JJ. Implications of adopting the Global Lungs Initiative 2012 all-age reference equations for spirometry. *Eur Respir J* 2013;42:1046–1054.
20. Brazzale DJ, Hall GL, Pretto JJ. Effects of adopting the new global lung function initiative 2012 reference equations on the interpretation of spirometry. *Respiration* 2013;86:183–189.
21. Linares-Perdomo O, Hegewald M, Collingridge DS, Blagev D, Jensen RL, Hankinson J, Morris AH. Comparison of NHANES III and ERS/GLI 12 for airway obstruction classification and severity. *Eur Respir J* 2016;48:133–141.
22. Quanjer PH, Weiner DJ. Interpretative consequences of adopting the Global Lungs 2012 reference equations for spirometry for children and adolescents. *Pediatr Pulmonol* 2014;49:118–125.
23. Miller MR. Spirometry in primary care. *Prim Care Respir J* 2009;18:239–240.
24. Culver B. Defining airflow limitation and chronic obstructive pulmonary disease: the role of outcome studies. *Eur Respir J* 2015;46:8–10.
25. 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;139:52–59.
26. Brazzale D, Hall G, Swanney MP. Reference values for spirometry and their use in test interpretation: a position statement from the Australian and New Zealand Society of Respiratory Science. *Respirology* 2016;21:1201–1209.
27. Swanney MP, Miller MR. Adopting universal lung function reference equations. *Eur Respir J* 2013;42:901–903.
28. Miller MR. Choosing and using lung function prediction equations. *Eur Respir J* 2016;48:1535–1537.
29. Stanojevic S, Graham BL, Cooper BG, Thompson BR, Carter KW, Francis RW, Hall GL; Global Lung Function Initiative TLCO Working Group; Global Lung Function Initiative (GLI) TLCO. Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur Respir J* 2017;50:1700010.
30. Gutierrez C, Ghezzi RH, Abboud RT, Cosio MG, Dill JR, Martin RR, McCarthy DS, Morse JLC, Zamel N. Reference values of pulmonary function tests for Canadian Caucasians. *Can Respir J* 2004;11:414–424.
31. Enright PL, Johnson LR, Connett JE, Voelker H, Buist AS. Spirometry in the lung health study. 1. Methods and quality control. *Am Rev Respir Dis* 1991;143:1215–1223.
32. Banks DE, Wang ML, McCabe L, Billie M, Hankinson J. Improvement in lung function measurements using a flow spirometer that emphasizes computer assessment of test quality. *J Occup Environ Med* 1996;38:279–283.
33. Stoller JK, Buist AS, Burrows B, Crystal RG, Fallat RJ, McCarthy K, Schluchter MD, Soskel NT, Zhang R;  $\alpha$ -1 Antitrypsin Deficiency Registry Study Group. Quality control of spirometry testing in the registry for patients with severe  $\alpha$ -1-antitrypsin deficiency. *Chest* 1997;111:899–909.
34. Pellegrino R, Decramer M, van Schayck CPO, Dekhuijzen PN, Troosters T, van Herwaarden C, Olivieri D, Del Donno M, De Backer W, Lankhorst I, et al. Quality control of spirometry: a lesson from the BRONCUS trial. *Eur Respir J* 2005;26:1104–1109.
35. Goss CH, McKone EF, Mathews D, Kerr D, Wanger JS, Millard SP; Cystic Fibrosis Therapeutics Development Network. Experience using centralized spirometry in the phase 2 randomized, placebo-controlled, double-blind trial of denufosal in patients with mild to moderate cystic fibrosis. *J Cyst Fibros* 2008;7:147–153.
36. Pérez-Padilla R, Vázquez-García JC, Márquez MN, Menezes AMB; PLATINO Group. Spirometry quality-control strategies in a multinational study of the prevalence of chronic obstructive pulmonary disease. *Respir Care* 2008;53:1019–1026.
37. Enright P, Vollmer WM, Lamprecht B, Jensen R, Jithoo A, Tan W, Studnicka M, Burney P, Gillespie S, Buist AS. Quality of spirometry tests performed by 9893 adults in 14 countries: the BOLD Study. *Respir Med* 2011;105:1507–1515.
38. Hankinson JL, Eschenbacher B, Townsend M, Stocks J, Quanjer PH. Use of forced vital capacity and forced expiratory volume in 1 second quality criteria for determining a valid test. *Eur Respir J* 2015;45:1283–1292.
39. Beydon N, Davis SD, Lombardi E, Allen JL, Arets HG, Aurora P, Bisgaard H, Davis GM, Ducharme FM, Eigen H, et al.; American Thoracic Society/European Respiratory Society Working Group on Infant and Young Children Pulmonary Function Testing. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *Am J Respir Crit Care Med* 2007;175:1304–1345.
40. Haynes JM, Kaminsky DA. The American Thoracic Society/European Respiratory Society acceptability criteria for spirometry: asking too much or not enough? *Respir Care* 2015;60:e113–e114.
41. Enright PL. How to make sure your spirometry tests are of good quality. *Respir Care* 2003;48:773–776.