Reference values for pulmonary function test: suggestions for a correct use and interpretation

Riccardo Pistelli[1]
Sandra Sammarro[2]

[1] Respiratory Physiology Unit, Columbus Hospital, Catholic University, Rome, Italy
[2] Department of Respiratory Diseases, San Camillo-Forlanini Hospital, Rome, Italy

Address for correspondence:
Riccardo Pistelli, MD, PhD
Respiratory Physiology Unit
Columbus Hospital, Catholic University
Via Moscati 31
00168 Roma, Italy
Phone: +39 06 3503678
E-mail: riccardopistelli@h-columbus.it

Summary

This paper describes the complex reasoning activity the respiratory clinician performs while using the respiratory function tests for diagnosing a disease. The probability each functional parameter is belonging to a healthy subject is the first useful measure to start with that reasoning. In a second step, a set of parameters is evaluated to define a functional syndrome and, finally, the identification of the disease underlying that syndrome is the last step of this journey. The comparison of measured parameters with reference data in healthy subjects is crucial to measure the probability at the beginning of the diagnostic reasoning. Reference data are often used also to define the severity of a respiratory disease or to evaluate the efficacy of a therapeutic intervention. Clearly, reference data are used to address quite different tasks, the methodological implications of which are discussed in this paper. In the last part of this paper the authors present their suggestions and some temporary solutions to what is a lack of knowledge for a more sensible utilization of respiratory reference values.

KEY WORDS: Pulmonary Function Test (PFT), spirometry, reference values, chronic obstructive pulmonary disease, lower limit of normality (LLN).

Introduction

The interpretation of Pulmonary Function Test (PFT) is usually focused on the comparison of data describing some aspects of pulmonary function obtained in each individual with data obtained from a population of healthy subjects. In other words, the pulmonary physiologist aims at defining the position of each individual measure obtained in a PFT session with reference to the distribution of the same measure in the reference population. In general, the distance and the direction of the deviation of each measure from the mean or median value in the reference distribution are used to derive a clinically meaningful interpretation of a PFT session. The rationale for this comparison derives from the assumption that the data actually measured can be attributed to a member of the population from which the standard reference values have been derived.

In general, the distance and the direction of the deviation of each measure from the mean or median value in the reference distribution are used to derive a clinically meaningful interpretation of a PFT session.

In this paper we discuss the factors that can challenge that assumption, the lack of reliable reference values for many functional parameters and how it is possible to sort out from. Secondly, we discuss some misuse or biased interpretation of the deviation from the reference values and what we hope it could be a better future use of the physiological evaluation.

The Standard Reference Values

Predicted values should be obtained from studies of "normal" or "healthy" subjects with the same anthropometric (gender, age and height) and ethnic characteristics of the patient being tested. In the past, predicted values were summarized in the cells of tables where it was possible to find the mean value of a functional parameter measured in a small sample of healthy people hopefully comparable with the subject on study. In the original publication of John Hutchinson, the inventor of the spirometer, it is possible to find the first example of tables of

As a consequence of this impressive change of human phenotype, the use of reference standards produced with data collected many decades ago may systematically bias our judgment towards an underestimation of possible anomalies.

© Copia non istituzionale

reference values (1). At that time, it was possible only to roughly evaluate the difference, either as absolute or percentage value, between a measured Vital Capacity and its standard reference included in the appropriate cell of the table. More recently, tables including in each cell a standardized measure of variability of the reference value have been published. Since then on, it was possible to estimate the position of any actual measure with reference to the distribution of the same measure in the population of comparable healthy subjects that is the probability of any measure to belong to a normal subject (2). In the last decades, the possibility of analyzing large amount of data with computers running sophisticated statistical programs made it possible to model the reference values according to multivariate models. By using the equations produced in this way, it is now possible to tailor a reference value according to the specific anthropometric characteristics of each subject and to evaluate the difference between the measured and predicted values according to the modeled variability of any functional variable.

Ideally, reference values are calculated with equations derived from measurements performed in a representative sample of healthy subjects in the general population. In the reality, reference equations are quite often derived from large groups of volunteers, provided that criteria for normal selection and proper distribution of anthropometric characteristics are satisfied. Criteria to define subjects as ''normal'' or healthy have been reported in previous ATS and European Respiratory Society (ERS) statements (3-5). A complete list of all papers reporting equations to calculate the reference values for data collected during a PFT session is beyond the limits of this paper. Examples of this list can be found in published papers for both the general population and some specific subsets such as elderly people (6, 7). From those lists it is quite easy to appreciate that, in most of cases, reference values have been produced for only few functional parameters, in small convenient but not representative samples of subjects, in many countries and different populations, with different methods and equipment. An updating of those lists does not appreciably change this general picture. It is, then, much more important to discuss which criteria should guide the choice of the proper reference equations to be used in a clinical setting.

In the first step of the clinical approach to the respiratory function data we need reference values to define the "normality" of measures obtained from a subject referred to a respiratory physiology unit. In a second step, the clinician put a small amount of all measured data in a rather formal (6) or informal diagnostic algorithm to obtain a global interpretation of the functional situation of a specific subject. In the last step, he tries to attribute that global interpretation to a peculiar disease or disorder, if any, affecting subject. Fortunately, in spite of the lack of reliable reference values for many functional parameters, we have many reference values for the small amount of data commonly used in the clinical interpretation reasoning. Those data are: Vital Capacity (VC), Forced Expiratory Volume in one second (FEV1), ratio FEV1/FVC, Total Lung Volume (TLC), and Lung Diffusion capacity for CO (DLCO). The following are the main factors we need to consider when selecting the proper reference values: Characteristics of the population from which the reference data were sampled, Range of age of subjects included in the sample, Calendar date of collection of data, Methods and Equipment used to perform the measurements. The software of instruments for PFT should allow for an easy installation of new equations to generate reference values, if necessary, and the reference values used should be documented on every pulmonary function report with the first author's last name (or organization) and the date of publication (5).

The comparability of methods and equipment used to produce reference data with methods and equipment used in a specific laboratory as well as the comparability of ethnic and exposure characteristics of the population to which the clinician hopes to meaningfully refer the data measured in each subject are quite obvious factors to guide the choice of reference equations. However, the calendar date of collection of data from which the reference standards were derived is a less obvious point that needs some discussion. The volume of lungs in humans living in Europe increased well above the contemporary increase of body size in the last two centuries. Using the VC measurements of healthy individuals performed by John Hutchinson in the middle of the 19th century and the most popular reference values used in Europe in the last twenty years (8), we can estimate a 1300 ml increase of VC for a 170 cm long 35-year-old male. As a consequence of this impressive change of human phenotype (formally "cohort effect on the relation between height and VC"), the use of reference standards produced with data collected many decades ago may systematically bias our judgment towards an underestimation of possible anomalies. This bias is certainly more relevant for elderly subjects, because the most popular reference equations for this age range were produced with data collected from subjects born around the first decade of the 20th century, that is before the socio economic related accelerated rate of change of the human phenotype in Europe. A less frequently appreciated consequence of this cohort effect is the systematic overestimation of the age related decrease of lung volumes in the reference equations derived from cross-sectional measurements of functional data in healthy subjects of different ages. Actually, the older subjects are also subjects born many decades ago so that their lung volume is affected by both aging and cohort effect. A further relevant point in the selection of reference equations is the age range of subjects from which those equations were derived. Statistical reasoning can produce an estimate of the probability of any functional measurement to belong to the population of measurements used as reference. If the subject actually studied is older than subjects in the reference population, that statistical reasoning cannot be formally applied. To overcome this problem, it can be assumed that the extrapolation of estimates of a predictive equation outside the limits of the sampled population does not in-

© Copyright Pagine Italiane

troduce an appreciable bias. However, if the cohort effect is a factor that biases the estimates for subjects included in the age range of the reference population, the linear extrapolation outside those limits certainly increase that bias. Moreover, the assumption of a linear or otherwise mathematically modeled trend outside the limits of the measured observations is frequently contradicted in biological sciences and must be considered always arbitrary and to be cautiously used when any alternative is not available. In any case, it should be clearly stated in the report of a PFT session that such an extrapolation was used and that any inference about the normality of the measurements is not based on a formally accurate statistical reasoning.

Current use and interpretation of standard reference values

A calculated reference value is usually the mean or the median value in a distribution of values. This distribution is interpreted as the residual biological variability of a functional parameter in the healthy normal population after the removal of variability due to the factors (i.e. age, height, and gender) included in the modeling process. Some statistical parameter produced by the model can be used to calculate any reference value in the same distribution. Usually, with the assumption of a normal distribution of data, the Residual Standard Deviation (RSD) is the parameter used to calculate any reference value in that distribution. It is possible than to multiply the RSD for any value (z) corresponding to a defined position in the normal distribution to calculate a quantity that can be subtracted from the mean value if we aim to calculate the value of the reference value corresponding to a specific position between the mean and - ∞. Of course (but this is a rare option in respiratory physiology) we may also add the same quantity to the mean value to calculate the value corresponding to a specific position between the mean and + ∞. If the assumption of normality is not plausible, it is necessary to use alternative non-parametric methods of modeling to give a reliable estimate and a numerical expression of the residual variability. Whichever the chosen statistical solution (the discussion of which is beyond the limits of the present paper), it is possible to calculate any value that can be used as a cut-off point to classify a measurement obtained in the clinical practice as "normal" or "abnormal". Usually, this value corresponds to the 5th centile in the distribution of standardized reference values and it is named Lower Limit of Normality (LLN). Of course, the value of the LLN is dependent from the data set and from the model used to calculate it. When a measured functional parameter is used to define the presence or absence of a disease according to its position with reference to the LLN, we may obtain different results by using different reference data set or different model applied to the same data set. Chronic Obstructive Pulmonary Disease (COPD) and other chronic pulmonary diseases are among the most impressive examples of this situation. The recent publication of new equations to produce "global" reference values (9) has been followed by many publications in which the impact of those equations on the diagnosis and staging of COPD (10-15), as well as their "global" validity in some areas (16), has been discussed. For example, Quanjer et al. (10) discuss the implications of adopting the new global equations on the diagnosis of a restrictive pattern, generating higher or lower prevalence rates than those obtained while using the ECSC or the NHANES reference equations, respectively. It is interesting that, in the same paper, the prevalence of the obstructive pattern, whose diagnosis is based on the FEV1/FVC ratio and not on the absolute values of the two parameters, was only marginally affected by the choice of the reference equations. However, the definition of gravity of the obstruction, based on the FEV1 value only, was still quite affected by the choice of the reference equations. In conclusion, the use of a value, whose frequency in a normal healthy population has been defined by modeling its variability in a representative sample of the same population, is the correct method to define a limit we may assume a the LLN for a specific functional parameter. However, we must consider that LLN as dependent from the sampled data, whose representativeness may never be "global", and from a modeling strategy. Most importantly, we have to be conscious that the definition of LLN based on a frequency evaluation is completely arbitrary and assuming that what is "rare" is "abnormal". In the last part of this paper there are some suggestions about a strategy to sort out from this tricky situation with a clinically meaningful use of reference equations and what the authors hope could be a more sensible use of respiratory functional data in the future.

When a measured functional parameter is used to define the presence or absence of a disease according to its position with reference to the LLN, we may obtain different results by using different reference data set or different model applied to the same data set.

The definition of LLN based on a frequency evaluation is completely arbitrary and assuming that what is "rare" is "abnormal".

A clinical interpretation of respiratory functional data

In the clinical practice, data collected in many facilities of any hospital are used to confirm or refuse a diagnosis, to check the efficacy of a therapeutic intervention and to define the prognosis of a patient. In respiratory medicine, reference equations for functional data are used in different ways to get all those goals: LLN to define the diagnosis, the value of a functional parameter as percent of predicted to define gravity and prognosis,

FEV1 is a good predictor of survival in the general population, but it is only one of predictors of prognosis in patients affected by COPD.

© CIC Edizioni Internazionali

and the pre-post treatment difference of a functional parameter normalized to the predicted scale to evaluate the efficacy of that intervention.

If the definition of a disease is the presence of a value of a functional parameter below the LLN, the clinician is in an apparent situation of certainty until the next change of the LLN, for example following the publication of new equations. A situation in which a diagnosis can be accepted or refused according to a changing external standard is clearly not acceptable and may be one of the reasons of the under use of the functional assessment in the clinical practice of respiratory medicine. Up to now, the possible solution is the use of the LLN according to its probabilistic definition: 5th centile of the distribution in the normal reference population. It is possible then to compare that probability of being a member of the normal healthy population with the probability of being affected by a disease as suggested by all the other already collected clinical data. If the probability of the presence of a disease is quite high, it could be sensible to push higher the LLN to accept that diagnosis. In the opposite case, it could be sensible to push the LLN in the opposite direction. Fortunately, as already said before, the ratios of functional parameters, among which FEV1/FVC is the most widely used, are less dependent from changing reference equations. If the unreasonable use of an age independent fixed ratio is avoided, the above reported comparison of the clinical suspicion of COPD *versus* the centile of the LLN can improve the reliability of that diagnosis.

The definition of severity and prognosis of COPD, based on a functional parameter only, was suggested by the GOLD Initiative (17) and it was successful beyond any reasonable expectation of the authors. Since the very beginning, it was clear that the distinction of different severity stages based on the FEV1 value as percent of predicted was not associated with meaningful differences of the clinical situation of patients (18). Recently, the same classification has been integrated with clinical data, but up to now no change has been adopted for the expression of FEV1 (19). Moreover, FEV1 is a good predictor of survival in the general population, because of its close association with Vital Capacity, but it is only one of predictors of prognosis in patients affected by COPD. Probably, it could be sensible to use FEV1 as a measure of severity and prognosis by adopting some dichotomous classification with reference to the most reliable recent reference equations or to a direct standardization of its absolute value using the anthropometric measures of each subject (i.e. height^3 or BMI). A reduction of the scale of severity to only two reasonable grades is the only sensible approach when the evidence about the meaning of more detailed scales is completely lacking.

If the changes of functional parameters to evaluate the efficacy of interventions are expressed as differences between two values calculated as percent of predicted, it must be realized that equal changes in two subjects may have quite different meanings in terms of recovery towards a healthy normal situation, depending from the value before the treatment. This is due to the shape of the distribution curve, the area of which is much greater as it get closer to the mean or median value. To avoid possible mistakes in the interpretation of results, that difference should be measured between values expressed as centiles of the reference distribution. However, if the efficacy of interventions has to be evaluated against what is commonly known as the Minimal Clinically Important Difference (MCID), it is worth to have some study to define this last parameter. Many techniques can be applied to this aim (i.e. Distribution-based, Anchor-based, Delphi method) but, unfortunately, quite rarely those studies have been performed for the functional respiratory parameters.

Summary of suggestions

Recently published reference data for interpreting the results of a PFT session are now available, but only for the most frequently measured parameters. No matter which methods and data set have been used to produce those references, it should be sensible to check their validity in the practice of any clinical physiology service. A shortcut to do this check is by measuring how many measured data in clinically normal subjects fall outside the 5th-95th range on the centile scale. If significantly more than 5%, it could be probably better to look for a different set of reference equations or to apply some correction factor when comparing measured and reference data.

The LLN should never be assumed as the absolute border between healthy people and those affected by a specific disease without any comparison with the pre-test probability of presence of that disease in any individual.

The position of any measurement with reference to the normal healthy population must be defined in centile units. Many publications suggest identify the 5th centile with the LLN for many functional parameters. However, the LLN should never be assumed as the absolute border between healthy people and those affected by a specific disease without any comparison with the pre-test probability of presence of that disease in any individual. It should be a common wish the predictive value of a respiratory functional parameter for the presence and prognosis of a disease could be formally defined by using the results of longitudinal studies in cohorts of patients (20). In the past, some quite popular numerical risk factors for cardiovascular diseases have been defined in famous cohort studies (21) and more or less regularly updated with revisions of new data (22). Up to now, unfortunately, nothing similar is available for the community of respiratory clinicians, which may rely only on a formal or rather informal use of probabilities (23), the well known “clinical art”, in their daily work with the data of a PFT session.

References

1. Spriggs EA. John Hutchinson, the inventor of the spirometer. Medical History; 1977;21:357-364.

© Copyright Edizioni Internazionali

2. Commission of the European Communities. Reference tables for spirometric examinations. Luxembourg, 1967.
3. American Thoracic Society. Lung Function Testing: Selection of Reference Values and Interpretative Strategies. Am Rev Respir Dis. 1991;144:1202-1218.
4. Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. Eur Respir J. 1995;8:492-506.
5. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF,Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J. 1993;6:Suppl. 16:5-40.
6. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson DC, MacIntyre N, McKay R, Miller MR, Navajas D, Pedersen OF, Wanger J. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.
7. Pistelli R, Andreani M, Baldari F, Sammarro S. Respiratory function standards in the elderly. Eur Respir Mon. 2009;43:18-24.
8. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J. 1993;6:Suppl. 16:5-40.
9. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J, Stocks J; ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95-yr age range: theglobal lung function 2012 equations. Eur Respir J. 2012 Dec;40(6):1324-43.
10. 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.
11. Sluga R, Smeele IJ, Lucas AE, Thoonen BP, Grootens-Stekelenburg JG, HeijdraYF, Schermer TR. Impact of switching to new spirometric reference equations onseverity staging of airflow obstruction in COPD: a cross-sectional observational study in primary care. Prim Care Respir J. 2014 Mar;23(1):85-91.
12. Kim N, Park MH, Kim SY, Suh C, Lee S, Kim KH, Lee CK, Kim DH, Lee JT. Discordance in spirometric interpretations based on korean and non-korean reference equations. Ann Occup Environ Med. 2013 Dec 27;25(1):42.
13. Stanojevic S, Stocks J, Bountziouka V, Aurora P, Kirkby J, Bourke S, Carr SB, Gunn E, Prasad A, Rosenfeld M, Bilton D. The impact of switching to the newglobal lung function initiative equations on spirometry results in the UK CF Registry. J Cyst Fibros. 2014 May;13(3):319-27.
14. Quanjer PH, Weiner DJ. Interpretative consequences of adopting the Global Lungs 2012 reference equations for spirometry for children and adolescents. Pediatr Pulmonol. 2014 Feb;49(2):118-25.
15. 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(3):183-9.
16. Ben Saad H, El Attar MN, Hadj Mabrouk K, Ben Abdelaziz A, Abdelghani A,Bousarssar M, Limam K, Maatoug C, Bouslah H, Charrada A, Rouatbi S. The recent multi-ethnic global lung initiative 2012 (GLI2012) reference values don't reflect contemporary adult's North African spirometry. Respir Med. 2013Dec;107(12):2000-8.
17. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS; GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med. 2001 Apr;163(5):1256-76.
18. Antonelli-Incalzi R, Imperiale C, Bellia V, Catalano F, Scichilone N, Pistelli R, Rengo F; SaRA Investigators. Do GOLD stages of COPD severity really correspond to differences in health status? Eur Respir J. 2003 Sep;22(3):444-9.
19. From the Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2014. Available from: http://www.goldcopd.org/.
20. Marks GB. Are reference equations for spirometry an appropriate criterion for diagnosing disease and predicting prognosis? Thorax. 2012 Jan;67(1):85-7.
21. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart diseases: the Framingham Study. Am J Public Health Nations Health. 1951;41(3):279-81.
22. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines Neil J. Stone, Jennifer Robinson, Alice H. Lichtenstein, C. Noel Bairey Merz, Conrad B. Blum, Robert H. Eckel, Anne C. Goldberg, David Gordon, Daniel Levy, Donald M. Lloyd-Jones, Patrick McBride, J. Sanford Schwartz, Susan T. Shero, Sidney C. Smith, Jr, Karol Watson, and Peter W.F. Wilson Circulation. 2013; published online before print November 12 2013.
23. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians use of quantitative measures of test accuracy. Am J Med. 1998;104(4):374-80.

© Copyright Pneumologia Internazionali