

How Should the Lower Limit of the Normal Range Be Defined?

Bruce H Culver MD

Introduction
The Lower Limits for Spirometry Data
The Lamda, Mu, Sigma Method
Indicators of Severity
Spirometry Reference Data
Accepting Uncertainty
Spirometry Based Outcome Studies
Lung Volume Reference Data
Diffusing Capacity Reference Data
Reporting the Numbers

Lung function parameters vary considerably with age and body size, so that, unlike many laboratory tests, the normal range of expected values must be individualized. For spirometry, only low values are considered to be abnormal, so the lower limit of normal (LLN) is taken to be equal to the 5th percentile of a healthy, non-smoking population. Simple and commonly used “rules of thumb,” such as an $FEV_1/FVC < 0.70$ to indicate air-flow obstruction, or assuming values $< 80\%$ of predicted to be abnormal, are inaccurate and will cause misclassification, specifically under-diagnosis of abnormalities in younger, taller individuals and over-diagnosis in those older or shorter. A much more accurate LLN for the FEV_1/FVC ratio, which recognizes the change with age of this measurement, can be easily determined by subtracting 10 (10% or 0.10) from the age specific FEV_1/FVC predicted for any individual. The analysis and mathematical descriptions of reference data have become increasingly sophisticated in recent years, but the interpretation of values near the LLN continues to carry uncertainty, due to an overlap in values between low normal values and those reflecting early disease. Among patients referred to a pulmonary function laboratory, the pre-test probability of disease may be relatively high, so that even individuals with values above the LLN may be more likely than not to have respiratory disease. A future goal for the pulmonary community would be the development of risk stratified outcome data that would allow an estimation of the probability of disease with progressive decrements in lung function. While interpreting spirometry results near the LLN will continue to be problematic, a more important task for the pulmonary community is to focus on finding the pool of individuals with clear-cut, but undiagnosed, COPD. And for this, good quality spirometry remains the best tool and must be widely available. Key words: lung function testing; spirometry; lung volume measurement; reference equations; lung diseases; obstructive; outcome assessment. [Respir Care 2012;57(1):136–143. © 2012 Daedalus Enterprises]

Dr Culver is affiliated with the Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, Washington.

Dr Culver presented a version of this paper at the 48th RESPIRATORY CARE Journal Conference, “Pulmonary Function Testing,” held March 25–27, 2011, in Tampa, Florida.

The author has disclosed no conflicts of interest.

Correspondence: Bruce H Culver MD, Division of Pulmonary and Critical Care Medicine, University of Washington School of Medicine, UWMC 356522, Seattle WA 98915. E-mail: beculver@uw.edu.

DOI: 10.4187/respcare.01427

Introduction

A primary step in the interpretation of lung function data is to determine if the values obtained fall within the expected normal range of values. It has been generally accepted in both research and clinical medicine to allow a 5% error rate (ie, $P < .05$ for the probability of a chance association and ± 2 standard deviations (SD) to encompass 95% of a value normally distributed in a population). In pulmonary function testing, where we are generally concerned only with low values and where not all parameters are normally distributed, this has been stated as: the lower limit of normal (LLN) is equal to the 5th percentile of a healthy population. This appeared in the 1991 ATS guidelines, after equivalent statements appeared in the early 1980s, then was reconfirmed in the 2005 American Thoracic Society/European Respiratory Society (ATS/ERS) document.¹⁻⁴ For a value with a normal distribution, the 5th percentile LLN is equivalent to the mean $- 1.645 \times SD$. For regressions the equivalent variance is expressed as standard error of the estimate (SEE) or residual standard deviation (RSD). For tests in which either high or low values may be considered abnormal, Mean $\pm 2 \times SEE$ would define a normal 95% CI from the 2.5 to 97.5 percentiles.

Unlike the great majority of clinical laboratory tests, where the normal range is the same for all individuals, or at least all adults, lung function parameters vary considerably with age and body size, so that the expected value must be individualized. This is done by calculating a reference or predicted value that represents the midpoint of the normal range for healthy individuals of the same age, height, sex, and ethnicity. Preferably, but less commonly in our reports, a normal range or lower limit value would also be stated. This added complexity has led to the common use of approximations for the LLN, which lead to errors of classification, especially among individuals younger, older, shorter, or taller than the midrange.

The issues around the use of a fixed ratio of 0.70, often expressed as 70%, to establish air-flow obstruction in COPD have been dealt with elsewhere. That the FEV₁/FVC ratio declines with age is not a new finding. In the classic Bates, Macklem, Christie textbook of 1971, *Respiratory Function in Disease*, it is noted that the value in normal subjects 20–30 years old is 87% and falls to 81% by age 50–60, referenced to a 1958 French paper (appropriately, as the FEV₁/FVC ratio was described there in 1949 and became known as the Tiffeneau index).⁵ A decrement in the ratio with age was also clearly demonstrated in the reference data reported by Morris in 1973 and by Crapo in 1981.^{3,6} The change of 2% per decade in the French data is almost exactly that found in the more recent large National Health and Nutrition Examination Survey (NHANES) III data set.⁷ Despite the pleas of many, including much of the

Table 1. Interval Below the Predicted FEV₁/FVC to the Lower Limit of Normal in the National Health and Nutrition Examination Survey III

	White	African-American	Mexican-American
Male	9.7	10.4	9.1
Female	9.8	10.7	9.3

group assembled here, the Global Initiative for Obstructive Lung Disease (GOLD) organization⁸ has elected to continue to use a ratio of 0.70 in their definition of COPD, but there is no excuse for any pulmonary function laboratory, or even office based practitioner, to use this as a general criterion for air-flow obstruction.

The Lower Limits for Spirometry Data

The NHANES III reference data are recommended by the ATS/ERS 2005 guidelines for establishing spirometry predicted values for North Americans.^{2,7} These include separate equations to define the LLN for each parameter, and, as noted above, this should be included in the report, but, for the FEV₁/FVC ratio, even this small step is unnecessary. Analysis of this data set showed that the variance (SEE) for the ratio did not vary significantly by age or height, so the interval from the predicted midpoint value to the 5th percentile LLN is constant and given by $1.645 \times SEE$. Conveniently, for the FEV₁/FVC this value is very close to 10% (or 0.10). Exact values for men and women in the 3 ethnic groups reported are shown in Table 1. Since all labs, and even most handheld spirometers, display an age specific predicted value for the FEV₁/FVC ratio, all the interpreter need do is subtract 10 to find a statistically accurate LLN. This same value would also be a reasonable approximation to the true LLN for any of the widely used reference sets, far more so than any fixed value of the ratio. One need not over-emphasize the exact value of the LLN, as there will always be some uncertainty around this point, and the interpreter would be wise to consider a range on either side (perhaps ± 1 ; ie, between 9 and 11 units below the predicted ratio) to represent a “borderline value” rather than normal or indicative of air-flow obstruction.

More prevalent in clinical pulmonary function labs than the fixed ratio is the use of a fixed percent of predicted to determine the LLN for most other parameters, but this can lead to similar misclassification. Over the years the prediction equations have become more sophisticated and complex, but still rely primarily on age, height, sex, and ethnicity.

Typical adult equations of the 1970s and 1980s were linear regressions of the form:

Table 2. Fifth Percentile Lower Limit of Normal as a Percent of Predicted

		25, Tall	40, Average	70, Short
NHANES III ⁷				
FVC	Male	83.3	82.3	77.8
	Female	81.9	81.4	75.2
FEV ₁	Male	82.8	81.3	74.3
	Female	81.9	80.6	73.3
FEF ₂₅₋₇₅	Male	65	60	32
	Female	63	58	35
Crapo ³				
FVC	Male	81.1	77.7	70.3
	Female	85.6	83.7	73.1
FEV ₁	Male	82.7	79.4	71.3
	Female	85.8	83.5	71.1

NHANES III = National Health and Nutrition Examination Survey III
 FEF₂₅₋₇₅ = forced expiratory flow during the middle half of the forced vital capacity maneuver

$$a + b \times \text{age} + c \times \text{Ht}$$

The normal data were generally found to have a homoscedastic distribution: that is, the variance around values of small magnitude is the same as that around large values. In this case, the 5th percentile LLN will occur at a lower percent of a small predicted value, and at a higher percent of predicted for a large value. Thus a fixed percent predicted LLN such as the commonly used 80% will under-diagnose abnormalities in younger, taller individuals and over-diagnose in those older or shorter.

The NHANES equations of 1999 use low order polynomials in the form:

$$a + b \times \text{age} + c \times \text{age}^2 + d \times \text{Ht}^2$$

(for FEV₁/FVC, only a + b × age was required).

In order to cover the age range from 8 to 80, it was necessary to provide separate equations above and below age 18 (F) or 20 (M). Unlike some earlier reference data, the large NHANES III data sample did show that the population variance of FVC and FEV₁ increased with the larger values associated with greater height, although it did not vary with age, and thus the equations for LLN take height into account. Even so, the calculated LLN for FVC and FEV₁ ranges from 83% of the predicted value in young adults to 75% at age 70, so the widely used 80% of predicted is correct, or nearly so, only in the midrange of age (Table 2). For the forced expiratory flow during the middle half of the FVC maneuver (FEF₂₅₋₇₅), the true LLN ranges from 65% of predicted in young individuals down to 35% by age 70, and misunderstanding this has contributed to much of its undeserved reputation for “sensitivity.”

The Lamda, Mu, Sigma Method

Recent efforts have used more complex statistical techniques to express lung function in smoothed curves through growth and decline.⁹ Most recently, Stanojevic et al have introduced to spirometry the use of the lambda, mu, sigma (LMS) method, previously used to develop pediatric growth charts with their bands of percentiles.¹⁰ At each point along the continuum, mu represents the median, sigma the coefficient of variance, and lambda is an index of skewness. The result is not an equation per se, but a set of tables, read by computer that creates a smooth continuous predicted value (given by the median, mu) from early childhood to old age. The sigma and lambda terms allow for the 5th percentile LLN to be independently determined throughout the age-height spectrum and further demonstrates the unsuitability of a fixed percent value for LLN. By this analysis, the normal range, as a percent of predicted, of both the ratio and its components is wider in young children and widens progressively above age 40 (Fig. 1).

This LMS method is being utilized in a current ERS Task Force effort to merge many data sets into new prediction models. One example of its flexibility is shown in the recent publication describing an unusual variation in the FEV₁/FVC ratio during the adolescent growth spurt, where the ratio, which falls rapidly during childhood, then transiently increases before resuming the steady decline of adulthood.¹¹ This apparently reflects differences in the rate of development of airway properties and thoracic dimensions. This task force, the Global Lungs Initiative group, has now gathered raw spirometric data from over 150,000 healthy subjects in 70 data sets from 34 countries, in an effort to form a single source for multiethnic, all age reference data. Early results from this large data pool show that there has been no secular trend in spirometry values over a 30 year span, and provide evidence that to reliably reflect a population, the reference sample needs to have at least 150 individuals of each sex.¹²

Indicators of Severity

It has been traditional to express deviations from the expected reference value as a percent of predicted, and these percentages are widely used to categorize grades of severity. Individual values can also be reported as a percentile of the normal population. While this provides some information within the low normal range that might be helpful in risk assessment, it breaks down for the broader distribution of abnormal values, as these would all be compressed from the 5th percentile to near zero. Since the 5th percentile LLN is equivalent to predicted - 1.645 × SEE, the LLN can also be described as a z-score of -1.645, with larger negative numbers showing increased severity. Reporting the z-score has the advantage that the LLN is

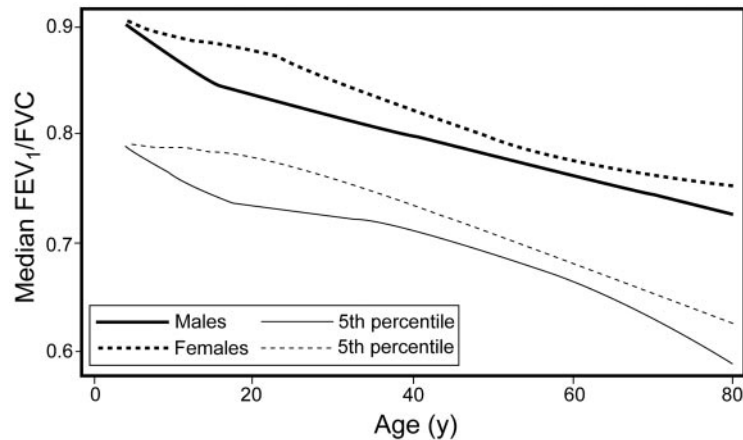


Fig. 1. Using the lambda, mu, sigma method, the sigma and lambda terms allow the 5th percentile lower limit of normal to be independently calculated throughout the age-height spectrum. For the FEV₁/FVC ratio (above) and for both FVC and FEV₁ (not shown), the normal range is wider in young children and in older adults. (From Reference 10, with permission.)

the same number (-1.64) for all spirometric parameters. This could be quite helpful in weaning practitioners away from the flawed 80% shortcut, but its utility for severity grading is less clear. The unit of the z-score tells us how likely this value would be in a normal population, but it is not intrinsically related to the impact of the disease. At this point, knowing the FEV₁ is 40% of normal conveys more meaning to me, and certainly to the patient, than a z-score of -4.9 .

Spirometry Reference Data

For spirometry a great many reference equations exist, some based on rather small samples and others on large populations, some with samples of convenience and others by statistically valid population sampling, and most, recently at least, with efforts to restrict to healthy non-smokers. Many reflect specific geographic or ethnic groups. The NHANES III reference data have been recommended because based upon a large group, appropriately sampled, with good quality spirometry, and provides separate equations for whites, African-Americans, and Mexican-Americans.⁷ The Stanojevic et al publication applies the LMS method to the NHANES data, and merges it with additional pediatric data to extend the age range down to 4.^{10,13} It is thus helpful in tracking the function of growing children without having to jump among equations from preschool, to school age, through adolescence to adulthood. However, these data are currently available only for whites, and, for adults, differ very little from NHANES. New multi-ethnic reference data spanning age 3–90, collated by an ERS Task Force, are expected to be published in 2012.

Accepting Uncertainty

The NHANES spirometry data match quite well with earlier publications by Crapo and others, so, although newer and larger, it does not appear that interpretive results would be much different using the earlier studies.^{3,14,15} We face much greater uncertainties than whether the LLN of a particular individual's FEV₁/FVC ratio should be 0.72 or 0.73. There are the issues of whether a particular reference set matches our patient, our equipment, and the skills of our technicians, but more dominant is the question of the clinical meaning of a value near the LLN in this individual. We know that 95% of healthy non-smokers would have higher values, but that does not tell us the likelihood that this person is normal or has disease. For that we would need to know something about the pretest probability of disease in this individual and among the population seen in our laboratory. If we are testing an asymptomatic non-smoker at the recent World Spirometry Day, then even a value a point or 2 below the LLN may be more likely than not to be "normal." However, in a hospital based referral lab, those with FEV₁/FVC at, or even a few points above, the LLN may be quite likely to have early air-flow obstruction, although we dare not say it. This led one physician many years ago to state, "The general tendency in pulmonary function testing at the clinical level is toward cautiously noncommittal over-interpretation in language replete with modifiers."¹⁶ Rather than the dichotomy of normal/abnormal, we need the risk profiles and outcome data that would allow us to make a rational statement about the probability of disease at any point on the spectrum of FEV₁/FVC values from mid-normal down to the LLN and below. Lacking that, we will have to continue to be "cautiously noncommittal," and although we can cer-

tainly do better than the tarnished GOLD standard of 70%, it seems unnecessary to obsess over the exact LLN.

This uncertainty is also not a new idea. It was the underlying message of the 1968 “cautiously noncommittal” editorial referred to above,¹⁶ and Clausen discussed this very nicely in a 1982 publication of the California Thoracic Society, stating that “the large and inherent overlap between normalcy and disease states will persist as a limitation in pulmonary function test interpretation.”¹⁴ The ATS Spirometry 1994 update made a distinction between the related tasks of “(1) the classification of the derived values with respect to a reference population . . . and (2) the integration of the spirometric values into the diagnosis, therapy and prognosis for an individual patient.” And, noting the difficulty of the latter, concluded, “Accordingly, no specific guidelines for interpretative procedures are recommended that would be applicable to all laboratories.”¹⁷

Spirometry Based Outcome Studies

We are beginning now to see some outcome data addressing the diagnostic criteria for COPD. Mortality follow-up to 22 years after the first NHANE Survey showed that even mild COPD, defined as $FEV_1/FVC < 0.70$ and $FEV_1 > 80\%$ (aka GOLD 1) carried a future risk of excess death, indicated by a hazard ratio (HR) of 1.3.¹⁸ This supports the recommendation to define obstruction by a low ratio, even when the FEV_1 is “normal,” but does not help us interpret borderline values. Although some in this cohort would have had a ratio < 0.70 but above the true LLN, less than 25% of the population was over 60 at the time of spirometry, so the majority, below the LLN, would have been sufficient to drive the outcome of increased mortality.

Subsequently, in response to the 70% controversy, Manino et al used the Cardiovascular Health Study population to report mortality outcomes comparing GOLD criteria to the LLN determined from NHANES III reference data.¹⁹ Baseline spirometry and follow-up for mortality and hospitalizations for up to 11 years were available for 4,965 participants: white and black (Table 3). Participants with spirometry considered normal (ratio > 0.70) but with respiratory symptoms did demonstrate an increased risk of death with HR 1.2. In this population, all > 65 at the time of spirometry, a full 75% of the GOLD 1 strata ($FEV_1/FVC < 0.70$, $FEV_1 > 80\%$) had an FEV_1/FVC above the 5th percentile LLN. The adjusted HR for death in this subgroup was not significantly different from those with normal spirometry, while those with FEV_1/FVC below the LLN, but with $FEV_1 > 80\%$, had a significantly elevated risk of death: HR 1.4. Even in the GOLD 2 strata ($FEV_1/FVC < 0.70$, $FEV_1 < 80\%$, $> 50\%$), one third had a ratio $> LLN$. This group did show an elevated risk of death (HR 2.2). The conclusion was that older individuals with

Table 3. All Cause Mortality by $FEV_1/FVC < 0.70$ Versus Fifth Percentile of the LLN

no.		Adjusted Hazard Ratio
921	$FEV_1/FVC > 0.70$, $FVC > 80\%$ but with respiratory symptoms	1.2
862	$FEV_1/FVC < 0.70$ but $> LLN$, $FEV_1 > 80\%$	1.1 (NS)
289	$FEV_1/FVC < 0.70$ and $< LLN$, $FEV_1 > 80\%$	1.4
266	$FEV_1/FVC < 0.70$ but $> LLN$, FEV_1 50–80%	2.2
424	$FEV_1/FVC < 0.70$ and $< LLN$, FEV_1 50–80%	1.7

LLN = lower limit of normal
(Adapted from Reference 19.)

$FEV_1/FVC < 0.70$ but “normal” by LLN were more likely to die, but this was shown only for the ~25% that also had a reduced FEV_1 . All of these groups did show an increased risk of “COPD related” hospitalizations in the ensuing years. These data do support the idea that individuals with FEV_1/FVC above the LLN, but with respiratory symptoms or borderline values, are at risk for morbidity. However, if we wish to incorporate that into our interpretations, it should be done in a more systematic, age independent, fashion, such as raising the LLN to the 7.5 or 10th percentile in those with risk factors for airway disease.

The latter strategy was recently considered by Fragoso et al, who used white participants in NHANES III aged 40–80 (not just the healthy non-smokers used for prediction equations), recalculated their predicted FEV_1/FVC ratio from the Stanojevic LMS data, then looked at mortality at 6–12 years follow-up (Table 4).²⁰ Those with a ratio less than the 5th percentile LLN were compared to those in successively higher 5 percentile strata, and to a reference group with ratio > 25 th percentile. By this analysis only those with ratio $< LLN$ (including all degrees of severity) had an elevated risk of death over this relatively short-term (adjusted HR of 1.68). The borderline group of interest, between the 5–10 percentiles, had a modestly elevated point estimate, but the confidence interval spanned 1.

Thus, while there are certainly people progressing toward COPD who have spirometry within the normal range, these studies do not suggest major short-term consequences. However, these were studies of general population samples, so not likely to reflect the much higher likelihood of early disease among those referred for pulmonary function testing. And, for COPD, the time from borderline spirometry to major morbidity or death is likely to be over 20 years—a serious challenge for outcome studies. The HRs calculated are adjusted for smoking exposure to reflect only the added impact of low spirometry, but, for prognosis, the combined risk might be more relevant. For example, in the Fragoso et al study, the unadjusted HR for death for participants with FEV_1/FVC between the 5–10th

HOW SHOULD THE LOWER LIMIT OF THE NORMAL RANGE BE DEFINED?

Table 4. All Cause Mortality Stratified by Centiles of FEV₁/FVC*

FEV ₁ /FVC†	Participants no. (%)	Deaths‡ no. (%)	Hazard Ratio for All Cause Mortality	
			Unadjusted	Adjusted
> LMS-LLN 25	2,123 (61.4)	335 (15.8)	1.00	1.00
LMS-LLN 20–25	191 (5.5)	38 (19.9)	1.26 (0.85–1.88)	1.24 (0.85–1.83)
LMS-LLN 15–20	210 (6.1)	43 (20.5)	1.28 (0.97–1.70)	1.24 (0.92–1.66)
LMS-LLN 10–15	210 (6.1)	34 (16.2)	0.99 (0.76–1.29)	1.13 (0.85–1.49)
LMS-LLN 5–10	248 (7.2)	63 (25.4)	1.59 (1.22–2.08)	1.26 (0.95–1.66)
< LMS-LLN 5	477 (13.8)	138 (28.9)	1.92 (1.52–2.42)	1.68 (1.34–2.12)

* National Health and Nutrition Examination Survey III participants aged 40–80 y.

† FEV₁/FVC lower limit of normal (LLN) by lambda, mu, sigma (LMS) data of Stanojevic.¹⁰

‡ Mortality follow-up after 6–12 y

(Adapted from Reference 20.)

percentiles was significantly elevated at 1.59, even with the relatively short 6–12 year follow-up.

So dealing with spirometry results near the LLN will continue to be problematic, but a more important task for the pulmonary community is to focus on finding the pool of individuals with clear-cut, but undiagnosed, COPD. And for this, good quality spirometry remains the best tool.

Lung Volume Reference Data

For lung volumes, the reference data question is quite different: not, which of many reference sets to use, but rather, are there any out there that should be used? The 2005 ATS/ERS documents included a table listing the lung volume reference data published 1993–2002, and among many small, ethnically specific studies there were none dealing with North American populations and only a few for specific European countries. For guidance, reference was made to a 1995 paper in which Stocks and Quanjer lamented the unsatisfactory state of lung volume reference data and republished the 1993 European Community for Steel and Coal reference equations.^{21,22} These, in turn were first published in 1983 and resulted from a merging of many previously published studies, mostly done in the 1960s and 1970s. Unlike the current international effort, which is using original data, the European Community for Steel and Coal equations were mathematically developed from the prior published equations: a method that adds an additional layer of uncertainty, particularly to the normal range. Also in common use are the data of Crapo et al,²³ done in the same population from which spirometry and D_{LCO} reference data were published. This has the advantage that the FVC predicted for spirometry matches the VC predicted with the lung volumes on the same report. The TLC was obtained from the single-breath helium dilution of the diffusing capacity test, and the number studied (123 M, 122 F) was, perhaps, marginal, but was well

distributed over the age range. Most PFT software programs still offer the Goldman-Becklake data of 1959. These are based on a small sample of convenience (44 M, 50 F) from Johannesburg, South Africa, with no mention of smoking status.²⁴ There are some newer data available, reporting plethysmographic lung volumes of healthy non-smokers, including 482 white adults from the Barcelona area²⁵ and 212 New Zealand adults of European background.²⁶ By collating these and others we can hope for better lung volume reference data if the Global Lungs Initiative group maintains its energy beyond spirometry.

Diffusing Capacity Reference Data

The reference data problem for diffusing capacity has been widely and long recognized. While spirometry predictions vary by up to a few percent, the variation among reference values for D_{LCO} can be up to 40%. This reflects the fact that the test itself, even when nominally done by the same technique, varies considerably from lab to lab. Recent equipment has become better standardized, at least out of the box, but with time and in different laboratories, the variation returns.²⁷ The latter problem can be addressed through careful biologic controls or by the simulator device available.²⁸ Still, labs are well advised to select from the available reference sources the set that best matches a diverse group of normal subjects tested in that lab. This is becoming an increasing problem as patients may be excluded from surgery or chemotherapy due to a D_{LCO} percentage cutoff, such as < 40%, but it is important to know, 40% of what? In reviewing the preoperative literature, I have usually not found a reference source for the predicted D_{LCO} stated. If our labs are not using reference data well matched to our equipment and technique, we will be doing a disservice to our patients and colleagues.^{29,30}

Reporting the Numbers

Having given due diligence to all of the above considerations, our final task in the PFT laboratory is to produce a report that conveys the data and our interpretation clearly and efficiently. The ATS/ERS guidelines make some recommendations for reporting, and the ATS PFT committee has further discussed more specific formatting. There is general consensus that only the more meaningful tests be reported (eg, for spirometry the FVC, FEV₁ and FEV₁/FVC ratio suffice). Unless careful attention is paid to the wide normal range, the FEV₂₅₋₇₅ may be more often misleading than helpful, and flows at specific expired volumes do not add utility. Vedal and Crapo showed that, with 14 parameters reported, at least one would fall below the 5th percentile LLN in 24% of a healthy group.³¹ Given the need for an individually specific predicted value for each parameter, it has been common to list this first, but most other clinical laboratory reports show the actual value first, and this variability occasionally leads to the wrong value being transcribed into notes. There is strong consensus that uniformity among PFT labs would be desirable, and a poll of the speakers at this conference showed a 70% preference to list the actual value first. As discussed above, an indicator of the LLN should be included, both for the interpreter and the subsequent readers of the report. Options include the numerical value of the LLN, a percentile or a z-score. Showing only the percent predicted encourages continued use of the 20% "rule," but this index is widely, and I think helpfully, used in severity assessment, so I do not think it can be abandoned. Perhaps seeing FEV₁ or TLC values of 75% but above the LLN, or 82% but below the LLN, will help in the educational process. It is more than a bit disappointing to reread the thoughtful "cautiously noncommittal" editorial written when I was in medical school and still be discussing, 40 years later, the issues of a fixed percentage FEV₁/FVC and \pm 20% normal range, and the overlap of health and disease so well outlined then.¹⁶

REFERENCES

- American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991; 144(5):1202-1218.
- Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26(5):948-968.
- Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981;123(6):659-664.
- Clausen JL. Prediction of normal values. in Clausen JL, editor. *Pulmonary function testing: guidelines and controversies*. New York: Academic Press; 1980.
- Bates DV, Macklem PT, Christie RV. *Respiratory function in disease*, 2nd edition. Philadelphia: Saunders; 1971.
- Morris J, Temple W, Koski A. Normal values for the ratio of FEV₁ to FVC. *Am Rev Respir Dis* 1973;108(4):1000-1003.
- Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999;159(1):179-187.
- Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al; Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;176(6):532-555.
- Pistelli F, Bottai M, Viegi G, et al. Smooth reference equations for slow vital capacity and flow-volume curve indexes. *Am J Respir Crit Care Med* 2000;161(3 Pt 1):899-905.
- Stanojevic S, Wade A, Stocks J, Hankinson J, Coates AL, Pan H, et al. Reference ranges for spirometry across all ages. *Am J Respir Crit Care Med* 2008;177(3):253-260.
- Quanjer PH, Stanojevic S, Stocks J, Hall GL, Prasad KV, Cole TJ. Changes in the FEV₁/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010;36(6):1391-1399.
- Quanjer PH, Stocks J, Cole TJ, Hall GL, Stanojevic S. Influence of secular trends and sample size on reference equations for lung function tests. *Eur Respir J* 2011;37(3):658-664.
- Stanojevic S, Wade A, Cole TJ, Lum S, Custovic A, Silverman M, Hall GL, et al; Asthma UK Spirometry Collaborative Group. Spirometry centile charts for young Caucasian children: The Asthma UK Collaborative Initiative. *Am J Respir Crit Care Med* 2009;180(6):547-552.
- Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983;127(6):725-734.
- Glindmeyer HW, Lefante JJ, McColloster C, Jones RN, Weill H. Blue-collar normative spirometric values for Caucasian and African-American men and women aged 18 to 65. *Am J Respir Crit Care Med* 1995;151(2 Pt 1):412-422.
- Poirier KP. A quantitative definition of obstructive lung disease (editorial). *Am J Med* 1968;45(3):329-335.
- American Thoracic Society. Standardization of spirometry: 1994 update. *Am J Respir Crit Care Med* 1995;152(3):1107-1136.
- Mannino DM, Buist AS, Petty TL, Enright P, Redd SC. Lung function and mortality in the United States: data from the first National Health and Nutrition Examination Survey follow up study. *Thorax* 2003;58(5):388-393.
- Mannino DM, Buist AS, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007;62(3):237-241.
- Fragoso CA, Concato J, McAvay G, Van Ness PH, Rochester CL, Yaggi HK, Gill TM. The ratio of FEV₁ to FVC as a basis for establishing chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;181(5):446-451.
- Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. ATS Workshop on Lung Volume Measurement. Official Statement of the ERS. *Eur Respir J* 1995;8(3):492-506.
- Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yemault JC. Lung volumes and 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):5-40.
- Crapo RO, Morris AH, Clayton PD, Nixon CR. Lung volumes in healthy nonsmoking adults. *Bull Eur Physiopathol Respir* 1982;18(3):419-425.

24. Goldman HI, Becklake MR. Respiratory function tests: normal values at median altitudes and the prediction of normal results. *Am Rev Tuberc* 1959;79(4):457-467.
25. Roca J, Burgos F, Barbera J, Sunyer J, Rodriguez-Roisin R, Castell-sague J, et al. Prediction equations for plethysmographic lung volumes. *Respir Med* 1998;92(3):454-460.
26. Marsh S, Aldington S, Williams M, Weatherall M, Shirlcliffe P, McNaughton A, et al. Complete reference ranges for pulmonary function tests from a single New Zealand population. *N Z Med J* 2006;119(1244):U2281. Erratum in: *N Z Med J* 2007;120(1254):U2551.
27. Jensen R. Personal communication.
28. Jensen R, Leyk M, Crapo R, Muchmore D, Berclaz PY. Quality control of D_{LCO} instruments in global clinical trials. *Eur Respir J* 2009;33(4):828-834.
29. Culver BH. Preoperative management of the thoracic surgery patient: pulmonary function testing. *Semin Thorac Cardiovasc Surg* 2001;13(2):92-104.
30. Ferguson MK, Vigneswaran WT. Diffusing capacity predicts morbidity after lung resection in patients without obstructive lung disease. *Ann Thorac Surg* 2008;85(4):1158-1164.
31. Vedal S, Crapo RO. False positive rates of multiple pulmonary function tests in healthy subjects. *Bull Eur Physiopathol Respir* 1983;19(3):263-266.

Discussion

Coates: Could we come back to one of the issues that I raised in pediatrics yesterday? That was the exclusion criteria in the NHANES III meant more than a quarter of the children were excluded. A few for smoking, most because they said they coughed occasionally or had physician-diagnosed asthma or wheezing or tightness in the chest. On the other hand, three quarters of the sample of adults were excluded. I'm not saying it is right or wrong, but we've talked a lot about not having a gold standard for asthma or COPD. Did we deal only with super-normals? What is the normal population? It seems very different than how we would come up with predictive values and LLN for a biochemical value such as serum sodium.

Culver: Yes, but it's not that different than what we've done with cholesterol; we don't just take a population average of everybody's cholesterol. There have been outcome data, and our lab says "less than 200," but I think there are now guidelines that say even less than 180 is what we should be aiming for. That's not a population average but an idealized number. The bone density people primarily relate to young healthy normals, not to the general population of their aged bone loss compatriots.

If we're trying to detect a disease, we don't want to include that same disease in the population, so I think we want a healthy population. Whether we've gone too far in excluding ev-

erything, I'm not sure. If you think the reference data aren't sufficiently inclusive, you might want to use the ± 1.96 standard deviation for your lower limit to be a little more cautious, at least in people who you're screening, if you're doing that kind of thing, with minimal symptoms. I still think it's rational to try and select a healthy non-smoking population for our reference data.

Coates: I guess the corollary to the question, then, is, when you start putting various series together, the criteria for deciding if you're healthy or not should be pretty similar, if not ideally the same.

Culver: Yes, and that's one of the things that this global initiative is dealing with, because that's defined differently for different studies, and better known in some than others.

MacIntyre: I want to focus on diffusing capacity, if I may. I'm not a statistician so I don't completely understand LLN calculations. I was told of a case of an older patient whose predicted D_{LCO} was 15, and if you do the LLN calculation, it comes out to be 7, and that means that 40% predicted is still normal. The referring physician asked the very valid question, "What do I do with that kind of information?" My first question is, does the calculation of LLN use the same number for every predicted value? You showed data that say that works for vital capacity, but does it work for D_{LCO} too?

Culver: You should know better than I for diffusion capacity, because you've looked at those data. If the distribution is fairly homoscedastic, if it doesn't narrow down as the numbers get smaller, then, yes, it's the same absolute decrement below the predicted value. And, yes, that does reduce the LLN. We see the same thing with the FEF_{25-75} , to be abnormal when you're 60 years old, it probably has to be about 40% predicted.

MacIntyre: But people don't make major decisions based on the FEF_{25-75} . They *do* make big decisions based on diffusing capacity.

Culver: I think that probably tells us that it's a pretty shaky thing to make big decisions on in older folks.

Miller: I think part of the reason is that most of the reference populations don't have a lot of people at the upper extreme of age. Most of the adult reference populations begin at age 18-20, so there are a lot of those, but there aren't a lot of presumably normal 70 and 80-year-olds. Then you get into the question of those who are normal: are they really super-normal? That's why the Thompson equation¹ for D_{LCO} , which is deliberately over-representing older people, may be useful, because you get shaky statistically when you get to extremes of age and heights. There aren't a lot of 80-inch people in the reference populations either.

1. Thompson BR, Johns DP, Bailey M, Raven J, Walters EH, Abramson MJ. *Predic*

tion equations for a single breath diffusing capacity T_{LCO} in a middle-age caucasian population. *Thorax* 2008;63(10):889-893.

Culver: And even shakier if, as often happens, the computer has simply extrapolated an equation that was designed with data up to age 65 or 70 and may have extrapolated to age 75 or 80: those may not necessarily be linear. If the regression line isn't exactly lined up right, it shoots off into the sky or into the basement as it goes further off the page. I think that in looking at some of these data, that may be the case with the Goldman lung volume data,¹ the female data has a fairly steep age regression, so that when I compared it to one of the others, it's 400 mL higher in younger individuals and 500 mL lower in older individuals.

1. Goldman HI, Becklake MR. Respiratory function tests: normal values at median altitudes and the prediction of normal results. *Am Rev Tuberc* 1959;79(4):457-467.

MacIntyre: I'm going to show my statistical stupidity here. To calculate the LLN, don't you take the predicted value and have a single number that is derived from the standard deviation of the prediction equation?

Enright: Indeed, that is true for many published studies.

MacIntyre: OK, it's a fixed value. And you use that same number whether it's a big diffusion capacity or a little diffusion capacity. I'm just trying to figure out, is that the right way to do this?

Culver: Yes. You can determine that number for any regression, but whether that was a valid number would depend on whether it met criteria for what they call homoscedasticity (ie, the variance is the same around small and large values).

MacIntyre: Now, that's a term I'm not familiar with, but I kind of like it because it explains my conundrum here. We're assuming the diffusion ca-

capacity is homoscedastic, but is that in fact true?

Culver: That might be a data set that, if one applied the LMS method,¹ which can adjust the variance all the way along the age-height spectrum, might come up with differences. This method would not say that the FEV_1/FVC was perfectly homoscedastic. If you fine-tune it, you can find variation as you go along the age spectrum. So you might see that in the diffusion capacity, but it also might get even wider in the elderly. It often tends to because you have fewer people, and maybe with survivor populations in their old age there can be quite a bit more variability. Be careful what you wish for, because it might get worse.

1. Cole TJ, Green PJ. Smoothing reference curves: the LMS method and penalized likelihood. *Stat Med* 1992;11(10):1305-1319.

MacIntyre: I hate to keep harping on this, but it's bothering me a lot. It just seems that this single value to calculate LLN using a predicted value of 35 versus using a predicted value of 15 yields results that seem quite different clinically. Subtracting 8 from 35 and your LLN is now 27. Subtracting 8 from 15 is now 7. Help me here.

Miller: These are the questions we asked our statisticians when we published our data. The answer there is, if you're concerned about these extremes of age and height, you have to get more people who are in that range. You're stuck otherwise. The basic conclusion, that when you're younger or taller, you're below 80% predicted to be abnormal, and when you're shorter or older, you're well below 80% to be abnormal—that still holds. Whether you could fine-tune that offset, that confidence interval, you need people in that age or height group.

Enright: There is light at the end of the tunnel. More than half of us around

the table are directors of PFT labs that test in one week several times the number of healthy people that comprise the reference equations¹ that we're currently using for D_{LCO} . It's very embarrassing. The Global Lungs Initiative group has recently begun gathering D_{LCO} data from large studies of healthy people. Not just the equations as published² in 1983 by the European Community for Steel and Coal, but the raw data. Al Miller has contributed his raw D_{LCO} data to it already, and we already have about 1,000 individuals, but we hope to have many times more by the end of next year. We'll use the updated methods that do not assume on homoscedasticity.

We will include data from older people. The PLATINO [Latin American Project for the Investigation of Obstructive Lung Disease] study³ in Mexico City and in South America have just bought D_{LCO} machines and are going to apply them to general population samples in 5 of their cities. The investigators have promised to offer those data to the Global Lungs Initiative group, and we'll have the data from Australia that you mentioned.

So there is some light at the end of the tunnel, and I urge all of you to, if you know a study of normal population samples with D_{LCO} (especially in children and elderly, because they're sorely lacking), please offer them to the group. I think it's really a shame that NHLBI [National Heart, Lung, and Blood Institute] is not going to fund a study of many lung volumes in population-based samples. The CDC [Centers for Disease Control and Prevention] with NHANES IV is not doing it; they're almost completely done gathering their spirometry data. It's a major problem that we should have addressed decades ago.

1. Roca J, Burgos F, Sunyer J, Saez M, Chinn S, Antó JM, et al. References for forced spirometry. Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998;11(6):1354-1362.

2. Schrader PC, Quanjer PH, van Zomeren BC, de Groot EG, Wever AM, Wise ME. Selection of variables from maximum expiratory flow-volume curves. *Bull Eur Physiopathol Respir* 1983;19(1):43-49.
3. Pérez-Padilla R, Vázquez-García JC, Márquez MN, Jardim JR, Pertuzé J, Lisboa C, et al; the Latin American COPD Prevalence Study (PLATINO) Team. The long-term stability of portable spirometers used in a multinational study of the prevalence of chronic obstructive pulmonary disease. *Respir Care* 2006;51(10):1167-1171.

Kaminsky: Bruce, I just wanted to make one comment. The theme here is how to interpret results. One thing you didn't mention but is very important is following results over time. As we all know, we can have someone come in with a FEV₁ of 110% predicted but a low ratio, and they have symptoms of cough or wheeze with exercise, they may still have asthma. Likewise for the diffusion capacity: if

it's dropping over time more than you would expect based on their age—I determine that just by looking at percent predicted—if that percent predicted is dropping where it shouldn't as the person ages and this correlates with something clinically, either desaturation with exercise or progression of disease radiographically, those are the things that help us in this gray zone as well.