Comparison of Fixed Percentage Method and Lower Confidence Limits for Defining Limits of Normality for Interpretation of Spirometry

Ashutosh N Aggarwal MD, Dheeraj Gupta MD, Digamber Behera MD, and Surinder K Jindal MD

BACKGROUND: The use of the lower 90% confidence limit of the lower limit of normal (LLN_{CI}), rather than a fixed percentage of the predicted value ($LLN_{\%}$), appears to be statistically more appropriate for interpretation of spirometry results. There has been no comparative assessment of these 2 definitions of the LLN in routine clinical practice. METHODS: We studied results of spirometry interpretations made with these 2 approaches, and assessed various factors that influence discordant classification of spirometry results. Spirometry records from 18,112 consecutive adult patients referred for spirometry were interpreted as normal, obstructive, or restrictive, based on both LLN_{CI} and $LLN_{\%}$. Discordant results were analyzed using multiple logistic regression techniques to identify variables that significantly affected discordant classification of results. RESULTS: Overall, 11.7% of the results were discordant between the 2 methods. Agreement between the 2 methods, calculated using the kappa estimate, was poorer with spirometry values from women and from patients at the extremes of height and age. Age, sex, and height independently influenced discordant classification. Limits of agreement between LLN_{CI} and $LLN_{\%}$ were wide for all the spirometric variables studied—more so in women and in shorter and older patients. CONCLUSIONS: LLN_{CI} and $LLN_{\%}$ yielded different interpretations of spirometry data in several instances, and the 2 methods cannot be used interchangeably. When interpreting spirometry data in routine clinical practice, LLN_{CI} should be preferred over $LLN_{\%}$. *Key words: lower limit of normal, pulmonary function test, PFT, spirometry, obstruction, restriction.* [Respir Care 2006;51(7):737–743.

Introduction

All algorithms for interpretation of spirometry values aim at identifying subjects whose values lie below the normal range. Despite extensive guidelines regarding principles and methods of interpreting pulmonary function test results, there is still is no uniformity among different laboratories in actual practice of comparing an individual observation to its predicted value. The simplest and traditionally used method in India is to express the observed value as a percentage of the predicted value, and to define the lower limit of normal (LLN) as an arbitrary fixed percentage.

SEE THE RELATED EDITORIAL ON PAGE 719

Many pulmonary-function laboratories define the LLN as 80% of the predicted value for vital capacity (VC) and forced expiratory volume in the first second (FEV_1), and 70% of the predicted value for the FEV_1/VC ratio.[1] Using a fixed percentage (such as 70%) of observed FEV_1/VC ratio completely ignores changes in predicted values in relation to age and differences in predicted values in relation to height and sex. Use of a fixed percentage of the predicted FEV_1 or VC can be justified only if the absolute data, from which a regression equation is derived, has a heteroscedastic distribution, such that the spread of data varies in proportion to a change in a predictor variable.

Ashutosh N Aggarwal MD, Dheeraj Gupta MD, Digamber Behera MD, and Surinder K Jindal MD are affiliated with the Department of Pulmonary Medicine, Postgraduate Institute of Medical Education and Research, Chandigarh, India.

Correspondence: Ashutosh N Aggarwal MD, Department of Pulmonary Medicine, Postgraduate Institute of Medical Education and Research, Sector 12, Chandigarh 160012, India. E-mail: ashutosh@indiachest.org.

Such a distribution may sometimes be seen in children.[2] In adults, however, ventilatory data tend to have a homoscedastic scatter; that is, the variance of data around the regression line tends to be uniform.[3,4] Thus, the limits of normality can be described by confidence bands that lie parallel to the regression line. Therefore, a more valid, and the currently recommended, approach is to define the LLN as the 5th percentile, or the lower 90% confidence limit, of the predicted value.[5,6] The 95% confidence limits can be statistically determined by subtracting or adding 1.96 times the standard error of the estimate from the predicted value. Since the upper limits are not relevant in clinical spirometry, the 5% error can be transferred to one end of the curve and, in this way, the confidence limit can be calculated at 90%, using 1.645 times the standard error of the estimate.[7] This is sometimes also referred as the 95% confidence limit by one-tailed test.

Some investigators have shown that the use of the fixed percentage leads to discordance in classification of spirometry data, when compared to the use of the lower 90% confidence limit.[7–11] However, there is very little information on the magnitude of these abnormalities or the factors that influence them in routine clinical practice. We hypothesized that there would be significant differences in interpretation of spirometry results using the 2 methods, and we studied the nature, contributing factors, and magnitude of such differences, using data from routine spirometry examinations performed at our pulmonary-function laboratory.

Table 1. Age, Height, and Sex Distribution in the Study Population

Height (cm)	Male	Female	Total
131–140	7	96	103
141–150	107	2,131	2,238
151–160	1,708	4,601	6,309
161–170	5,470	1,091	6,561
171–180	2,663	32	2,695
181–190	206	0	206
Age (y)			
16–25	1,107	857	1,964
26–35	1,452	1,545	2,997
36–45	1,831	1,865	3,696
46–55	2,250	1,789	4,039
56–65	2,080	1,252	3,332
66–75	1,136	517	1,653
> 75	305	126	431
Total	10,161	7,951	18,112

Methods

Over a 5-year period, 18,112 consecutive spirometry records from adult patients (age > 15 years) were analyzed. The reasons for performing spirometry and other clinical details were not analyzed further. There were 10,161 (56.1%) records from male patients and 7,951 (43.9%) from female patients. The mean ± SD age was 48.2 ± 16.2 years among the men and 44.8 ± 14.8 years among the women ($p < 0.001$). Mean ± SD height was 166.5 ± 6.9 cm among the men and 153.9 ± 6.0 cm among the women ($p < 0.001$) (Table 1).

Spirometry Interpretation

With all the patients, spirometry was performed using a dry rolling seal spirometer (Spiroflow, PK Morgan, Kent, United Kingdom). Spirometric indices such as VC, FEV_1, and FEV_1/VC ratio were measured by experienced technicians, using American Thoracic Society guidelines.[12] The highest values from among 3 technically acceptable and reproducible spirometry maneuvers were expressed at body temperature and pressure saturated with water vapor. The predicted value for each of these measured variables was calculated using regression equations for healthy adults from northern India, which were previously derived by us.[13] These equations were generated from spirometry studies performed on 962 healthy nonsmoker adult volunteers from northern India (540 men and 422 women), ages 15–74 years, using a water-seal spirometer. These equations express spirometry variables as a function of age, height, and sex of the subject (Table 2). LLN for each variable was calculated as the difference between the predicted value and 1.645 times the standard error of the estimate of the relevant regression equation. This value represents the lower 90% confidence limit of the predicted value (LLN_{CI}).[6] LLN was also calculated for each variable, using the fixed-percentage method ($LLN_{\%}$). VC or FEV_1 < 80% of the predicted value and FEV_1/VC ratio < 70% were considered abnormal.

FEV_1, VC, and FEV_1/VC were used as the basic variables to interpret spirometry data.[6] If the subject's FEV_1/VC was less than the LLN for that subject, he or she was categorized as having an obstructive pattern. If the subject's VC was less than the LLN associated with a normal FEV_1/VC ratio for that subject, he or she was categorized as having a restrictive pattern. Classification of spirometry data into normal, obstructive, and restrictive patterns was performed using both LLN_{CI} and $LLN_{\%}$.

Data Analysis

Spirometry results (normal, obstructive, or restrictive pattern) obtained using LLN_{CI} and $LLN_{\%}$ were compared after construction of contingency tables. Agreement between 2 methods was calculated using the kappa estimate for the entire population and for the different categories of age and height.[14] Spirometry results that

Table 2. Regression Equations for Prediction of Normal Lung-Function Values in North-Indian Adults

Variable	Sex	Regression Equation	SEE
FEV_1 (L)	Male	$-1.90 - 0.025A + 0.00006A^2 + 0.036H$	0.417
	Female	$-1.07 - 0.030A + 0.00013A^2 + 0.027H$	0.323
VC (L)	Male	$-3.44 - 0.013A - 0.00005A^2 + 0.048H$	0.497
	Female	$-2.05 - 0.014A - 0.00004A^2 + 0.035H$	0.447
FEV_1/VC (%)	Male	$103 - 0.35A + 0.002A^2 - 0.07H$	6.6
	Female	$111 - 0.36A + 0.003A^2 - 0.10H$	5.8

FEV_1 = forced expiratory volume in the first second
VC = vital capacity
A = age
H = height
SEE = standard error of the estimate

Table 3. Spirometry Results Categorized on Basis of Lower Limit of Normal Defined by Lower 90% Confidence Limit Versus Fixed-Percentage Method

Using $LLN_{\%}$	Using LLN_{CI}: Normal Spirometry		Obstructive Pattern		Restrictive Pattern		Total (*n*)	(%)
	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)		
Normal spirometry	7,297	85.8	807	15.8	39	0.9	8,143	45.0
Obstructive pattern	250	2.9	3,791	74.3	232	5.2	4,273	23.6
Restrictive pattern	958	11.3	507	9.9	4,231	94.0	5,696	31.4
Total	8,505	100.0	5,105	100.0	4,502	100.0	18,112	100.0

LLN_{CI} = lower limit of normal defined by lower 90% confidence limit
LLN% = lower limit of normal defined by fixed percentage method

had different interpretations using LLN_{CI} and $LLN_{\%}$ were identified. Comparisons were made between these and the results from other subjects, to assess the factors that influence discordance between results. Logistic regression analysis was performed to study the relative contributions of these factors.

The difference between the 2 methods for LLN for VC, FEV_1, and FEV_1/VC was computed for each subject. These differences were calculated by subtracting the lower 90% confidence limit from the 80%-of-predicted value. The precision of these estimates of differences was assessed by calculating limits of agreement, which are the 95% confidence limits for the mean difference.[15] The span of these limits was assessed to evaluate if these 2 methods could be used interchangeably.

Results

Overall, 8,505 (47.0%) subjects had normal spirometry, and 5,105 (28.2%) and 4,502 (24.9%) subjects, respectively, had obstructive and restrictive patterns, using LLN_{CI}. A substantial proportion of the subjects with normal spirometry and obstructive airflow pattern (11.3% and 9.9%, respectively) were classified as having restrictive pattern using $LLN_{\%}$ (Table 3). Conversely, 15.8% of subjects with an obstructive pattern interpreted based on LLN_{CI} were classified as having normal spirometry using $LLN_{\%}$. Of the 9,969 subjects classified as having abnormal spirometry using $LLN_{\%}$, 1,208 (12.1%) were classified as having a normal spirometry when calculated with LLN_{CI}. The diagnoses obtained by using LLN_{CI} and $LLN_{\%}$ were identical in 15,319 (84.6%) instances, with an overall kappa estimate of agreement of 0.761 (Table 4).

There were 2,793 discordant results, yielding an overall discordance rate of 15.4%. The kappa estimate showed fair or poor agreement between the 2 methods in almost all categories of height and age among women (see Table 4). More women had discordant results than did men, in all categories of age and height. Among men, agreement was poor in shorter and older subjects (see Table 4). Subjects at extremes of age and height had more discordant results.

Age, height, and sex significantly and independently influenced discordance of spirometry results (Table 5). Women, subjects older than 55 years, and those who were $<$ 160 cm tall had significantly higher odds of having a discordant result using $LLN_{\%}$.

Among women, LLN for VC was consistently overestimated by the fixed-percentage method. The agree-

Table 4. Discordance and Agreement Between Spirometry Interpretation Based on Lower Limit of Normal Defined by Lower 90% Confidence Limit Versus Fixed-Percentage Method

Height (cm)	Men			Women			Total		
	Discordant Results		Kappa Estimate	Discordant Results		Kappa Estimate	Discordant Results		Kappa Estimate
	(*n*)	(%)		(*n*)	(%)		(*n*)	(%)	
131–140	2	28.6	0.533	44	45.8	0.335	46	44.7	0.354
141–150	17	15.9	0.764	579	27.2	0.581	596	26.6	0.590
151–160	196	11.5	0.826	1,050	22.8	0.641	1,246	19.7	0.692
161–170	478	8.7	0.867	212	19.4	0.686	690	10.5	0.839
171–180	190	7.1	0.888	4	12.5	0.787	194	7.2	0.886
181–190	21	10.2	0.840	NA	NA	NA	21	10.2	0.840
Age (y)									
16–25	91	8.2	0.863	172	20.1	0.675	263	13.4	0.779
26–35	47	3.2	0.945	347	22.5	0.635	394	13.1	0.780
36–45	78	4.3	0.933	417	22.4	0.641	495	13.4	0.786
46–55	140	6.2	0.905	398	22.2	0.656	538	13.3	0.795
56–65	272	13.1	0.804	328	26.2	0.598	600	18.0	0.728
66–75	215	18.9	0.718	178	34.4	0.487	393	23.8	0.643
> 75	61	20.0	0.699	49	38.9	0.433	110	25.5	0.620
Total	904	8.9	0.864	1,889	23.8	0.627	2,793	15.4	0.761

NA = Not applicable, because none of the women tested were ≥ 181 cm tall.

Table 5. Multiple Logistic Regression Model to Evaluate Influence of Age, Height, and Sex on Discordant Spirometry Results, As Assessed by Lower Limit of Normal Defined by Lower 90% Confidence Limit Versus Fixed-Percentage Method

	Regression Coefficient (B)	Standard Error of B	Odds Ratio (e^B)	95% Confidence Limits for Odds Ratio
Constant	−1.450	0.056	0.235	
Age (y)				
16–25	0.027	0.083	1.028	0.874–1.209
26–35	−0.079	0.073	0.924	0.801–1.066
36–45	−0.050	0.069	0.951	0.832–1.088
46–55*	–	–	1.000	–
56–65	0.453	0.067	1.573	1.380–1.793
66–75	0.911	0.077	2.487	2.136–2.894
> 75	1.028	0.125	2.795	2.186–3.573
Height (cm)				
131–140	0.913	0.206	2.492	1.663–3.733
141–150	0.148	0.059	1.160	1.032–1.302
151–160*	–	–	1.000	–
161–170	−0.219	0.062	0.803	0.711–0.907
171–180	−0.389	0.095	0.678	0.562–0.817
181–190	0.053	0.240	1.055	0.660–1.687
Sex				
Male*	–	–	1.000	–
Female	1.046	0.062	2.846	2.519–3.214

*Reference category

ment between LLN_{CI} and $LLN_{\%}$ worsened with decreasing stature and advancing age (Fig. 1). Among men, LLN_{CI} was consistently less than $LLN_{\%}$ for shorter and older subjects.

Among women, LLN_{CI} for FEV_1 was consistently less than $LLN_{\%}$ for subjects < 160 cm tall or > 45 years old; in other groups the limits of agreement were spread on either side of zero (Fig. 2). Among men, LLN_{CI} was consistently less than $LLN_{\%}$ for subjects < 160 cm tall or > 55 years old. The agreement between LLN_{CI} and $LLN_{\%}$ worsened with decreasing stature and advancing age.

LLN_{CI} for FEV_1/VC ratio was consistently > 70% among women, and agreement between LLN_{CI} and $LLN_{\%}$ worsened with decreasing stature and at extremes of age (Fig. 3). Among men, consistent differences between both definitions of normality were more noticeable for subjects older than 45 years (see Fig. 3).

Limits of agreement between LLN_{CI} and $LLN_{\%}$ for FEV_1, VC, and FEV_1/VC were wide for all age and height groups, for both men and women. Overall, the percentage of predicted that would match the LLN_{CI} for VC ranged from 54.4% to 84.7% in men and from 44.2% to 80.4% in women. Similarly, the percentage of predicted that would match the LLN_{CI} for FEV_1 ranged from 52.5% to 84.6% in men and from 47.9% to 83.1% in women.

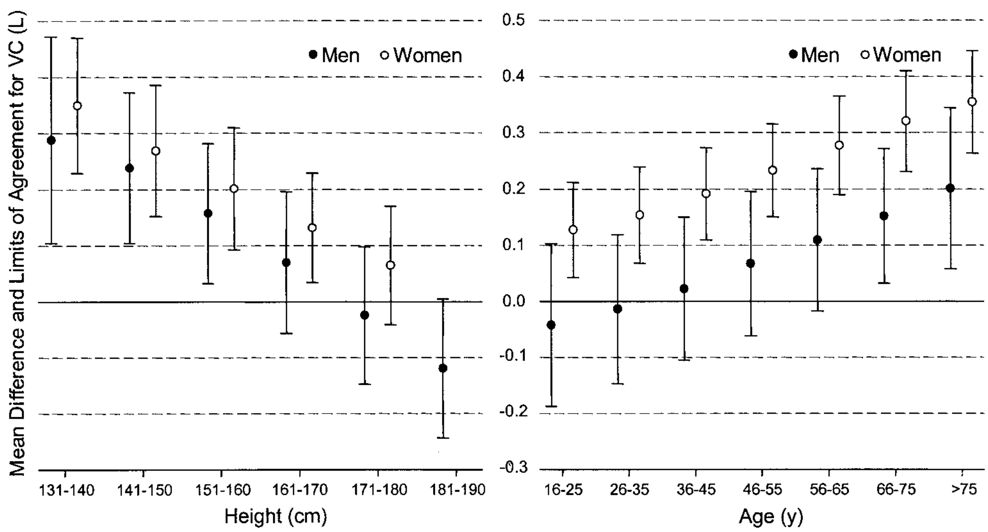

Fig. 1. Mean difference (bias) and limits of agreement between lower limit of normal vital capacity (VC), as calculated by subtracting the lower 90% confidence limit from the 80%-of-predicted value, for various categories of age and height in men and women.

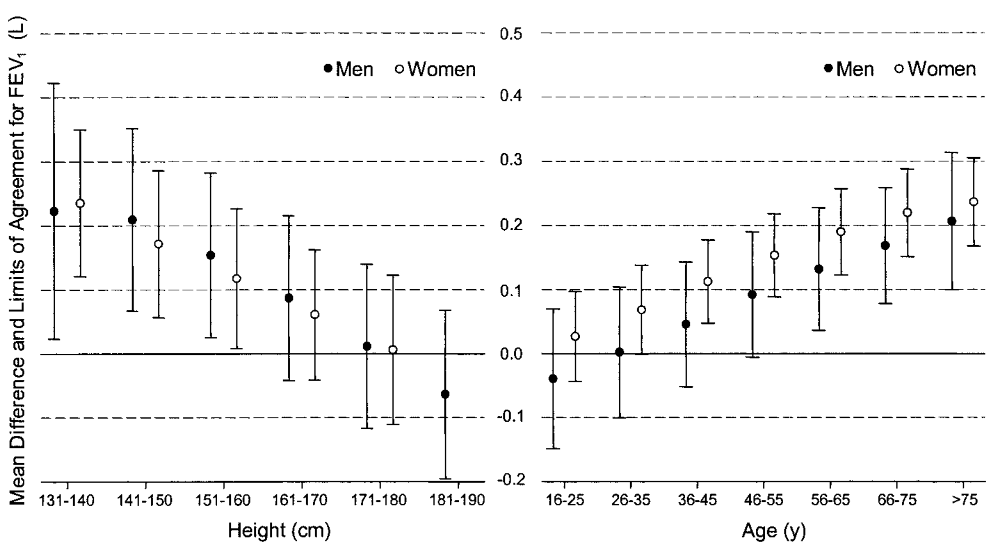

Fig. 2. Mean difference (bias) and limits of agreement between lower limit of normal forced expiratory volume in the first second (FEV_1), as calculated by subtracting the lower 90% confidence limit from the 80%-of-predicted value, for various categories of age and height in men and women.

Discussion

Correct interpretation of spirometry data requires appropriate application of both measurement technique and statistics. The latter is needed primarily to identify individuals whose spirometry values are less than normal. Most clinicians are familiar with the fixed-percentage method to classify results as normal or abnormal. Although this practice is ingrained by usage, there is very little statistical or physiological basis for such an approach. Current guidelines recommend defining LLN as the lower 90% confidence limit of the predicted value.[5,6]

Each of the 2 methods of defining LLNs has its own advantages and problems. Perhaps the greatest drawback of using LLN_{CI} is the inherent complexity involved in calculating these values, and this is the single most important factor responsible for the continued use of the fixed-percentage method. This method may also create some problems in categorizing the severity of impairment observed on spirometry (eg, mild, moderate, or severe defect), which traditionally is based on a fixed decremental percentage of predicted normal value. The use of LLN_{CI} is, however, a more valid approach statistically, and this has been well recognized for several years. Conceptually,

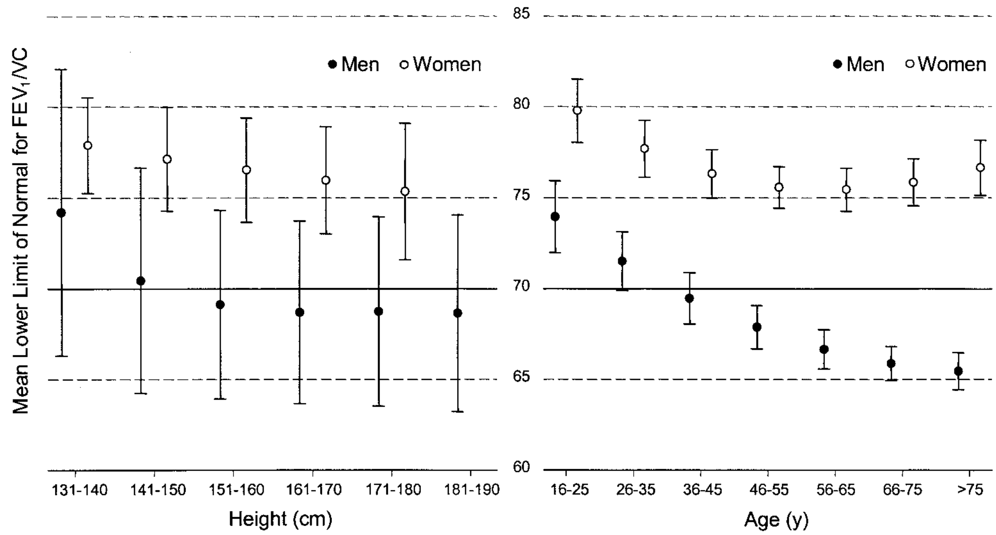

Fig. 3. Mean lower limit of normal for the ratio of forced expiratory volume in the first second (FEV_1) to vital capacity (VC), as defined by the lower 90% confidence limit and the corresponding 95% confidence intervals, for various categories of age and height in men and women.

any fixed percentage of a predicted value designated as a cut-off between normal and abnormal values is an arbitrary figure, with no statistical or epidemiological basis. Even though this fact has long been known, it has been difficult to change the accustomed thinking of clinicians, and some pulmonary-function laboratories continue to rely on the fixed-percentage method. Although LLN_{CI} can closely approximate $LLN_{\%}$ in some situations, such an occurrence is merely fortuitous and occurs over a very narrow range of values.

For example, using our regression equations, LLN_{CI} and $LLN_{\%}$ for a man having a predicted VC of 2.73 L are identical (1.91 L).[13] At all other values, $LLN_{\%}$ differs from LLN_{CI}, with the extent of deviation being a function both of the scatter of the original data and the slope of the regression line.[10] $LLN_{\%}$ for which the predicted value is smaller (for example in short and elderly subjects) will deviate more readily from LLN_{CI} in persons for whom the predicted value is larger (as in tall and young subjects). Although such observations have been emphasized by a few investigators,[7–11] a detailed analysis of the extent of discordance between LLN_{CI} and $LLN_{\%}$ in clinical practice had been lacking.

Another equally important issue is that of using a fixed value (eg, 70%) to separate abnormal FEV_1/VC values. Conceptually, this approach suffers from similar drawbacks, as highlighted previously, in that any single arbitrary fixed cut-off value cannot be a criterion to diagnose airflow obstruction in patients with widely different age and body habitus. Even though standard guidelines advocate the use of LLN_{CI} to define an obstructive pattern, the use of a fixed 70% value is still promoted by some as a standard of diagnosis in patients with asthma and chronic obstructive pulmonary disease. Our data clearly indicate that LLN_{CI} for FEV_1/VC is noticeably different from a 70% figure in a large proportion of subjects, especially women. Use of a fixed 70% value will underdiagnose airway obstruction in young and short men, and all women (see Fig. 3) and should therefore not be recommended as the basic criterion to define an obstructive pattern.

In the patient population we studied, 11.7% of the spirometry results were discordant when the 2 methods of defining LLN were compared. It is true that subjects with unequivocal and severe pulmonary-function defects, having observed values below both LLN_{CI} and $LLN_{\%}$, will be classified similarly, irrespective of the definition of normality used. However, the interpretation of values in the "borderline area" between the 2 LLN definitions will be markedly influenced by the LLN classification scheme. In this group, the results could be categorized as normal using one method or abnormal using the other. It is this group that also requires a precise reporting of results, since a distinction between normal and abnormal pattern may decide whether a particular patient requires further investigation for a possible respiratory disorder. Management decisions (such as suitability for surgery or grant of compensation) may also depend on this interpretation.

Regression equations for ventilatory function in adults have age, height, and sex as accepted variables. Each of these variables may independently contribute to discordance in interpretation. This was also observed in our analysis (see Table 4). When these variables were adjusted for each other in a multiple logistic regression model, they still independently influenced discordant classification (see Table 5). The odds ratio increased progressively for age $>$ 55 years or height $<$ 160 cm. The risk of discordant classification, therefore, increases steeply with decreasing stature or increasing age.

There was no definite pattern of discordance in interpretation of spirometry results when $LLN_{\%}$ was used instead of LLN_{CI}. Although normal and obstructive patterns identified with LLN_{CI} were more commonly classified as restrictive when $LLN_{\%}$ was used, all possible combinations of spirometry results were encountered when these 2 different normality criteria were compared (see Table 3). This occurred because different combinations of age, height, and sex affect discordant classification of VC, FEV_1, and FEV_1/VC differently and to a varying extent. We also studied the influence of choice of normality criteria individually for each of these 3 variables. These results again indicate that the risk of discordant classification for each of these variables is more with women, the elderly, and those of shorter stature. More importantly, limits of agreement for the 2 systems of classification were wide for each variable across most categories of age and height, and more so in women. This suggests that the 2 methods cannot be used interchangeably in the patient population.

Conclusion

The use of different schemes to define LLN when interpreting pulmonary-function data leads to discordant interpretation of results in a significant proportion of subjects in routine clinical practice. Based on statistical considerations, application of confidence limits to define LLN should be preferred over the traditional method of using fixed percentages.

REFERENCES

1. Morris JF, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. Am Rev Respir Dis 1971;103(1):57–67.
2. Dockery DW, Berkey CS, Ware JH, Speizer FE, Ferris BG Jr. Distribution of forced vital capacity and forced expiratory volume in one second in children 6 to 11 years of age. Am Rev Respir Dis 1983; 128(3):405–412.
3. Miller MR, Pincock AC. Predicted values: how should we use them? (editorial) Thorax 1988;43(4):265–267.
4. Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. Thorax 1966;21(5):445–449.
5. Crapo RO, Morris AH, Gardner AM. Reference spirometric values using techniques and equipment that meet ATS recommendations. Am Rev Respir Dis 1981;123(6):659–664.
6. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. Am Rev Respir Dis 1991; 144(5):1202–1218.
7. Quadrelli S, Roncoroni A, Montiel G. Assessment of respiratory function: influence of spirometry reference values and normality criteria selection. Respir Med 1999;93(8):523–535.
8. Harber P, Tockman M. Defining "disease" in epidemiologic studies of pulmonary function: percent of predicted or difference from predicted? Bull Eur Physiopathol Respir 1982;18(6):819–828.
9. Harber P, Schnur R, Emery J, Brooks S, Ploy-Song-Sang Y. Statistical "biases" in respiratory disability determinations. Am Rev Respir Dis 1983;128(3):413–418.
10. Sobol BJ, Sobol PG. Per cent of predicted as the limit of normal in pulmonary function testing: a statistically valid approach. Thorax 1979;34(1):1–3.
11. Oliver LC, Eisen EA, Sprince NL. A comparison of two definitions of abnormality on pulmonary outcome in epidemiologic studies. Am Rev Respir Dis 1986;133(5):825–829.
12. American Thoracic Society. Standardization of spirometry, 1994 update. Am J Respir Crit Care Med 1995;152(3):1107–1136.
13. Jindal SK, Wahi PL. Pulmonary function laboratory in the tropics: needs, problems and solutions. In: Sharma OP, editor. Lung disease in the tropics. New York: Marcel Dekker; 1991: 523–542.
14. Fleiss JL. Statistical methods for rates and proportions. New York: Wiley; 1981: 140–154.
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476): 307–310.