Using spirometry results in occupational medicine and research: Common errors and good practice in statistical analysis and reporting

Abstract

Spirometry appears to be a simple and inexpensive method to measure disorders of the respiratory tract. In reality however, a simple spirometry test requires knowledge and skill to correctly conduct and evaluate the test and its results. This review addresses common misunderstandings in using, evaluating and reporting spirometry results in Occupational Health practice, clinical medicine and research. Results of spirometry need to be evaluated in relation to reference values. The factory medical officer has to decide first whether the test was technically correctly executed and is acceptable for medical interpretation. The next step is to compare results of the individual to published reference values. A 10% reduction of reference values for North Indians and Pakistanis and a 12 to 13% reduction for South Indians is recommended when Caucasian reference tables are used. In occupational health practice the worker's spirometry performance over time needs to be considered. Common errors in reporting summarized results, for instance from groups of workers, are the incorrect use of tests of significance and incorrect presentation of aggregated spirometry results. The loss of respiratory function is recommended as an indicator of difference between two groups. That way, early changes in function can seen without waiting for a drop of function below the usually used 80%-of-predicted limit. This procedure increases the sensitivity of medical surveillance. In research the more precise Lower Limit of Normal should be calculated and used. Correct reference equations, good patient coaching, decision on the technical quality (acceptability) of each spirometry test and critical reevaluation of the machine's readout are essential parts of a correct spirometry test. A good understanding how results are calculated is crucial for further statistical evaluation.

Key words: Epidemiological studies, evaluation, lower limit of normal, lung function test, medical surveillance, spirometry, standards, statistical analysis

INTRODUCTION

Spirometry is commonly used in clinical medicine and research to evaluate effects of exposure on the respiratory system. Factory medical officers use it in their practice to monitor the company workforce. This review addresses some common problems in using, evaluating and reporting spirometry results in companies or for research purposes. Difficulties for Factory Medical Officers and researchers are addressed. The necessary steps for the correct evaluation, reporting and inference from the aggregated results of spirometries are described.

In accordance with the Indian Factories Act, 1948, and the State Factories Rules, company doctors are using spirometry for periodic medical surveillance programs to

- a) monitor healthy workers at risk for respiratory diseases,
- b) monitor chronically ill workers with respiratory disorders as part of their diseases management,
- c) detect early effects of lung diseases in the workforce at a stage where intervention is still advantageous, and in preemployment examinations, we use it to
- d) establish baseline values for new employees who will be exposed to hazards causing respiratory diseases (e.g. dust)^[4] and
- e) monitor and measure respiratory health effects in research.

N. L. Wagner, W. S. Beckett¹, R. Steinberg Department of

Review

Article

Department of Environmental Health Engineering, Sri Ramachandra Medical College and Research Institute, Chennai, India. ¹Department of Environmental Medicine, University of Rochester, Rochester, New York, USA

For correspondence:

N. L. Wagner, Department of Environmental Health Engineering Sri Ramachandra Medical College and Research Institute (DU) 1, Ramachandra Nagar, Porur, Chennai 600 116, India. E-mail: NLWagner@web.de



On first sight, it appears to be a simple and inexpensive method to measure respiratory effects. In reality however, a simple spirometry test requires a high level of knowledge and skill to correctly evaluate (accept or dismiss) the test in reference to its technical quality, to evaluate the results of individual tests for the correct medical evaluation, and to correctly report aggregated test results of groups of workers or research subjects.

Most commonly spirometry is used to evaluate and counsel the *individual* person and their lung function, for instance, in a general practitioner's practice or an Occupational Health Center. Reporting the spirometry results of a whole *group* however is a different issue. This however is a common task for factory medical officers in company surveillance programs and researchers in their studies.

Use of reference values in spirometry

As with any biological test, the results of spirometry need to be evaluated in relation to "normal" or reference values. One step is to compare results of the individual to published reference values that have been measured in a similar population of healthy individuals without lung disease and who do not smoke.

The most important determinants of lung function are

- 1. Height: taller people have greater lung function
- 2. Age: younger adults have higher lung function, and lung function is lost with aging after the age of 25 years
- 3. Sex: women of the same age and height as men have slightly lower lung function.

In order to compare the individual results to healthy persons, reference tables or reference equations have been established by testing several hundred or even thousands of healthy people.^[2] Consensus conferences and professional societies have then published recommendations on how to use these reference tables.

A male worker's individual spirometry result – for instance FEV1: 3.50 l – can be compared to healthy males of his age and height to determine how many per cent of the predicted value of the reference population his results are.

If reference tables are not established for a specific population or they are not part of the integrated program of a spirometer, another approach can be used. The published recommendations explain how to adapt these published reference values for different human races like Europeans, Orientals, Hong Kong Chinese, Japanese, Polynesians, North Indians, South Indians, Pakistanis, and Africans. This adaptation is necessary in order to correctly evaluate the lung function and not to over- or under-estimate the prevalence of a respiratory disorder. This approach is practical and useful when using a spirometer that automatically compares results to, for instance, a Caucasian population.

Several sets of reference values and prediction equations have been published for populations in different parts of India^[3-10] including for Indians living in other countries.^[11] Based on trends in the FVC, the European Respiratory Society recommends a 10% reduction of reference values for North Indians and Pakistanis and a 12 to 13% reduction for South Indians to account for the slightly different body shape when Caucasian reference tables are used in the spirometer's database.^[12]

At least one commercially available spirometer in India has normal values for Indians programmed on its microprocessor so that these comparisons can be made automatically^[13], others perform the above mentioned calculation automatically if the setting for different races is used.^[14]

The reference tables are created by breaking down the study population by sex, height, age, and sometimes weight. Ideally, each single cell in this four-dimensional matrix should be filled with a sufficient number of results from individuals. Then the point estimates are calculated assuming a normal distribution of the results *inside* each single data cell.^[16] An example is given in Box 4.

In reality, there are often not enough healthy individuals tested to fill all cells of this matrix sufficiently. For practical purpose, available data are evaluated; the results for the other cells are interpolated by performing linear regression analysis to plot a line which estimates the normal values for males of a certain height at various ages. The reference equations are thus calculated with the available data. This interpolation creates potential problems for the quality of these resulting equations. They may lead to prediction values that are not accurate because of small numbers, particularly for extremes of age or height. These uncertainties are discussed in depth by other publications.^[16,17]

Evaluating individual results

Detailed guidelines on the technical conduct and the quality of spirometry have been widely published and are common standard of care for doctors in general $practice^{[45-19]}$ and for company doctors using spirometry in occupational settings.^[20,24]

The spirometer computes the best results in absolute numbers and in "percentage of predicted" by comparing the actual results to a reference table of so-called normal values for this specific population. The range of normal values differs according to body proportions which, in turn, differs to some degree between human races. If the spirometer has reference values for different races, in order to have a correct reference population, we have to enter the origin of the patient into the spirometer or computer.

Commercially available spirometers often calibrate themselves according to BTPS-body conditions: normal body temperature (37° C) and ambient temperature, ambient pressure, saturated with water vapor. This requires the spirometer to have a thermometer to measure ambient temperature and a barometer to measure barometric pressure. If these are not built into the spirometer, the examiner has to measure temperature and pressure every day in the examination room and set these parameters manually so that the measured spirometry values can be corrected for variations in temperature and pressure.

The doctor or technician has to decide first whether the test was technically correctly executed and is therefore acceptable for medical evaluation. Criteria for this important step, the acceptability evaluation, are given in Table 1. After acceptance, the results of the tests are interpreted for medical purpose. The medical evaluation is done in the usual way by comparing the individual spirometry results to the "predicted value" for a person of her/his age, sex, height, (and sometimes weight) having determined the race before the test.

Spirometry results (FEV1 and FVC) in healthy male adults are declining in absolute numbers at a rate of about 25 ml per year with age. The reference values or equation include this age-related change. The "percent of predicted" results always refer to the age-adjusted reference value. Therefore, an individual should, in theory, maintain the same "percent of predicted" value throughout life. A drop of 15% compared to the predicted value of, e.g. FEV1 is a significant early warning sign of a beginning obstructive disease.

In occupational health practice though, the worker's spirometry performance over the course of time should therefore be considered. This method of evaluating the change over time is more sensitive to detect early changes than focusing on the printout of the spirometer giving only the actual "percent of predicted" each year.^[22] It is estimated that about half of all workers may benefit from this refined evaluation of spirometries over time.^[23] An example is given in Box 2.

Evaluating and reporting of aggregated results

Often aggregated spirometry results of groups of workers or cohorts are evaluated and reported in publications or company reports. However, the presentation of those aggregated results is often not correct (Table 2). In most cases, these errors occur because of a misunderstanding of the nature of spirometry results.

Table 1: Criteria for technical quality

Criteria for technical quality and acceptability of spirometry tests

- 1. There were no reasons to postpone the spirometry: e.g. a recent respiratory infection, smoking or eating in last hour
- 2. Patient had understood the instructions and has performed with maximum effort
- 3. No sign of hesitation or false starts at the beginning
- No sign of cough: coughing during the first second of the maneuver or at other times that might affect the results
- 5. No sign of variable effort of patient during test (air flow at an inconsistent rate)
- 6. No sign of glottis closure: Sometimes the glottis closes involuntarily, temporarily cutting off the flow of air completely
- 7. No sign of early termination, before an obvious plateau is reached *Definition of "plateau:* no change in air flow for at least 1 second after an exhalation time of at least 6 seconds (10 s are optimal)
- 8. No sign of leaks, which are caused when the person does not have a tight seal around the mouthpiece
- 9. No baseline error. The recording curve must start the tracing of the subject's effort at zero for the volume line
- 10. Reproducibility: the two highest values for FVC and FEV1 taken from acceptable forced expiratory maneuvers show minimal variability after three acceptable maneuvers and the size and shapes of the volume-time curves and flow-volume curves similar

When we deal with results from spirometries, it is not appropriate to assume that the raw aggregated results (for instance, the means or averages measured in liters) of the exposed group can directly be compared with the results of the unexposed group. It is recommended to first express each individual's results as "per cent of predicted" before comparing the exposed to the controls; otherwise we might either *overlook* respiratory disorders or *raise false alarms* in the company (See example in Box 3).

Suggested statistical analysis of aggregated spirometry results

In occupational settings, we actually have two control groups: the reference population, which allows us to calculate the "percent of predicted" values; and an internal control group, for instance, an unexposed group of workers, which allows us to calculate the risks in a specific department of a company. If the group being studied has aggregated results – for instance, a FEV1 of 92% of predicted in average - that are lower than the average for the healthy population, then some factors have caused this reduction in lung function. It is therefore particularly useful to have a baseline set of spirometry results at the beginning of employment. Then, if there is a new abnormality later on, the loss of lung function must have occurred *during* the period of employment. The reasons however can be occupational or extra-occupational. In research studies with industrial cohorts, it is useful to use both, the published "normal values" and also a local control group of unexposed individuals from the same region, the same workplace and similar socioeconomic status for a better additional comparison.

Source: 15, 20

Table 2: Incorrect	evaluation	and	reporting	parameters
--------------------	------------	-----	-----------	------------

Parameters	Exposed workers	Control subjects	P values
	(n=50)	(n=60)	
FVC (I)	3.87±0.85	3.96±0.66	Non significant
FEV1 (I)	3.46 ± 0.84	3.57±0.62	Non significant
FEV1	87.54±10.43	89.79±5.11	Non significant
PEFR (I/min)	469.61 ± 94.05	507.86 ± 69.39	P<0.05

Note on Incorrect Use of Tests of Significance: Here, the exact *P*-values (or the point estimates with confidence intervals) are not given. This, however, is considered good academic standard. the reader cannot form a proper opinion about the scientific value and validity of results if these details are not reported. Tests of statistical significance such as *P*-values depend - among other things - on the sample size of the study. The author might have found something important but the test is below the "significance threshold" just because there were not enough subjects taking part in the study. *P*-value calculations are not appropriate in evaluating spirometry results. They are parametric tests presuming a normal (Gaussian) distribution. Their use assumes that all persons in that group have the identical age, weight and height and that their results in the spirometry should therefore be normally distributed. That is usually not the case

Table 3: Reporting of details results of one parameter
--

FVC	Category of percent of predicted	Unexposed numbers	Exposed numbers
Normal	80-100%	103	69
Mild abnormality	70-79%	4	12
Moderate abnormality	50-69%	1	2
Severe abnormality	<50%	0	1
Total		108	84

The following procedure of evaluating and reporting spirometry results is recommended

The loss of function, expressed in "percent of predicted," is taken as an indicator of difference between the two groups instead of the absolute numbers. That way, we can demonstrate early changes without waiting to see that the results fall below the 80 or 70% of predicted to indicate obvious disease. It increases the sensitivity of the study. The two groups need therefore not be identical in height, weight or age.

The distribution of the loss of function can be reported in detail (Table 3) or in more aggregated clinical groups (Table 4). The necessary descriptive statistics (e.g. risk ratio, attributable fraction), inferential statistical calculations for hypothesis testing (e.g. tests of significance) or interval estimation (point estimates and confidence intervals) can then be calculated using standard methods.

For further simplification, the prevalence of abnormal results alone in each cohort (exposed *vs* unexposed) can be reported and the risks and the risk ratios calculated using the standard methods (Table 4). This data reduction, however, reduces the resolution of the study and might delete important details. You cannot detect early changes and warning signs. Usually 80 or 70% cutoff is used to declare a result "not normal" depending on the specificity and sensitivity needed. Following the consensus guidelines and standard textbooks is recommended.^[16,17]

On the other hand, even this method is actually incorrect and only used in clinical practice because of convenience. The 80% cutoffs are actually not the deviation from the mean by two Standard Deviations (SD). It is a widely accepted definition in science that a result is "not normal" when it lies outside the 2SD range (outside the approximate 95% range) in a normally distributed curve. All laboratory results in medicine, for instance, are derived by this method. In spirometry, this lower limit of the 2SD as the limit below which we call a result "abnormal," is called the "Lower Limit of Normal" (LLN). A more detailed demonstration on how to calculate the LLN is found in the published literature (see Box 4).^[20]

When using spirometry for research purposes, taking the more exact approach of LLN should be seriously considered and discussed. If that approach is not taken, the reasons should be discussed in the research report.

Accounting for false calculations of spirometers

After dismissing the technically unacceptable tests, we usually trust the spirometers to give us the correct "best" values and respective "percent of predicted." But is this true? Spirometers often use one parameter (e.g. the FVC) as the most important parameter for ranking a series of tests for one individual. For calculating the "percent of predicted," they sometimes do not use *other* test results of the same patient, who showed better results in other categories (e.g. FEV1 or PEF). The machine uses only parameters from the "best test"

Table 4:	Results	grouped	in	disease	categories
----------	---------	---------	----	---------	------------

	Exposed group		Group unexposed		Risk ratio of exposed vs unexposed
Categories	Number of workers	Percent	Number of workers	Percent	
Mild restriction	4	4	14	15	4.29
Moderate restriction	1	1	6	6	7.35
Severe restriction	0	0	3	3	
Mild obstruction	4	4	6	6	1.84
Moderate obstruction	1	1	2	2	2.45

it has selected. If the spirometer uses this method to select the best test results, we then have to calculate *by hand* the correct "percent of predicted" of other parameters to get the correct "percent of predicted" of all best parameters.

This procedure is recommended when a patient has borderline abnormal test results and a medical intervention such as initiation of treatment or change of job depends on the correct interpretation of the spirometry. Sometimes previously "sick" persons become suddenly "healthy" when we look at all results and not only the machine-picked preselected one. The worker might just had been graded "sick" by the insufficient machine.

The American Thoracic Society discusses in their consensus paper that it would be ideal if all results could be taken from the one "best" test of an individual.^[15] But in practice, that is

Text boxes

Box 1: Example of an ideal situation to establish reference values

A sufficient number of persons to establish a normally distributed curve and come to a point estimate is found for this purpose. They are all of male sex, height of 173 cm, and the age of 25 years. Now we can calculate the point estimate of the s—called 'predicted value' for the group of persons. Now, the same procedure with the same sex, same height and same weight but with a different age (for instance 26 years) has to be repeated. Later, the height is changed leaving other parameters the same so on.

Box 2: Example for evaluation of spirometry results over time

A worker might start working with e.g. 102% of predicted in FEV1. He is exposed to allergens (e.g. organic dust). One year later his FEV1 is only 84% of predicted. He would be considered "normal" if the company doctor only looks at the percent of predicted. However, his drop of 18% compared to previous FEV1 results indicates a significant loss in his lung function. Severe asthma may develop. If the factory medical officer only looks at the print-out every single year the disease and the presence of allergens at the workplace might be overlooked. No prevention will be done in time to prevent further damage in the worker and his co-workers.

Box 3: Example for overlooking respiratory disorders

If the exposed subjects have, by chance, a lower mean age or a higher mean height than the workers used as controls, the results of the spirometry in absolute numbers (such as FVC measured in liters) might by chance be the same as the control group even though they have already an impairment. A significant effect of occupational exposure could be missed by this false analysis and reporting.

Box 4: Explanation of lower limit of normal range in spirometry

From the Spirometry Training Guide of the National Institute of Occupational Safety and Health NIOSH/USA $^{\rm [20]}$

"The lower limit of the normal range (LLN) is the threshold below which a value is considered abnormal - usually the value is set so that 95% of a "normal" population will have values above the LLN value and correspondingly, 5% of a "normal" population will have values below the LLN. The LLN is about 80% of the predicted value for FEV1 and for FVC, but about 90% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the FEV1/FVC ratio, and about 60% of the predicted value for the predicted using the reference equations. ... Although the comparisons with the LLN are preferred, results that are at least 80% of the predicted normal value are often considered within the normal range if no other data are available for comparison. ... The specific value of the LLN is dependent on the study population and methods used to derive the reference values. LLN's should be available from the reference value source."

often not possible, and we have to look at several tests to find the best results with the maximum effort of the patient.

CONCLUSIONS

In general practice, occupational health practice and in medical research, a critical look at the quality of spirometry results is necessary.

Correct reference equations, good patient coaching, decision on the technical quality (acceptability) of each spirometry test, and critical re-evaluation of the machine's read-out are essential parts of a good spirometry test.

Every spirometry result has to be thoroughly cross-checked and sometimes re-calculated or corrected by a competent person who is actually better than the machine. Otherwise we will rely on either technically insufficient or false test reports from the machine.

A good understanding how results are calculated is crucial for further statistical evaluation and correct reporting, especially if aggregated results of groups are described.

ACKNOWLEDGMENT

The discussions with our colleagues in the Department of Environmental Health Engineeringhelped to develop the topic. The article was supported in part by a grant from NIEHS, USAP30 ES01247.

REFERENCES

- See Government of India, The Factories Act, 1948, section 41C and Tamil Nadu Factories Rules, section 95 schedules IV, V, VIII, XIX, XX, XXII, XXXII.
- Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. American Journal of Respiratory and Critical Care Medicine 1999; 159:179-87.
- Patel RK, Bhagat GR, Kaji BC, Dalsania VD, Thanvi SS, Patel DV. Study of Pulmonary Function. Tests in 2000 Healthy Persons in Gujarat. J Assoc Physicians India 1998;46:689-94.
- Kamat SR, Sharma BS, Raju VR, Venkatraman C, Balakrishna M, et al. Indian Norms for Pulmonary Function. J Assoc Physicians India 1997;25:531-40.
- Cohen Ashok Fulambarker, Ahmet Sinan Copur, Asavari Javeri, Sujata Jere and Mark E. Cohen. Reference Values for Pulmonary Function in Asian Indians Living in the United States. Chest 2004;126;1225-33.
- Chatterjee S, Nag SK, Dey SK. Spirometric standards for non-smokers and smokers of India (eastern region). Jpn J Physiol 1988;38:283-98.
- Vijayan VK, Kuppurao KV, Venkatesan P, Sankaran K, Prabhakar R. Pulmonary function in healthy adult Indians in Madras. Thorax 1990;45:611-5.
- Vijayan VK, Sankaran K, Venkatesan P, Kuppurao KV. Prediction equations for maximal voluntary ventilation in non-smoking normal subjects in Madras. Indian J Physiol Pharmacol 1993;37:138-40.
- 9. Vijayan VK, Kuppurao KV, Venkatesan P, Sankaran K. Reference values and prediction equations for maximal expiratory flow rates in

non-smoking normal subjects in Madras. Indian J Physiol Pharmacol 1993;37:291-7.

- Behera D. Normal spirometry values. *In*: Shankar PS, editor. Pulmonary function tests in health and disease. The National Book Depot, Mumbai 2003.
- Fulambarker A, Copur AS, Javeri A, Jere S, Cohen ME. Reference values for pulmonary function in Asian Indians living in the United States. Chest 2004;126:1225-33.
- ATS Workshop on Lung Volume Measurements Official Statement of the European Respiratory Society. Reference Values for Residual Volume, Functional Residual Capacity and Total Lung Capacity. Eur Respir J 1995;8:492-506
- 13. RMS Medspiror, Recorders & Medicare Systems, Chandigarh, India, www.rmsindia.com
- 14. MIR Spirobank, MIR Medical International Research, Italy, www.spirometry.com
- American Thoracic Society, Medical Section of the American Lung Association. Standardization of spirometry - 1994 update. Am J Resp Crit Care Med 1995;152:1107-36
- American Thoracic Society, Medical Section of the American Lung Association. Lung Function Testing: Selection of Reference Values and Interpretative Strategies. Am Rev Respir Dis 1991;144:1202-18.
- 17. Working Party on Standardization of Lung Function Tests. European

Community for Steel and Coal. Standardized Lung Function Testing. Official Statement of the European Respiratory Society. Eur Resp J 1993;6.

- Quanjer PH, Lebowitz MD, Gregg I, Miller MR, Pedersen OF. Peak expiratory flow: conclusions and recommendations of a Working Party of the European Respiratory Society. Eur Respir J 1997; 10:2s–8s
- American Thoracic Society (ATS). Pulmonary Function Laboratory Management and Procedures Manual. Accessed 15. Oct 2005 at http://www.thoracic.org/statements
- Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. NIOSH Spirometry Training Guide. Morgantown, 2003. Accessed on 9. Feb 2005 at http://www.cdc.gov/ niosh/docs/2004-154c/pdfs/2004-154c.pdf
- American College of Occupational and Environmental Medicine. Evidence Based Statement on Spirometry in the Occupational Setting. Accessed on 17. Oct.05 at http://www.acoem.org under http:// www.acoem.org/guidelines/evidence
- Wang ML, Petsonk EL. Repeated measures of FEV1 over six to twelve months: what change is abnormal? J Occup Environ Med 2004;46:591-5.
- Hankinson JL, Wagner GR. Medical screening using periodic spirometry for detection of chronic lung disease. *In*: Occupational Medicine: State of the Art Reviews. Philadelphia, Pa: Hanley & Belfus. Cited from [20]; 1993. p. 118.