

Standardization of Spirometry in Assessment of Responders Following Man-made Disasters: World Trade Center Worker and Volunteer Medical Screening Program

Paul Enright, MD,¹ Gwen Skloot, MD,² and Robin Herbert, MD³

¹The University of Arizona, Tucson, AZ, USA

²The Division of Pulmonary and Critical Care Medicine and

³Department of Community and Preventive Medicine, Mount Sinai School of Medicine, New York, NY, USA

ABSTRACT

Background: Spirometry is the most commonly used pulmonary function test to screen individuals for suspected lung disease. It is also used for screening workers with exposures to agents associated with pulmonary diseases. Although the American Thoracic Society (ATS) provides guidelines for spirometers and spirometry techniques, many factors are not standardized, so that results from individual pulmonary function laboratories vary substantially. These differences can create substantial difficulties in using data pooled from multiple sites to understand health consequences of disasters that involve exposures to pulmonary toxins. This article describes the approach used to minimize these differences for a consortium of institutions who are providing medical monitoring examinations to World Trade Center (WTC) responders. The protocol improved upon the minimal ATS guidelines.

Methods: Spirometric measurements were obtained before and after use of a bronchodilator. A fourth-generation spirometer was chosen that exceeded ATS spirometer accuracy standards. The accuracy was verified at the beginning of each day of testing.

Technologists who performed the spirometry tests were centrally trained and certified and received regular reports on their performance. Reference values and normal ranges were obtained from the National Health and Nutrition Examination Survey (NHANES III) data set. A standardized interpretation flowchart was followed to reduce misclassification rates for airway obstruction and restriction. Patients with spirometric abnormalities were referred for more extensive diagnostic testing.

Results: More than 12,000 spirometry tests were performed during the first examination. The 20 spirometers used at the 6 participating institutions maintained accuracy within 3% for more than 4 years. Overall, more than 80% of the test sessions met ATS quality goals. Spirometry abnormality rates exceeded those obtained for adults who participated in the NHANES III survey.

Conclusions: The program allowed standardization of the performance and interpretation of spirometry results across multiple institutions. This facilitated reliable and rapid diagnosis of lung disease in the large number of WTC responders screened. We recommend this approach for postdisaster pulmonary evaluations in other settings. *Mt Sinai J Med* 75:109–114, 2008. © 2008 Mount Sinai School of Medicine

Key Words: spirometry, quality control, WTC, occupational screening, disaster.

Address Correspondence to:

Robin Herbert

Department of Community and Preventive Medicine

Mount Sinai School of Medicine
New York, NY

Email: robin.herbert@mssm.edu

BACKGROUND

Pulmonary specialists and respiratory therapists (RTs) rushed to Ground Zero within weeks following

September 11th, 2001, and realized that workers who were inhaling smoke, fumes, and dust would likely suffer symptoms and, potentially, loss of lung function from resulting airway inflammation. They brought portable spirometers to measure lung function. During October 2001, these therapists performed 5-minute spirometry tests for more than 1000 workers in a union hall and gave them the reports with dozens of numbers and a computerized interpretation.¹ About half of the reports indicated a lung function abnormality, consistent with restriction of lung volume or airway obstruction.

In November, copies of the 1000 spirometry reports were scanned and sent for review to a pulmonary function expert at 1 of the top 5 medical institutions in the United States. He viewed all of the flow-volume curves and computer interpretations and was disturbed by the overall poor quality of the tests and the very high misclassification rate. A second opinion from Paul Enright agreed. About half of those who were told that they had restriction (a low vital capacity) actually did not (a false positive). On the other hand, some workers with a chronic cough were told that their spirometry result, were normal (according to the computerized interpretation) but that their flow-volume curves showed airway obstruction (a false negative).

During 2002, the National Institute of Occupational Safety and Health (NIOSH) recommended a spirometry Quality Assurance (QA) program for the World Trade Center (WTC) Worker and Volunteer Medical Screening Program. This program addressed all factors that can cause errors in spirometry tested results, such as instruments, technologists, reference equations, interpretation scheme, and medical decisions based on test results (referrals, diagnoses, and treatment plans). This article describes these factors and the standardization used to minimize misclassification rates. Spirometry results from the first examination have been published.²

METHODS

Spirometry Maneuver

Spirometry has been used by doctors for more than 150 years.³ The maneuver is fairly simple to describe and involves 3 separate phases. The first phase is a complete, deep inhalation, the second is a forceful expiration, and the third is maintenance of exhalation effort for at least 6 seconds. Multiple, precise numbers describing instantaneous flows and relative volumes are generated from this maneuver along with a graphic display of all data from the 3 phases.

Unfortunately, despite elaborate presentation, in practice, spirometry is often fraught with error.⁴⁻⁶

Instruments

Current models of spirometers measure airflow and convert it to exhaled volume (in liters). This allows them to be relatively portable when compared with the older models of volume spirometers, which had to contain up to 10 L of exhaled air and, thus, were rather large.⁷ Unfortunately, the compact size of modern spirometers was often at the expense of long-term accuracy. Condensation of moisture from exhaled breath, as well as phlegm coughed into the flow sensors, clogged many models, and falsely increased the measured flows (forced expiratory volume in 1 second or FEV₁) and volumes (forced vital capacity or FVC).⁸ Falsely high FVCs were reported by the automated spirometer used in October 2001 for dozens of workers whose fingers partially occluded the flow sensors during the breathing maneuvers. This was not recognized by those who tested Ground Zero workers at that time.

We carefully chose a modern hand-held spirometer with an ultrasonic flow sensor that does not become inaccurate when moisture or phlegm is exhaled into the mouthpiece (EasyOne Diagnostic model, ndd, Switzerland). The mouthpiece is simply an open tube.⁹ The American Thoracic Society (ATS) has guidelines for spirometers and spirometry methods,¹⁰ including daily calibration checks to ensure accuracy within 3%. Therefore, the accuracy with which spirometers measured volumes was checked every day by using a 3-L calibration syringe. The 20 spirometers purchased for this project have maintained better than 3% accuracy for the first 4 years, similar to findings of a large multicenter study in Latin America that also used the same spirometer model.¹¹

Technologists

The technologist who administers the spirometry test is the most important determinant of test quality.¹² Patients remain passive during most medical tests, but the 3 phases of the spirometry test require athletic-type breathing maneuvers. The technologist must be an enthusiastic “cheerleader” to coax the patient for maximal effort during all 3 phases. Poor coaching causes inaccurate results.

For spirometry performed in occupational clinics, NIOSH certifies 16-hour spirometry courses that train occupational nurses and spirometry technologists.¹³ However, there is no review of the

quality of spirometry tests performed by a technologist subsequent to the spirometry course. Technologists may choose to take an examination to become a certified pulmonary function technologist (CPFT).¹⁴ However, less than half of the technologists working in hospital-based pulmonary function laboratories in the United States have this credential, because it is not required by most employers. The CPFT exam is not based on actual performance but only on knowledge of spirometry and arterial blood gas tests. Therefore, in our experience, only a minority of respiratory technologists are able to perform good quality spirometry tests. This requires both training and experience.

Technologists for the current program were trained at Mount Sinai Hospital at the inception of this project, and they followed a written manual of procedures. In addition, the spirometer checked the quality of each maneuver and graded the quality of the spirometry test session in real time.¹⁵ However, the technologist could ignore the messages displayed on the spirometer. The quality grade (A–F) was printed on every report for the physician to consider. Each month, Gwen Skloot, the Director of the WTC Pulmonary Function Quality Assurance Program reviewed the quality of all spirometry tests performed by each technologist and reported their success rate in meeting ATS goals for test session quality (equivalent to a quality grade of A or B), as in the multicenter Lung Health Study.¹⁶ Such feedback has been demonstrated to improve spirometry quality.¹⁷ WTC spirometry technologists achieved high success rates throughout this project.

Reference Equations

For 3 decades, a wide range of spirometry reference equations has been available. Predicted values for FEV₁, FVC, and FEV₁/FVC vary according to the reference equation chosen by the spirometer manufacturer as the default or changed by the user according to preference. Choices include Crapo from Salt Lake City, Knudson from Tucson, Morris from Portland Oregon, Miller from Michigan, and many others.¹⁸ None of these investigations included African Americans, Hispanics, or Asian Americans, so that appropriate reference values for these groups were not available. The ATS did not recommend a single source of reference equations for use in the United States until late in 2005,¹⁰ and few pulmonary function test (PFT) labs have adopted them. As a result of variability in reference equations employed, a patient with borderline abnormal lung function could have an interpretation of normal lung function in 1 laboratory, mild restriction in another, and mild airway obstruction in a third laboratory.

From the beginning of the WTC Worker and Volunteer Screening Program, spirometry reference equations from the third National Health and Nutrition Examination Survey (NHANES III) have been used for all sites, including the FDNY. Separate equations for African Americans, Hispanics, and Caucasians are provided, along with the lower limit of the normal range (LLN) for each key spirometry variable, on the basis of the fifth percentile of the healthy subgroup from NHANES III.¹⁹ Although the NHANES III spirometry reference equations were recommended by the ATS PFT standards committee in 2005, this choice was based on opinions of experts on the committee. No comparison was made of differences in the prevalence rates of spirometry abnormalities when using the popular reference equations. These differences are small when testing typical patients evaluated in a hospital-based PFT lab, because most of these individuals have moderate to severe lung disease and are far from the LLN. However, abnormality rates for mild restriction and mild airway obstruction can differ widely when samples of the general population, or blue-collar working groups such as the WTC responders, are compared by using the major reference equations.²⁰

RESULTS AND DISCUSSION

Interpretation Scheme

After comparison with a specific set of reference equations, a list of results displayed as a percentage of the predicted value is produced for key spirometry variables FEV₁, FVC, and FEV₁/FVC. However, reference studies do not suggest how various patterns of percentage of predicted results should be interpreted, and several schemes have been published.²¹ The physician interpreting results is free to choose any scheme. Therefore, the same set of percentage of predicted results from the same PFT laboratory could inadvertently be interpreted differently depending on which pulmonary specialist interpreted results on that day.²²

Because inhaled medication for asthma and chronic obstructive pulmonary disease (COPD) is a highly profitable market, there has been financial incentive during the past decade to push the threshold of spirometry abnormality from normal to abnormal and from a pattern of mild restriction (a slightly reduced vital capacity, which is often due to obesity) to a pattern of mild airway obstruction.²³ The industry-sponsored 2001 GOLD guidelines for COPD,²⁴ for example, consider an FEV₁/FVC <0.70 to indicate airway obstruction, even if the FEV₁ is

more than 100% predicted. If the pattern persists after an inhaled bronchodilator, it is considered due to COPD. However, this guideline ignores studies that show that the FEV₁/FVC declines with normal aging;²⁵ thus, airway obstruction is being underreported in young workers, while the false-positive rate for “COPD” increases dramatically in people older than age 55 years.²⁶

The WTC Screening Program used a standardized interpretation method used by CDC NIOSH, which (considering the high rate of respiratory symptoms of WTC responders) assumes a modest increase in pretest probability of airway obstruction. For mild airway obstruction to be interpreted, the FEV₁/FVC must be below the age- and sex-corrected LLN (the fifth percentile, which is about 90% of the predicted value) and the FEV₁ must also be below its LLN (about 80% of predicted). When the FEV₁/FVC is below the LLN but the FEV₁ is not, then an interpretation of borderline airway obstruction is given. The FEF 25%–75% is not used to interpret “small airway disease” when the FEV₁/FVC is normal, because this concept has never been validated.¹⁰

When a patient has airway obstruction, the presence of a low FVC is often called a “mixed pattern” suggesting a superimposed restriction of lung volumes, but about 90% of the time, the total lung capacity (TLC, as measured in a body plethysmograph) is normal, ruling out a restrictive abnormality.²⁷ The low FVC in such cases is due to early airway closure at the end of forced expiration with resultant air trapping and hyperinflation. The spirometry interpretation algorithm used by the WTC Screening Program does not make the mistake of interpreting this as a “mixed pattern,” but instead designates it as a low FVC.

Medical Decisions

Results of medical tests, even when very accurate and reproducible, do not, on their own, establish a diagnosis. Physicians also consider the patient's symptoms, risk factors (such as smoking, family history, and workplace exposures), medication responses, and consequences of falsely positive versus falsely negative diagnoses²⁸ (i.e., overdiagnosis versus delayed diagnosis, respectively).

A common flaw in medical decisions when evaluating a patient with a history of asthma-like symptoms but with normal spirometry results (lack of airway obstruction or lack of a bronchodilator response) is to rule out asthma. Mild asthma, by definition, is intermittent, and spirometry will usually be normal between exacerbations.²⁹ If the diagnosis is needed quickly in such patients, a methacholine

challenge test³⁰ or an exhaled nitric oxide test³¹ can be performed.

On the other hand, another common mistake when evaluating spirometry results in adult smokers (with or without respiratory symptoms) is to attribute mild airway obstruction to COPD. The patient should be given an inhaled bronchodilator. Ten minutes later (post-bronchodilator administration), if the airway obstruction is no longer present, COPD is ruled out, and asthma should be considered.³² If the patient still has airway obstruction, a diffusing capacity (DLCO) test should be ordered. A low DLCO (corrected for anemia, if present) in an adult smoker with moderate to severe airway obstruction post-bronchodilator administration, is highly likely to be due to COPD.²¹ However, other causes for fixed airway obstruction are possible, such as cystic fibrosis in young adults and bronchiolitis obliterans in workers exposed to some types of fumes.³³

In the absence of airway obstruction, a mildly low FVC (below the LLN) is often called “a restrictive impairment,” but the clinical correlates of this pattern are not described in medical literature. About half of patients referred to a PFT lab with this pattern do not have restriction of lung volumes when body plethysmography is performed;^{34,35} thus, this pattern is better described as a “nonspecific pattern.” If the FVC is normal after bronchodilator (or the slow inspiratory vital capacity is normal), then a restrictive impairment is quickly ruled out. If the FVC is moderately to severely low (below 60% predicted) or the patient has symptoms suggesting an interstitial lung disease, such as dyspnea on exertion or a chronic cough, then a DLCO test and chest x-ray (or perhaps a high-resolution lung CT scan) should be ordered.

Challenges

This program was established at a time when there was concern about potential health consequences due to inhalational exposures to a poorly characterized and unprecedented mix of aerosolized chemicals, combustion products, and micronized building materials. Challenges in assessment of the impact of the disaster response effort on respiratory health at the time of program inception included:

1. Limited information on the nature of exposures,
2. Inability to predict the most likely outcomes (given the limited toxicological information),
3. Absence of baseline (pre-exposure) pulmonary function testing and absence of a control population,
4. Provision to individual responders of meaningful clinical evaluations while also obtaining data

to permit population-based study of respiratory outcomes following an unprecedented disaster,

5. Rapid establishment of a multicenter program, staffed primarily by occupational and internal medicine physicians,
6. Rapid creation of standardized protocols and procedures, with clear guidelines about PFT performance and interpretation for health care providers with variable knowledge about performance and interpretation of PFTs,
7. Emphasis on the importance of using standardized equipment and quality assurance (QA) approaches with which no provider had prior familiarity,
8. Interpretation schema (LLN versus the traditional predicted percentage) that was just being introduced in the pulmonary world and with which most occupational medicine providers and internists were unfamiliar,
9. Logistical, administrative, staffing and other operational challenges faced by each clinical center in their efforts to rapidly expand clinical capacity to accommodate a complex and multifaceted clinical operation.

Thus, it was critical to include pulmonologists who had expertise in standardization and QA in PFT performance/interpretation, as well as a dedication to regularly communicate the importance of standardization and QA to healthcare professionals who were striving to establish and provide programs that were patient friendly, while simultaneously addressing a broad spectrum of possible physical health and mental health outcomes.

Despite many challenges, the WTC Screening Program was successful in incorporating a rigorous PFT QA program into a rapidly developed screening program for disaster responders. To minimize the risk of high misclassification rates and inconsistent interpretations present in other settings, this program standardized all steps in spirometry testing and interpretation and maintained a proactive QA program. This facilitated reliable and rapid diagnosis of lung disease in the 12,000 WTC responders who were screened.

SUMMARY

We have demonstrated the feasibility of incorporating a rigorous program of standardized PFT performance, interpretation, and QA into a setting of postdisaster screening, where, by definition, extensive planning is not possible. This model has two major strengths: (1) it has proven to be logistically feasible for use in the difficult period of postdisaster health response,

and (2) it has yielded high-quality PFTs for use in both individual clinical care and epidemiologic study of disaster-related health effects.

ACKNOWLEDGMENTS

This work was supported by the Centers for Disease Control and Prevention and the National Institute for Occupational Safety and Health, contract 200-2002-0038 and grant 5U1O-0H008232. This information has not been previously presented or published.

REFERENCES

1. QRS Spirometer. <http://www.qrssystem.com/233462.html> Accessed September 25, 2007.
2. Herbert R, Moline J, Skloot G, et al. The World Trade Center disaster and the health of workers: five-year assessment of a unique medical screening program. *Environ Health Perspect* 2006; 114: 1853–1858.
3. Hutchinson J. On the capacity of the lungs and on the respiratory functions, with a view to establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Tran* 1846; 29: 137–252.
4. Hankinson JL. Pulmonary function testing in the screening of workers: guidelines for instrumentation, performance, and interpretation. *J Occup Med* 1986; 28: 1081–1092.
5. Kunzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV1 due to technician, team, device, and subject in an 8 center study. *Eur Respir J* 1995; 8: 371–376.
6. Ng'ang'a LW, Ernst P, Jaakkola MS, et al. Spirometric lung function: distribution and determinants of test failure in a young adult population. *Am Rev Respir Dis* 1992; 145: 48–52.
7. Hankinson JL. State of the art of spirometric instrumentation. *Chest* 1990; 97: 258–259.
8. Townsend JC, Hankinson JL, Lindesmith LA, et al. Is my lung function really that good? Flow-type spirometer problems that elevate test results. *Chest* 2004; 125: 1902–1909.
9. ndd Easy One spirometer specifications. http://www.ndd.ch/English/Products/EasyOne_fs.html Accessed August 17, 2007.
10. Miller MR, Hankinson J, Brusasco V, et al. Standardization of spirometry. *Eur Respir J* 2005; 26: 319–338.
11. Perez-Padilla R, Vazquez-Garcia JC, Marquez MN, et al. The long-term stability of portable spirometers used in a multinational study of the prevalence of chronic obstructive pulmonary disease. *Respir Care* 2006; 51: 1167–1671.
12. Enright PL. How to make sure your spirometry tests are of good quality. *Respir Care* 2003; 48: 773–776.
13. NIOSH/CDC. What is the NIOSH-approved spirometry training course? <http://www.cdc.gov/niosh/topics/spirometry/> Accessed August 17, 2007.
14. NBRC. The Entry Level CPFT Examination given by the National Board of Respiratory

- Care. <http://www.nbrc.org/ExamsCPFT.htm> Accessed August 17, 2007.
15. Banks DE, Wang ML, McCabe L, et al. Improvement in lung function measurements using a flow spirometer that emphasizes computer assessment of test quality. *J Occup Environ Med* 1996; 38: 279–283.
 16. Enright PL, Johnson LR, Connett JE, et al. Spirometry in the Lung Health Study. 1. Methods and quality control. *Am Rev Respir Dis* 1991; 143: 1215–1223.
 17. Malmstrom K, Peszek I, Botto A, et al. Quality assurance of asthma clinical trials. *Control Clin Trials* 2002; 23: 143–156.
 18. American Thoracic Society. Lung function testing: selection of reference values and interpretive strategies. *Am Rev Respir Dis* 1991; 144: 1202–1218.
 19. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
 20. Celli BR, Halbert RJ, Isonaka S, Schau B. Population impact of different definitions of airway obstruction. *Eur Respir J* 2003; 22: 268–273.
 21. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
 22. Raghunath AS, Innes A, Norfolk L, et al. Difficulties in the interpretation of lung function tests in the diagnosis of asthma and chronic obstructive pulmonary disease. *J Asthma* 2006; 43: 657–660.
 23. Enright P. Flawed interpretative strategies for lung function tests harm patients. *Eur Respir J* 2006; 27: 1322–1323.
 24. Pauwels RA, Buist AS, Calverley PMA. Global strategy for the diagnosis, management, and prevention of COPD. *Am J Respir Crit Care Med* 2001; 163: 1256–1276.
 25. Quanjer PH, The Pulmonaria Group. Summary equations for spirometry LLN values, based on quality reviews. www.spirxpert.com/GOLD.html accessed April 11, 2007.
 26. Hansen JE, Sun X-G, Wasserman K. Spirometric criteria for airway obstruction. Use percentage of FEV₁/FVC ratio below the fifth percentile, not <70. *Chest* 2007; 131: 349–355.
 27. Dykstra BJ, Scanlon PD, Kester MM, et al. Lung volumes in 4774 patients with obstructive lung disease. *Chest* 1999; 115: 68–74.
 28. Sox HC. Decision-making: a comparison of referral practice and primary care. *J Fam Pract* 1996; 42(2): 155–160.
 29. Ulrik CS, Postma DS, Backer V. Recognition of asthma in adolescents and young adults: which objective measure is best? *J Asthma* 2005; 42(7): 549–554.
 30. Crapo RO, Casaburi R, Coates AL, Enright PL, For The American Thoracic Society. Guidelines for methacholine and exercise challenge testing-1999. *Am J Respir Crit Care Med* 2000; 161: 309–329.
 31. Taylor DR, Pijnenburg MW, Smith AD, De Jongste JC. Exhaled nitric oxide measurements: clinical application and interpretation. *Thorax* 2006; 61: 817–827.
 32. Perez-Padilla R, Hallal PC, Vazquez-Garcia JC, et al. Impact of bronchodilator use on the prevalence of COPD in population-based samples. *COPD* 2007; 4(2): 113–120.
 33. Kreiss K, Gomaa A, Kullman G, et al. Clinical bronchiolitis obliterans in workers at a microwave popcorn production plant. *N Engl J Med* 2002; 347: 330–338.
 34. Glady CA, Aaron SD, Lunau M, et al. A spirometry-based algorithm to direct lung function in the pulmonary laboratory. *Chest* 2003; 123: 1939–1946.
 35. Aaron SD, Dales RE, Cardinal P. How accurate is spirometry at predicting restrictive pulmonary impairment? *Chest* 1999; 115: 869–873.