Is My Lung Function Really That Good?*

Flow-Type Spirometer Problems That Elevate Test Results

Mary C. Townsend, DrPH; John L. Hankinson, PhD; Larry A. Lindesmith, MD, FCCP; William A. Slivka, RPFT; Gregg Stiver, RRT, CPFT; and Gerald T. Ayres, RPFT

Most spirometry errors reduce test results, and it is widely assumed that measurement accuracy is guaranteed by frequent spirometer calibrations or calibration checks. However, zero errors and changes in flow-type spirometer sensors may occur during testing that significantly elevate test results, even though the spirometer was calibrated recently. To draw attention to these often-unrecognized problems, this report presents anomalous spirograms and test results obtained from occupational medicine clinics and hospital pulmonary function laboratories during quality assurance spirogram reviews. The spurious results appear to have been caused by inaccurate zeroing of the flow sensor, or by condensation, mucus deposition, or unstable calibration of various flow-type spirometers. These errors elevated some FVCs to 144 to 204% of predicted and probably caused 40% of 121 middle-aged working men in respirator medical clearance programs to record both FVC and FEV_1 > 120% of predicted. Since spirometers report the largest values from a test, these errors must be recognized and deleted to avoid false-negative interpretations. Flow-type spirometer users at all levels, from the technician to the interpreter of test results, should be aware of the potential for and the appearance of these errors in spirograms.

(CHEST 2004; 125:1902–1909)

Key words: diagnostic errors; forced expiratory flows; forced expiratory volume; maximal expiratory flow-volume curves; peak expiratory flow; quality assurance; respiratory function tests; spirometry

Abbreviations: ATS = American Thoracic Society; PEF = peak expiratory flow; QA = quality assurance

Since 1978–1979, the Occupational Safety and Health Administration and the American Thoracic Society (ATS) have required/recommended that the largest FEV_1 and the largest FVC be chosen to summarize a spirometry test, and that flows be measured from the curve with the largest sum of FEV_1 plus FVC.[1–4] This algorithm was selected because most testing errors[4,5] and many spirometer malfunctions[6] cause results to be underrecorded. Since small erroneous curves will be replaced by larger, more accurate curves in a test session, erroneously reduced values will not affect the reported test results.

However, reporting the largest values makes spirometry results highly vulnerable to errors that elevate test results. If erroneously large curves are not recognized and deleted, the inflated values will be saved and reported in the test summary, replacing accurate but lower values from the test. This problem is compounded by the following limitations of many currently available flow-type spirometers: (1) lack of a large real-time display meeting ATS recommendations[4] to help technicians detect errors during testing; (2) saving only the largest three curves from a test; and (3) having no mechanism for editing curves recorded earlier in the test session (to delete an earlier erroneous curve, the entire test session must be deleted).

Examples

Spirogram Reviews

Spirograms were submitted for quality assurance (QA) review on a consulting basis or by National

*From M.C. Townsend Associates (Dr. Townsend), Pittsburgh, PA; Hankinson Consulting (Dr. Hankinson), Valdosta, GA; Department of Internal Medicine (Dr. Lindesmith), Gundersen Lutheran Medical Center, La Crosse, WI; the Department of Medicine (Mr. Slivka), University of Pittsburgh School of Medicine, Pittsburgh, PA; Latrobe Area Hospital (Mr. Stiver), Latrobe, PA; and University of Pittsburgh Medical Center (Mr. Ayres), Pittsburgh, PA.

This work was performed at the office of M.C. Townsend Associates, Pittsburgh, PA.

Manuscript received January 28, 2003; revision accepted October 20, 2003.

Correspondence to: Mary C. Townsend, DrPH, M.C. Townsend Associates, 289 Park Entrance Dr, Pittsburgh, PA 15228-1824; e-mail: mary.townsend4@verizon.net

Institute for Occupational Safety and Health-approved spirometry course participants from industrial medical departments, occupational medicine clinics, and hospital pulmonary function laboratories. The reviews found elevated test results and anomalous graph shapes recorded by several different flow-type spirometers. All problem graphs were recorded on calibrated spirometers, indicating that changes in the sensor or calibration occurred after calibration, during subject testing. These real-life problem graphs illustrate two errors that can occur in many different flow-type spirometers: (1) problems in zeroing the sensor before a maneuver, and (2) problems caused by condensation, mucus, or fingers blocking the sensor during a test. Though these errors have been discussed occasionally,[7–10] they have not been illustrated previously, allowing spirometry users to see the patterns to be recognized and deleted. As presented in the figures below, these errors can have profound effects that are obvious, or the effects may be subtle, and more difficult to detect.

Zero Errors

Flow-type spirometers measure airflow indirectly and then integrate the measured flows over time to obtain volumes. Many flow-type spirometers have pressure sensors, which measure the pressure or pressure drop across a resistance element as air moves through that element, *eg*, screen, Fleisch, and Pitot tube pneumotachometers. The measured pressure gradient is directly proportional to the air flow through the sensor. Other flow-type spirometers measure the electrical current that must be applied to a hot wire to prevent it from cooling as air flows across it (mass flow spirometer), the rotation of a turbine as air moves across it (turbine spirometer), or the transit time of an ultrasound beam as air passes through a sensor (ultrasonic spirometer).[7–12]

Before measuring expired air, most flow-type spirometers first measure the pressure gradient (or other signal) that exists when no air is passing through the sensor. Spirometers may be "zeroed" before each expiration, or only once before a test session. During zeroing, the pressure drop (or other signal) corresponding to no airflow is established, setting the intercept of the calibration curve relating flow to pressure drop (or other signal). Errors occur when the transducer or electronics falsely measures a pressure gradient (or other signal) when, in fact, no air is passing through the sensor. The calibration curve is shifted so that airflow is falsely "detected" during the subject test when the measured pressure drop is zero, and all measured flows and volumes are falsely elevated.[7,8] Though zero errors also occur in which no pressure gradient (or other signal) is measured while air is, in fact, passing through the sensor, causing reduced flows and volumes to be measured, this error is beyond the scope of this report.

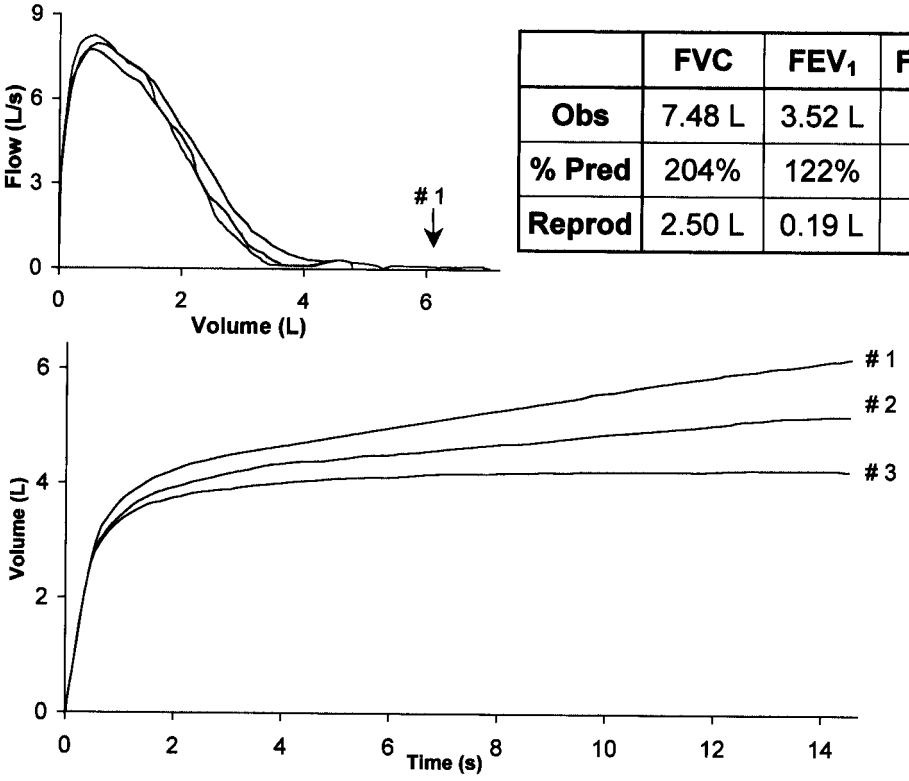
The spirograms in Figures 1, 2 illustrate zero errors of varying degrees. The spirogram in Figure 1 was produced by zeroing the spirometer before each expiration: curves 1 and 2 show varying zero errors, while the zeroing was apparently accurate in curve 3. Figure 2 presents a more extreme example of a zero error.

Though zero errors inflate all flows throughout the maneuver, the errors are best seen at the end of the maneuver, at what should be the FVC plateau, when expired air no longer enters the sensor. At that point, the falsely "detected" airflow of the zero error causes the volume-time curve to climb gradually as a straight line, as seen in curves 1 and 2 in Figure 1. This pattern is subtly different from airways obstruction, since the volume-time curves climb at a constant low rate, seen in the tail of very low flow in the flow-volume curve. (Though this spirometer stopped plotting the volume-time curve at 15 s, the flow-volume curve was recorded until the end of expiration.) In Figure 2, the zero error is larger, causing the volume-time curve to climb more steeply than in Figure 1, increasing at a rate of 0.25 L/s as seen in the flow-volume curve. Because the expiration was stopped after 6 s, the FVC in Figure 2 has reached only 145% of predicted, compared with 204% of predicted in Figure 1. Note that a hesitating start, which is a common technical error, is also apparent in Figure 2.

Finally, it is important to note that if the programming of the spirometer prevented editing of previously recorded curves, curves 1 and 2 in Figure 1 could not be deleted after curve 3 was recorded, even if curves 1 and 2 were recognized as flawed. Since the values of curve 1 were the largest recorded, those values were selected for the spirometry summary report. In addition, without a visual display, it would be difficult to recognize this error and delete the curves. In fact, even with an acceptable real-time display and an FVC of 204% of predicted, the spirometry test in Figure 1 was reviewed and accepted by the reviewing physician. A real-time display is only helpful when the viewer is aware of the potential for and the appearance of the zero-error problem.

Obstruction of the Sensor

Flow-type spirometers may use a different sensor for each subject or have one fixed sensor with



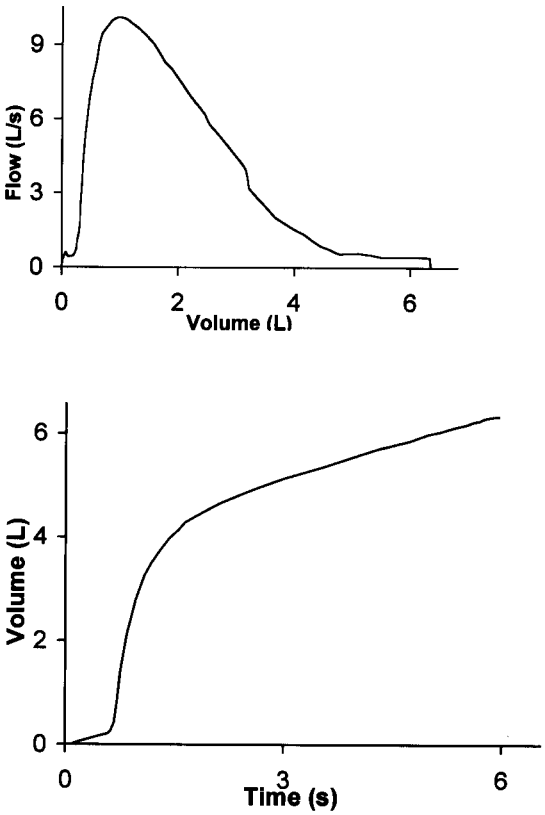
	FVC	FEV_1	FEV_1,%
Obs	7.48 L	3.52 L	47
% Pred	204%	122%	60%
Reprod	2.50 L	0.19 L	--

FIGURE 1. Zero errors. The zero error is largest for curve 1, smaller for curve 2, and the sensor was apparently correctly zeroed for curve 3. The inaccurately high values from curve 1 are reported in the summary, and the zero error causes the test reproducibility to be completely unacceptable. Note that the flow-volume curve was recorded until the end of the maneuver, though the volume-time curve was terminated at 15 s. Obs = observed; % Pred = percentage of predicted; Reprod = reproducibility.

replacement of filters in front of the sensor for each subject. In either case, sensors may become partially blocked during one subject's repeated blows into one sensor or when a series of subjects exhales into a spirometer with a fixed sensor. Exhaled water vapor can condense onto the sensor, mucus can deposit, or a subject's fingers can partially block the airflow through the sensor. In the case of a pressure-differential pneumotachometer, this blockage causes larger pressure drops to be measured across the resistance element, which are equated to increased flows and integrated to give elevated volumes.

Figure 3 shows an extreme example of probable contamination of the sensor by condensation, mucus deposition, or fingers partially obstructing the sensor. The FVC in curve 8 is 2.29 L larger than the FVC in curve 7, suggesting that the sensor was grossly contaminated by condensation or mucus deposition after the seventh maneuver. (The lead author has produced similar patterns experimentally, by slightly dampening the resistance element with water drops.) When the resistance element is partially blocked, the pressure drops measured across the resistance element throughout the expiration are greatly increased. These large pressure drops correspond to high flows, which when integrated, yield the large volumes seen in Figure 3. The peak expiratory flow (PEF) is 15.9 L/s, which exceeds the ATS instrumentation limit of 14 L/s for accurate flow measurement.[4] This very large PEF causes the flow-volume curve to exceed the scale, a characteristic that is not uncommon with moisture and sensor-blockage problems. However, a smaller person could have an inflated PEF that would not exceed the scale, making the problem more difficult to recognize. Finally, the PEF, FVC, and FEV_1 are all 130 to 170% of predicted, which is the range that we have seen for many blocked-sensor problems.

The spirogram in Figure 4 was known to have been affected by water vapor condensation in the sensor. The FVC and FEV_1 reproducibility of 0.60 L is unacceptable,[4] but more reasonable than that seen in Figure 3, and the percentage of predicted values are high but less extreme. Some athletes may have accurate percentage of predicted values that are this large. The error in Figure 4 is more subtle, and therefore more difficult to label as erroneous, rather than perhaps reflecting a learning effect.



	Obs	% Pred
FVC	6.72 L	145
FEV1	4.53 L	119
PEF	10.2 L/s	128

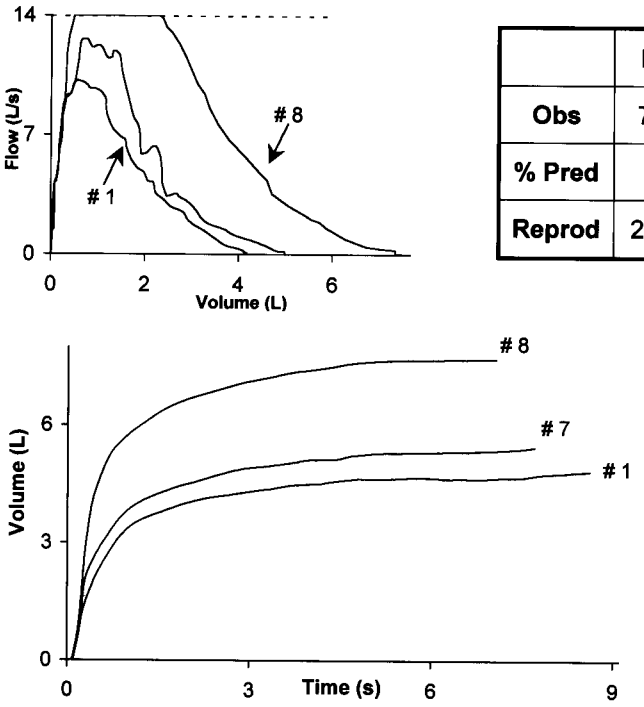
FIGURE 2. Larger zero error. A more extreme shift in the calibration curve during zeroing causes this zero error to be larger than that shown in Figure 1. Note that a hesitating start, a common technical error, is also apparent in Figure 2. See Figure 1 legend for expansion of abbreviations.

EFFECTS OF ERRORS ON RESPIRATORY SCREENING PROGRAMS

Complete sets of spirometry results were obtained from three small respirator medical clearance programs: one from an industrial facility that manufactured pet food, and two from mobile testing programs for employees of household waste and sewage management facilities. Since the programs had comparable spirometers and appeared to have similar problems with zero errors and sensor contamination, the three groups were combined for this report. There were 121 men in the three programs (age range, 20 to 60 years; mean, 40 years); 47% of the 121 men smoked 1 to 40 cigarettes per day (median, 20 cigarettes per day), and 53% were nonsmokers. Though the average percentage of predicted expected for a normal unexposed population is approximately 100%, lower values would be expected in this respirator medical clearance group since nearly half of the men were cigarette smokers.

However, Figure 5 shows that 94 of the 121 men (78%) recorded FEV_1 and FVC $> 100\%$ of predicted; 40% of the group had FEV_1 and FVC $> 120\%$ of predicted. The FEV_1 percentage of predicted ranged from 64 to 287%: 3 men had FEV_1 values $\leq 80\%$ of predicted, 22 men had FEV_1 values 81 to 100% of predicted, 45 men had FEV_1 values 101 to 120% of predicted, and 51 men had FEV_1 values $> 120\%$ of predicted. The FVC percentage of predicted ranged from 83 to 316%: 16 men had FVC values 81 to 100% of predicted, 40 men had FVC values 101 to 120% of predicted, and 65 men had FVC values $> 120\%$ of predicted.

Such spirometry results are highly improbable for a middle-aged working population, of whom 47% are current smokers. These results are probably false-negatives for many subjects, with pulmonary function that is sometimes grossly overestimated. Since the object of respirator medical clearance programs is to find reductions in pulmonary function, these results are "not only useless, but also convey false information which could be harmful to the employee"[1] if a worker is placed in a job with personal protective equipment that he cannot tolerate well. Probable zero errors and sensor contamination have made these screening programs ineffective.



	FVC	FEV1	PEF
Obs	7.6 L	5.8 L	15.9 L/s
% Pred	144	135	166
Reprod	2.29 L	1.95 L	3.9 L/s

FIGURE 3. Sensor contamination. On curve 8, the sensor was probably partially blocked by water vapor condensation, mucus, or the subject's fingers. Note that the PEF of curve 8 exceeds the ceiling recommended by ATS for equipment accuracy, and that obstruction of the sensor causes the reproducibility to be completely unacceptable. See Figure 1 legend for expansion of abbreviations.

RECOMMENDATIONS

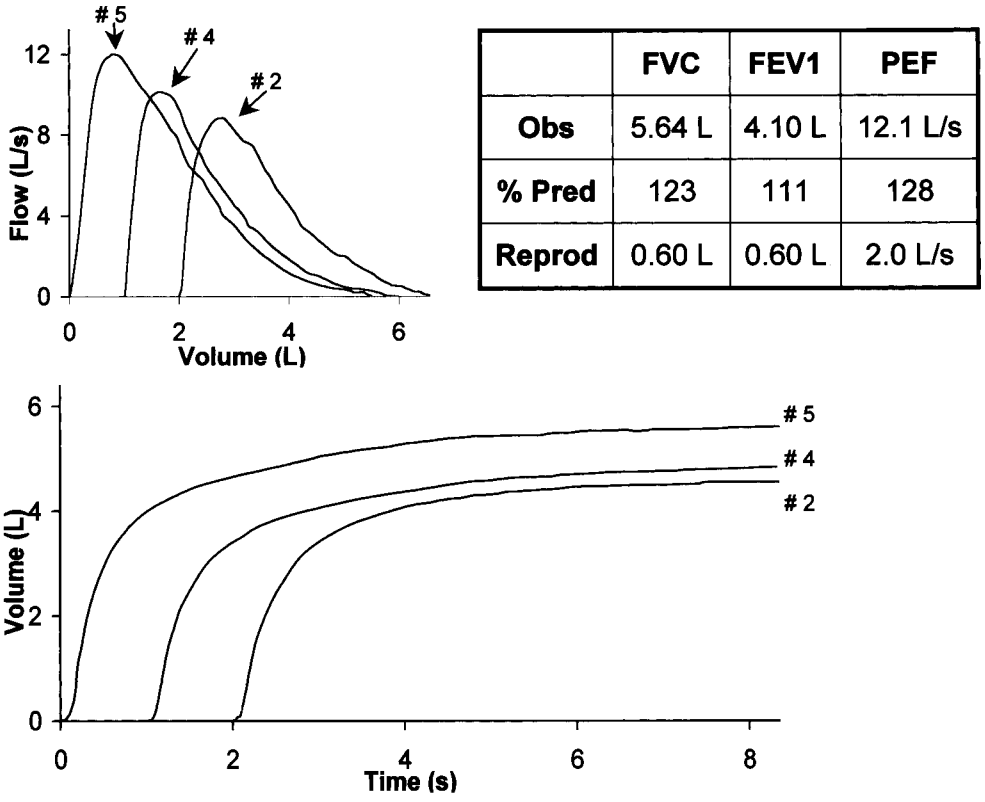
The prevalence of these zero and sensor obstruction errors cannot be documented from the examples presented here, but in the authors' experience, they are not rare: approximately 5 to 10% of the participants in the lead author's National Institute for Occupational Safety and Health-approved spirometry courses submit graphs with problems like these. A survey including spirogram QA reviews of occupational medicine clinics and pulmonary function testing laboratories would permit the prevalence of these errors to be estimated, but such a study is beyond the scope of this report, the intent of which was to draw attention to the problems. Seven recommendations are made based on these observed errors.

(1) Spirometry users at all levels, from the technician to the interpreter of the results, should be aware of the potential for and the appearance of these errors in spirograms. Illustrations of zero errors and contaminated sensor errors should be included along with standard figures of cough and early termination,[4] so that these errors are included in assessments of technical quality and QA protocols. If QA protocols do not include such errors, users will derive false assurance from having a QA program in place, though errors may occur that the QA protocol is not programmed to detect.

(2) Failure to achieve FEV_1 and FVC reproducibility may not be due to the subject's failure to give maximal or consistent efforts, as is often assumed, but may be due to unsuspected technical problems: zero errors (Fig 1), sensor contamination, or blockage (Fig 3, 4), or changing sensor temperature with consecutive maneuvers.[13]

(3) Spirometry users should query the accuracy of unusually high results, *eg*, $>$ 120 to 130% predicted, particularly if elevated values appear in clusters or if trends of increasing values are seen over a short time period. Many users already scrutinize graphs for technical errors or equipment failures if the FEV_1 and FVC are unexpectedly $<$ 70 to 80% of predicted, since it is generally assumed that spirometry errors reduce test values. However, users tend to be less skeptical of elevated test results, often accepting them at face value.

(4) Subjects with below-average lung function are unlikely to record erroneous results $>$ 120% of predicted; elevating a subject's 90% of predicted to 110% or inflating 70% of predicted to 90% is unlikely to cause suspicions of equipment errors. Therefore, biological calibration checks should be performed with maximal effort at least weekly, more often if large volumes of tests are conducted, and whenever spirometer problems are suspected. Detailed recommendations for using biological



	FVC	FEV1	PEF
Obs	5.64 L	4.10 L	12.1 L/s
% Pred	123	111	128
Reprod	0.60 L	0.60 L	2.0 L/s

FIGURE 4. Less severe sensor contamination. The sensor was partially blocked by water vapor condensation, causing unacceptable reproducibility, but presenting a pattern that is less clearly erroneous than in Figure 3. See Figure 1 legend for expansion of abbreviations.

controls are included in the National Lung Health Education Program Consensus Statement.[14] Unexpected changes in the biological calibration results may signal equipment problems that were not present when the calibration was set or checked.

(5) Real-time displays and printouts that show all curves and that are large enough to meet ATS recommendations and Occupational Safety and Health Administration standards are important for recognizing the errors presented in this report; a limited printout of the single "best curve" is inadequate.[1,4,5] The zero errors shown in Figures 1, 2 were not flagged by computerized QA indexes, and the sensor contamination problems in Figures 3, 4 were flagged only as nonreproducible curves. If there is no acceptable real-time display available, we recommend printing out curves during the test session to evaluate whether zero errors or blocked-sensor errors are developing, as well as to evaluate subject effort.

(6) Spirometer software should permit earlier curves to be reviewed, and eliminated at least from the summary report. Manufacturers should inform their users of the pitfalls that are peculiar to their spirometers. Very sensitive pressure transducers should be zeroed by completely covering the mouth-end of the sensor, placing the sensor mouth-end down on a tabletop. When the transducer is located in the sensor assembly (*eg*, in the handle), the sensor should not be tipped or rotated during testing, to avoid zero errors caused by gravitational effects on the transducer. If a zero error is noted, the test should be stopped, and the sensor should be replaced and re-zeroed. Narrow cigar-shaped sensors must be held carefully so that fingers do not occlude the sensor outlet. If one sensor is used for all subjects, it should be cleaned frequently, following manufacturer instructions.

(7) Real-life field testing of spirometers is recommended to identify types of spirometers that will function well, or are inappropriate for use in high-volume testing settings. Testing by an independent laboratory using the ATS 24 standard waveforms and limited human subject testing[3,4] indicates that a spirometer unit worked under ideal laboratory conditions, which is valuable. However, as with routine calibration (or calibration checks), such testing does not guarantee that the changes described in this report will not occur as subjects are tested throughout the day.

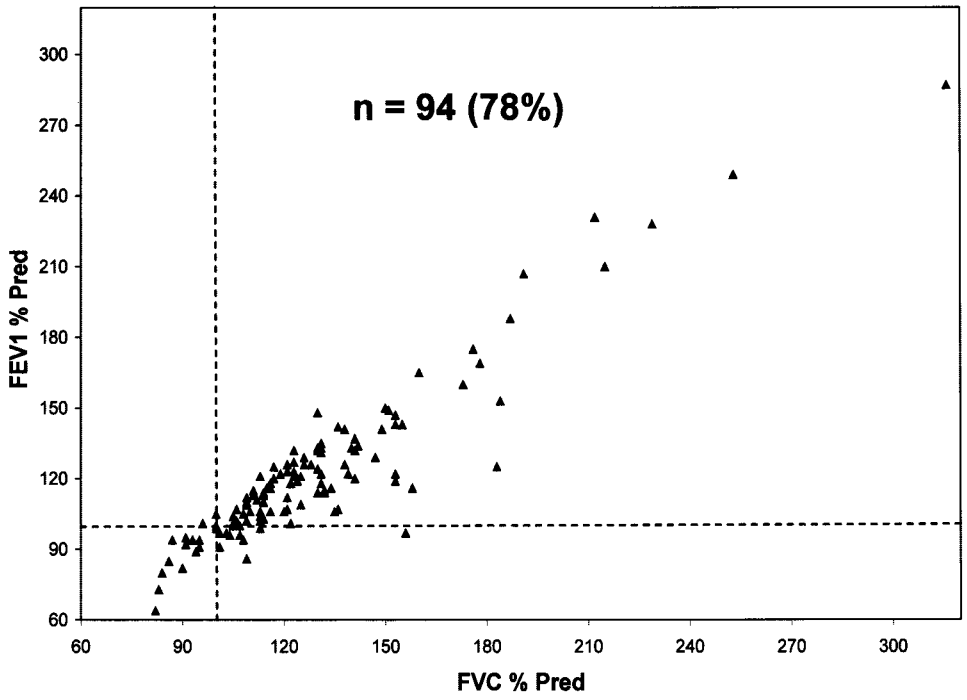


FIGURE 5. Effect of errors on respiratory screening and surveillance programs. Of 121 men screened using spirometers with zero errors and/or sensor contamination problems, 78% recorded both FEV_1 and FVC > 100% predicted, and 40% had both measurements > 120% predicted. See Figure 1 legend for expansion of abbreviation.

CONCLUSION

Unlike poor testing technique, which often reduces spirometry test results, errors due to faulty zeroing or partial blockage of flow-type spirometer sensors can seriously elevate test results, potentially leading to false-negative interpretations rather than false-positives.[15] The existence of these errors and their appearance should become widely recognized among spirometry users so that falsely elevated test results can be avoided or deleted. QA programs should scrutinize spirograms for these zero and sensor-related errors, in addition to evaluating subject effort and calibration records. This is particularly important since these errors can occur in spirometers that have been successfully calibrated on the day of the subject test. Biological calibration checks should be conducted at least weekly, more often if possible, and whenever a problem is suspected. Clusters of high FVC or FEV_1 values and patterns of progressively increasing test results within a short time period should alert users to probable spirometer malfunctions even when calibrations (or calibration checks) have been completed successfully.

ACKNOWLEDGMENT: Dr. Townsend acknowledges the important role played by her students and clients in improving teaching about spirometry, as they continue to bring real-life spirometry problems to her attention.

REFERENCES

1 U. S. Code of Federal Regulations. Title 29, Part 1910.1043 (h) and Appendix D, Cotton Dust, revised 1985
2 Gardner RM, chairman. ATS statement: snowbird workshop on standardization of spirometry. Am Rev Respir Dis 1979; 119:831–838
3 American Thoracic Society. Standardization of spirometry: 1987 update. Am Rev Respir Dis 1987; 136:1285–1298
4 American Thoracic Society. Standardization of spirometry: 1994 update. Am J Respir Crit Care Med 1995; 152:1107–1136
5 Townsend MC. ACOEM position statement: spirometry in the occupational setting. J Occup Environ Med 2000; 42:228–245
6 Townsend MC. The effects of leaks in spirometers on measurements of pulmonary function: the implications for epidemiologic studies. J Occup Med 1984; 26:835–841
7 Hankinson JL. Instrumentation for spirometry: state of the art reviews. Occup Med 1993; 8:397–407
8 Scanlon PD, Hankinson J. Physiologic methods. In: Harber P, Schenker MB, Balmes JR, eds. Occupational and environmental respiratory disease. St Louis, MO: Mosby-Year Book, 1996; 39–54
9 Snow M. Instrumentation. In: Clausen JL, ed. Pulmonary function testing guidelines and controversies: equipment, methods, and normal values. New York, NY: Academic Press, 1982; 33–34
10 Enright PL, Hyatt RE. Office spirometry. Philadelphia, PA: Lea & Febiger, 1987; 76–82
11 Ruppel GL. Manual of pulmonary function testing. St. Louis, MO: Mosby-Year Book, 1998; 252–258

12 Cooper BG, Madsden F. Spirometry. Eur Respir Buyers Guide 2000; 3:40–43
13 Hankinson JL, Viola JO, Petsonk EL, et al. BTPS correction for ceramic flow sensor. Chest 1994; 105:1481–1486
14 Ferguson GT, Enright PL, Buist AS, et al. Office spirometry for lung health assessment in adults: a consensus statement from the National Lung Health Education Program. Chest 2000; 117:1146–1161
15 Townsend MC, Hankinson JL, Lindesmith L, et al. Is my lung function really that good? Spirometer problems that inflate test results [abstract]. Am J Respir Crit Care Med 2002; 165:A200