Pulmonary Function Testing in the Screening of Workers: Guidelines for Instrumentation, Performance, and Interpretation

John L. Hankinson, PhD

Medical surveillance of workers exposed to potential respiratory hazards may be a valuable tool in early recognition and prevention of certain occupational lung diseases. The use of pulmonary function tests, particularly spirometry, has been widely accepted as an integral part of respiratory surveillance. A National Aeronautics and Space Administration contract report on the Occupational Safety and Health Administration medical and workplace surveillance requirements and recommendations by the National Institute for Occupational Safety and Health is a recent detailed study of medical surveillance requirements and recommendations (unpublished study, 1983).

This paper is a brief guide for those in the medical profession attempting to establish or improve their medical surveillance programs for occupational respiratory diseases. It describes procedures to use and techniques for interpreting test results, and finally includes a study of normal reference values. In addition, the references should provide additional information for establishing a respiratory medical surveillance program.

In screening asymptomatic workers, pulmonary function tests serve three purposes: to identify preexisting pulmonary disorders for proper job placement, to detect early changes in pulmonary function in individual workers while intervention may still be effective, and to accumulate data to evaluate how well the exposure controls are working. The results obtained from pulmonary function tests used in a medical surveillance program are usually not sufficient for making a medical diagnosis; therefore, results should be viewed as a means for screening workers to identify those requiring further clinical evaluation.

Of the many different tests of pulmonary function, few are appropriate for medical surveillance purposes. To be appropriate, a pulmonary function test must be simple to administer, inexpensive, safe, and, ideally, both sensitive and specific. The more sensitive tests are usually more variable and require more sophisticated instrumentation, training, and interpretation. The pulmonary function test most widely accepted for medical surveillance is spirometry. Spirometry, which meets all of the criteria above, is probably sufficient for occupational exposures where the only adverse respiratory effect[1] of concern is an obstructive lung disease pattern. However, in those occupational exposures where a restrictive lung disease pattern is of concern, an additional test of single-breath diffusion capacity of the lung for carbon monoxide (DL_{CO}) may be appropriate.

More sophisticated tests of lung function may be useful in the follow-up evaluation of selected workers, but for most surveillance programs, spirometry is sufficient. For example, tests for total lung capacity (TLC), functional residual capacity (FRC), airways resistance (Raw), and residual volume (RV), although useful in diagnosing lung disease, are too expensive and complex for medical surveillance. Sophisticated tests for small airways disease, such as the single breath nitrogen closing volume determinations, helium-oxygen spirometry, and perhaps even the forced midexpiratory flow rate ($FEF_{25\%-75\%}$), are not appropriate because of their large variability within a normal population, and therefore, their potential overinterpretation when used in a surveillance program. In addition, these tests are usu-

From the Clinical Investigations Branch, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, 944 Chestnut Ridge Rd, Morgantown, WV 26505 (address correspondence to Dr Hankinson, chief).

0096-1736/86/2810-1081$02.00/0

ally less reproducible and have a larger within-subject variability. For example, the forced vital capacity (FVC) has a within-subject coefficient of variation (CV) of approximately 3%, whereas the $FEF_{25\%-75\%}$ has a CV of approximately 6%.

Medical and Occupational History

A comprehensive medical and occupational history is an essential component of any occupational medical surveillance program. A respiratory questionnaire providing a record of all employee respiratory symptoms, as well as smoking history, should be included in the medical history. Several valid, reliable, and standardized questionnaires are available, including those of the Medical Research Council (MRC) and the American Thoracic Society (ATS).[2,3] An occupational history should also be taken noting all jobs with specific duties and the nature of all materials to which the employee has been exposed. To develop an environmental exposure history, all periods of exposure up to and including present employment must be recorded.[4,5]

Single-Breath Diffusion Capacity

In those occupational exposures where a restrictive lung disease pattern is of concern, an additional test of single-breath diffusion capacity may be appropriate. Of the three methods of measuring diffusion capacity (single-breath, steady-state, and rebreathing), the single-breath method (DL_{CO}) is recommended because it is most widely used and is being or has been standardized.[6-8] In terms of instrumentation, standardization, and interpretation of results, however, DL_{CO} is 5 to 10 years behind spirometry.

Like spirometry, there are two methods of interpreting DL_{CO} results: the percent predicted and the residual methods. Miller and Scacci[9] recommend the use of greater than 85% of predicted DL_{CO} as normal. However, Crapo and Morris[10] state that "the common practice of designating normal as ±20% may lead to large errors in predicting normal DL_{CO}, because the 95% confidence limits are relatively constant over the entire range of predicted values." Both Crapo and Morris and the ATS[11] recommend that a single value (constant confidence interval) be used for the lower 95% confidence limit. This method is probably preferred unless conclusive studies refuting the constant confidence interval are published.

The predicted reference values of Crapo and Morris[10] and the ATS[11] have been normalized to standard hemoglobin concentrations. Although it is not recommended that venous specimens be routinely obtained, a reduced DL_{CO} may reflect a reduced hemoglobin rather than structural lung damage. Also, smoking is associated with a reduced DL_{CO} both due to elevated carboxyhemoglobin and other changes. Therefore, an abnormal DL_{CO} should be followed by a thorough clinical evaluation before a diagnosis is rendered.

Spirometry Procedures

Spirometry tests should be carried out according to the recommended spirometry standards of the ATS[12] and the National Institute for Occupational Safety and Health (NIOSH).[13] These standards not only establish minimum equipment requirements, but also the procedures to use in administering the tests. Because spirometers have not been subjected to comprehensive instrument compliance testing by an independent test laboratory, spirometer users must perform their own quality assurance and compliance testing. A simplified method for the initial testing of a spirometer or as part of a quality assurance program is given in Appendix A. This simplified method has been found to be almost as effective as more complex testing.[14] It has also been shown that for volume spirometers, any significant ambient temperature changes between two measurements (as in longitudinal or shift change studies) can introduce an error in forced expiratory volume in one second (FEV_1), causing a false significant change in FEV_1 to be observed.[15] The absolute error ranges from approximately 10% at ambient temperatures of 3°C to 2% at 23°C and results from the false assumption in BTPS calculations that a subject's exhaled air cools instantly as it enters the spirometer. The simplest solution to this potential problem is to maintain the spirometer temperatures relatively constant, with a range of no more than 3°C.

Number of Trials or Curves

The ATS and NIOSH spirometry standards[12,13] require that a minimum of three acceptable forced expiratory volume maneuvers (trials) be obtained from each subject, and that the largest and second largest FVCs must be within 5% of each other. Figure 1 shows the percentage for which the largest FVC occurred for each successive trial or curve. These data were obtained from a study of 551 normal subjects, where a minimum of five trials or curves were obtained. For example, the largest FVC occurred on the first trial or curve approximately 12% of the time. More than 20% of the time, the largest FVC occurred on the fourth and fifth trials. Therefore, there may be some benefit in obtaining more than three trials. However, the average difference between obtaining three acceptable curves and a minimum of five curves was less than 50 mL for the FVC. Obtaining more than seven trials is rarely of benefit because of subject fatigue.

Using Largest or Best-Curve Values of Trials

Crapo et al[16] found an approximately 8-mL difference between the best-curve FVC and FEV_1 (single curve with the largest sum of FVC and FEV_1) and the largest FVC and FEV_1, regardless of the curve(s) on which they occurred. Likewise, in the study of 551 normal subjects, only a 10-mL average difference was found between the largest value and best-curve value, based on five or

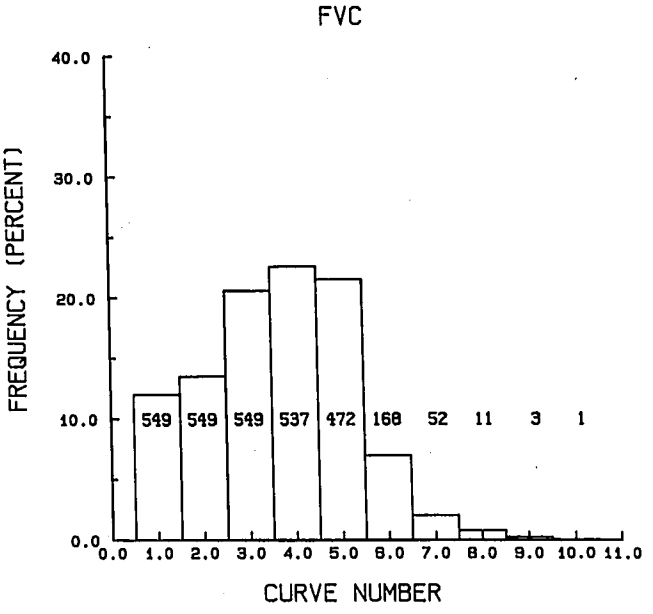

Fig. 1. Frequency or the number of times (in percent) the largest forced vital capacity (FVC) occurred for a given trial (forced expiratory maneuver or curve).

more maneuvers. Table 1 shows these mean FVCs, FEV_1s, and peak flows using three different methods: the largest, the best (sum of FVC and FEV_1), and the average of all (five or more) maneuvers for the population. Although the difference between the largest and best curve methods appears to be slight, we recommend (as does the ATS) that the largest FVC and FEV_1 be reported regardless of the curve(s) on which they occur and all values be corrected to BTPS. Tracings of all FVC maneuvers, either volume-time or flow-volume curves, should be kept as part of an employee's permanent medical record. Without these tracings, the results may be difficult to interpret or validate.

Spirometry Technicians: Critical Component

A critical component for successful spirometry testing is the technician who administers the test. Because valid spirometry depends on the understanding, cooperation, and maximal effort of the subject, the technician's role is critical. They must elicit vigorous subject effort and be able to detect faulty technique by carefully observing the subject and the curves. For this reason, the Institute strongly recommends that technicians complete a NIOSH-approved course. A NIOSH workbook and a manual on spirometry are available from the Institute's Division of Training and Manpower Development.[17,18]

Interpretation of Spirometry Results

As described in the *NIOSH Manual of Spirometry in Occupational Medicine*,[18] the four major objectives of occupational surveillance are to: "1) identify the job applicant with preexisting functional impairment and facilitate his proper placement; 2) detect both occupational and non-occupational disease in an early stage when corrective measures are most likely to be beneficial; 3) identify hazardous working conditions and underscore the need for improvements in industrial hygiene; and 4) reduce the human and economic toll of occupational disease." To achieve these objectives, it is desirable to obtain the best possible test results and interpret them accurately. A variety of methods use spirometry results to achieve these objectives.

TABLE 1
Mean Values for Entire Group Using Three Different Methods: Largest Value, Values from Single Best Curve (largest sum of Forced Vital Capacity [FVC] and Forced Expiratory Volume in 1 Second [FEV_1]), and Average of Five or More Trials (N = 551)

	FVC (L)	FEV_1 (L)	Peak Flow (L/s)
Largest	3.959	3.301	8.115
Best	3.949	3.291	7.486
Average	3.794	3.155	7.161

The three basic methods for interpreting spirometry results in a surveillance program are evaluation of lung function at baseline or at a given point in time; evaluation of changes in lung function during a work shift with exposure to hazardous substances; and evaluation of longitudinal changes in lung function. Comparisons of baseline lung function with appropriately selected predicted normal values are useful in identifying job applicants with preexisting functional impairment. Currently, however, no guidelines or criteria are widely accepted for medical decisions concerning proper job placement or determination of fitness to wear a respirator. Significant decrements in lung function during a work shift with exposure are useful in identifying those workers who may be hypersusceptible to a particular respiratory agent. In terms of chronic respiratory disease, a prudent assumption is that continued exposure to agents that cause acute reductions in lung function may also accelerate longitudinal declines and with continued exposure result in chronic lung disease. Comparisons of actual annual decline with expected values are useful in identifying hazardous working conditions and in detecting both occupational and nonoccupational disease at an early stage, when intervention is most likely to be beneficial. Follow-up examinations are generally more sensitive than a single examination, because intrasubject variability is considerably less than intersubject variability.[19] For this reason, workers' initial values may drop significantly before they fall below the lower limit of predicted values from a normal population.

Comparison of Observed Values With Predicted Normal Values

Two basic methods may be used for comparing a subject's observed with predicted value to determine if the observed value is within the normal range for age, height, sex, and ethnic background. One method is to express the observed value as a percent of the predicted value, and as a value in terms of the FEV_1/FVC percent (percent predicted method). For example, in clinical

situations, the lower limits of normal have typically been taken as 80% of the predicted value for FVC, and/or an FEV_1/FVC percent of 70%. Table 2 gives the criteria for abnormality originally recommended by the Intermountain Thoracic Society.[7] The second, more recent, method is to determine if the subject's value falls within the 95% confidence interval of a reference population (residual method). With this technique, the value, one confidence interval, (1 CI) in Table 3 is subtracted from the predicted value; an observed value that falls below this critical value is considered abnormal. The simplest formula for approximating the 95% confidence interval using a one-tailed *t* test is to multiply the SE of the estimate (SEE) by 1.645. Crapo,[16] the ATS,[11,20] the American College of Chest Physicians (ACCP),[21] and the most current draft of the revised Intermountain Thoracic Society manual[6] recommend use of the residual method. Table 3 gives the criteria for abnormality using the residual method.[22]

Both methods have advantages and disadvantages. The simpler percent predicted method has been used for many years, and a large amount of clinical experience has been acquired. In contrast, the 95% confidence interval (residual method), although appearing on the surface to be statistically more sound, has had only a short period of clinical experience. The 95% confidence interval method has the advantage of accounting for the variability of the measurement within a normal population. Therefore, those measurements that have greater variability within a normal population must exhibit a greater change for a subject to be classified as abnormal. For example, a larger percent change must be observed for the $FEF_{25\%-75\%}$ than for the FVC for a subject to be classified as having an abnormal test result, because of the larger variability of the $FEF_{25\%-75\%}$ than the FVC within a normal population. In addition, the residual method of classifying subjects with respect to obstructive lung disease corrects for age, which is often not considered when the $FEV_1/FVC\%$ is used to classify subjects.

TABLE 2
Criteria for Mild, Moderate, and Severe Obstructive and Restrictive Lung Disease Using Percent Predicted and Forced Expiratory Volume in 1 Second/Forced Vital Capacity Percent ($FEV_1/FVC\%$)[7]

	Obstructive Lung Disease $FEV_1/FVC\%$	Restrictive Lung Disease FVC/FVC Predicted (%)
Normal	>69	>80
Mild	61–69	66–80
Moderate	45–60	51–65
Severe	<45	<51

TABLE 3
Criteria for Mild, Moderate, and Severe Obstructive and Restrictive Lung Disease Using 95% Confidence Intervals[22]*

	Obstructive Lung Disease $FEV_1/FVC\%$	Restrictive Lung Disease Predicted FVC Minus Measured FVC
Normal	< 1 CI	< 1 CI
Mild	≥ 1 to 2 CI	≥ 1 to 1.75 CI
Moderate	≥ 2 to 4 CI	≥ 1.76 to 2.5 CI
Severe	≥ 4 CI	≥ 2.5 CI

* Abbreviations used are FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity.

Two assumptions, as yet to be validated for the 95% confidence interval method, are that the confidence interval for all ages and heights is a constant and that the distribution about the predicted value is Gaussian. The assumption of a constant confidence interval means that, to be classified as abnormal, those subjects with smaller FVCs or FEV_1s must have larger percent deviations from the predicted value than subjects with larger FVCs and FEV_1s. For example, tall subjects must exhibit the same deviation in liters from their predicted value as shorter subjects with a predicted FVC and FEV_1 of several liters less. Also, younger and older subjects must exhibit the same deviation from the predicted value, even though their predicted values are different. Figure 2 shows the predicted linear regression for FEV_1 *v* height for white males. Figures 3 through 6 show the predicted linear regressions for FEV_1 *v* age in white and black males and females studied. The dashed lines show the lower limits of normal using the residual and the percent predicted methods. Notice in Fig. 2 and 3 that for younger and taller white males (larger FEV_1s), the residual method classifies more of the normal white males as abnormal than the percent predicted method. On the other hand, for older and shorter white males (smaller FEV_1s), the residual method classifies fewer white males as abnormal. Crapo et al[16] found "the 95 percent confidence intervals (residual method) for all predicted spirometric values are relatively constant as height and age vary from the mean height and age of the sample population." They recommended that 95% confidence intervals be used to define the lower limit of normal and that the practice of subtracting 20% be abandoned.

Knudson et al[23] questioned whether the distribution about the predicted value was Gaussian; therefore, they did not recommend use of the lower 95% confidence

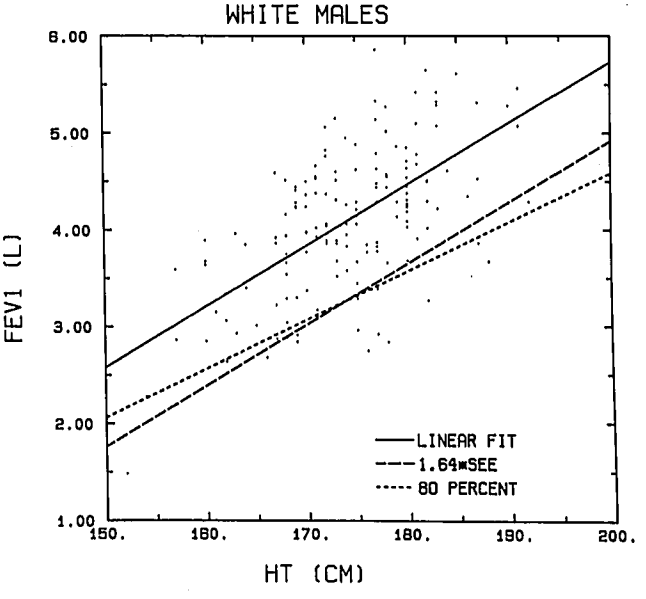

Fig. 2. Observed forced expiratory volume in 1 second (FEV_1) (liters) *v* height (cm): solid line is linear least square fit; dashed lines are the lower limits of normal using residual and percent predicted methods.

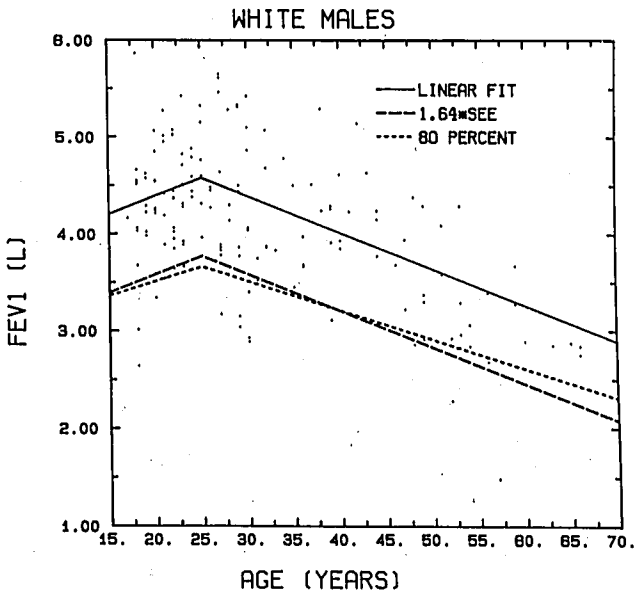

Fig. 3. Observed forced expiratory volume in 1 second (FEV_1) (liters) *v* age (years) for white males using two linear curve model: solid line is linear least square fit; dashed lines are the lower limits of normal using residual method (1.645 SE of the estimate [*SEE]) and percent predicted methods (80% of predicted).

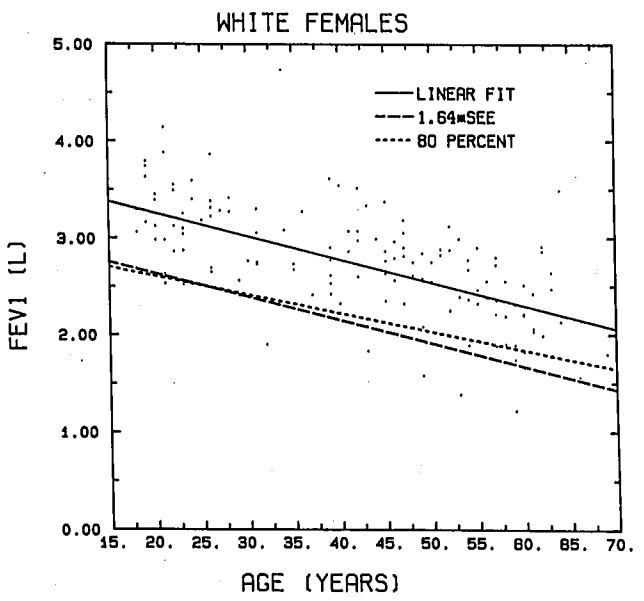

Fig. 5. Observed forced expiratory volume in 1 second (FEV_1) (liters) *v* age (years) for white females: solid line is linear least square fit; dashed lines are the lower limits of normal using residual method (1.645 SE of the estimate [*SEE]) and percent predicted methods (80% of predicted).

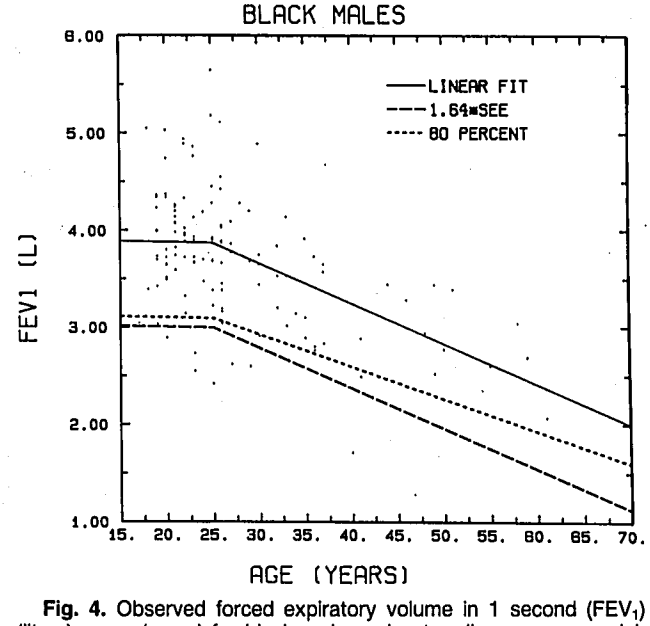

Fig. 4. Observed forced expiratory volume in 1 second (FEV_1) (liters) *v* age (years) for black males using two linear curve model: solid line is linear least square fit; dashed lines are the lower limits of normal using residual method (1.645 SE of the estimate [*SEE]) and percent predicted methods (80% of predicted.).

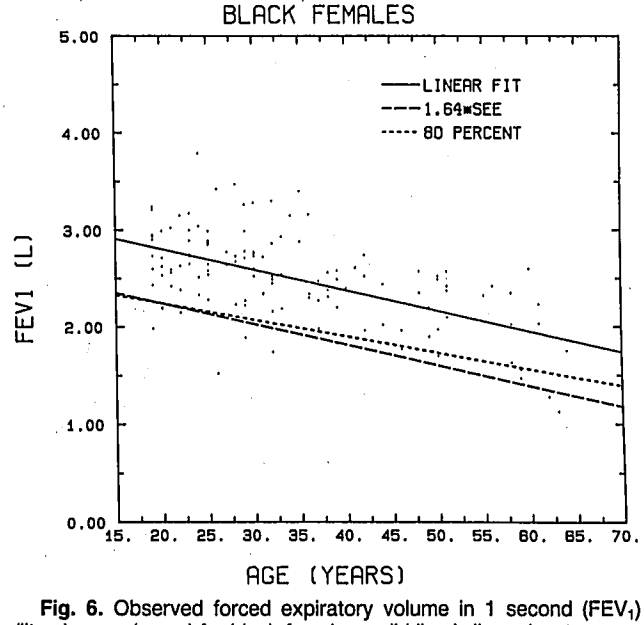

Fig. 6. Observed forced expiratory volume in 1 second (FEV_1) (liters) *v* age (years) for black females: solid line is linear least square fit; dashed lines are the lower limits of normal using residual method (1.645 SE of the estimate [*SEE]) and percent predicted methods (80% of predicted).

interval. Instead, they recommended, "to deal with skewness and variability, we regard the lower 95th percentile as a more consistent and preferable criterion for defining the limits of normal."

Because both the ATS[12, 21] and the ACCP[23] recommend the use of the residual method, it is clear that this method will find wider use, and additional clinical experience will be obtained. However, complete acceptance of the residual method should await additional studies with sufficient subjects in all age and height groups to determine the distribution of normal values about the predicted value. Until these studies have been completed, both the residual and percent predicted methods are regarded as acceptable.

Changes in Lung Function During a Work Shift

Significant decrements in lung function during a work shift with exposure are useful in identifying those work-

ers who have acute reactions in lung function to particular respiratory agents. A prudent assumption is that continued exposure to agents that cause such acute reductions in lung function may also result in chronic lung disease with continued exposure. Therefore, limiting these workers' exposure to such agents may prevent them from developing chronic respiratory disease. Decrements in FEV_1 over a work shift of more than 5%, provided this decrement is repeated on more than one occasion, should warrant follow-up clinical evaluation and possible limitation of further exposure. For those workers with an FEV_1 of less than 3 L, a decrement of greater than 150 mL should warrant follow-up.

To establish the limits for a significant change over a work shift, the day-to-day variability of the lung function parameter must be considered. Several studies have found the day-to-day within-subject coefficient of variation to be approximately 3%.[24-27] In addition, Pennock et al[26] and Rozas et al[27] regard a greater than 5% change within a day as significant. However, small decrements greater than 5% should be repeated before any action is taken. If only one shift examination is available, then a greater than 10% decrement is more prudent to prevent a large number of false-positive results in the absence of repeat examinations. The 5% criteria cannot be used for those workers with smaller FEV_1s (less than 3 L), because the absolute change in milliliters can become extremely small or within the background noise for the measurement.

Comparison of Observed Values With Previous Measured Values

Comparing a subject's observed value with a previously determined value is one method advocated to improve the sensitivity of spirometry without losing specificity. The day-to-day coefficient of variation of FVC within a subject is approximately 3%,[24-27] although the week-to-week variability is larger.[25,26] In contrast, the population coefficient of variation among subjects of the same age and height is approximately 10%. Because the coefficient of variation is smaller using the subject as his own control, this approach should give better sensitivity than comparisons with a normal reference population. However, since the expected yearly decrements in FVC and FEV_1 are small (25 to 30 mL/yr in nonsmokers and 40 to 50 mL/yr in smokers), an abnormal decrement must either be large or observed over a long period of time to be detected. For example, a coefficient of variation of 3% with an FEV_1 of 4.00 L suggests that an excess annual decrement of at least 120 to 240 mL (3 to 6% of FEV_1) must be observed to classify a subject as abnormal. This is four to ten times greater than would be expected in 1 year due to the normal aging process. In addition, because of the larger week-to-week coefficient of variation, Pennock et al[26] recommended a 12% change in FEV_1 from baselines as significant (after correcting for aging).

The problem of large variability in short-term longitudinal studies of lung function has been noted by Burrows et al[28] and Hankinson et al.[29] Hankinson, in a 2-year longitudinal study of 116 coal miners, found a substantial variability in FVC and FEV_1. The conclusion was that the practical usefulness of short-term (less than every 2 years) examinations of pulmonary function in a medical surveillance program seemed questionable, at least with regard to detecting small excesses in annual decrements in lung function.

To illustrate the problems in short-term longitudinal studies, Fig. 7 shows the FEV_1 results of 17 examinations, which were conducted at 6-week intervals over a 2-year period for seven different subjects. In Fig. 7 (far right-hand column), notice the considerable variability in annual decrements in FEV_1 between subjects (from +14 mL/yr to −150 mL/yr). Although a cyclical pattern appears for some subjects, no consistent patterns emerge when the data from all subjects are combined. Also notice that a wide range of annual decrements could be obtained in the same subject depending on which two data points were used to compute the annual decrement. This problem is illustrated in Fig. 8 where the FEV_1 results for subject B, Fig. 7 (male, nonsmoker, 34 years old), are analyzed for different time periods. Three different annual decrements are obtained depending on which data points are used in the analysis. When points "A" and "B" are used, subject B has an annual increase in FEV_1 of 36 mL/yr. When points "X" and "Y" are used, subject B has an annual decrement of 123 mL/yr. These results demonstrate one of the major difficulties in interpreting short-term longitudinal spirometric data.

A more accurate estimate of annual decrement can be obtained if multiple tests are conducted over a longer period of time. For example, two additional tests of FEV_1 were obtained for subject B approximately 5 years after his initial exam. Notice in Fig. 9 that his annual

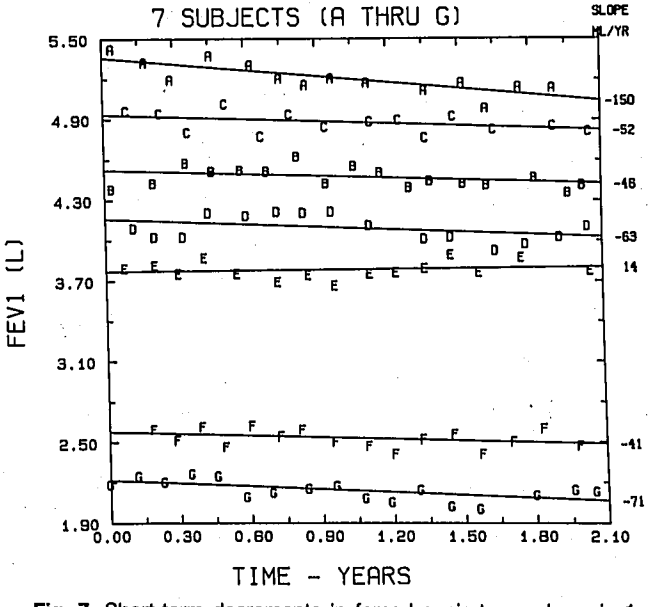

Fig. 7. Short-term decrements in forced expiratory volume in 1 second (FEV_1) (liters) for seven different subjects (A through G) over a 2-year period. Values were obtained approximately every 6 weeks; numbers to right of figure are estimated annual decrements in FEV_1 in milliliters per year.

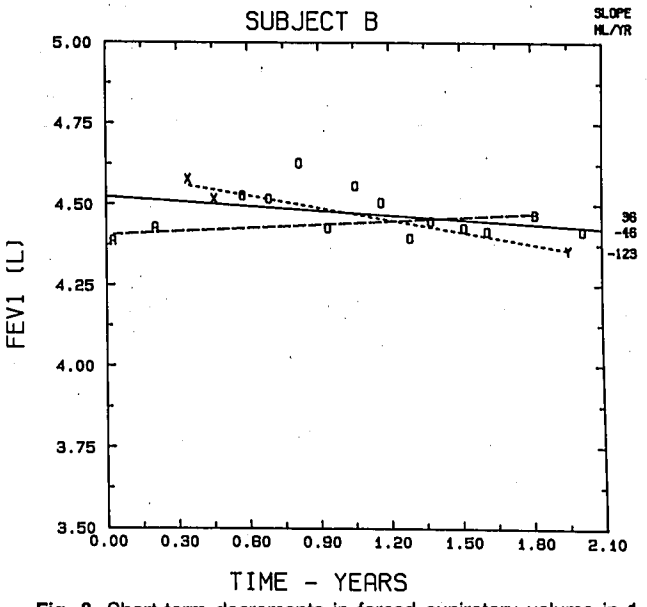

Fig. 8. Short-term decrements in forced expiratory volume in 1 second (FEV_1) for subject B during a 2-year interval: solid line is original line estimated using all of the data (−46 mL/yr). Dashed line between points A and B yields an estimated increase in FEV_1 of 36 mL/yr. Dashed line between points X and Y yields an estimated annual decrement in FEV_1 of 123 mL/yr. Numbers to right of figure are estimated annual decrements in FEV_1 in milliliters per year.

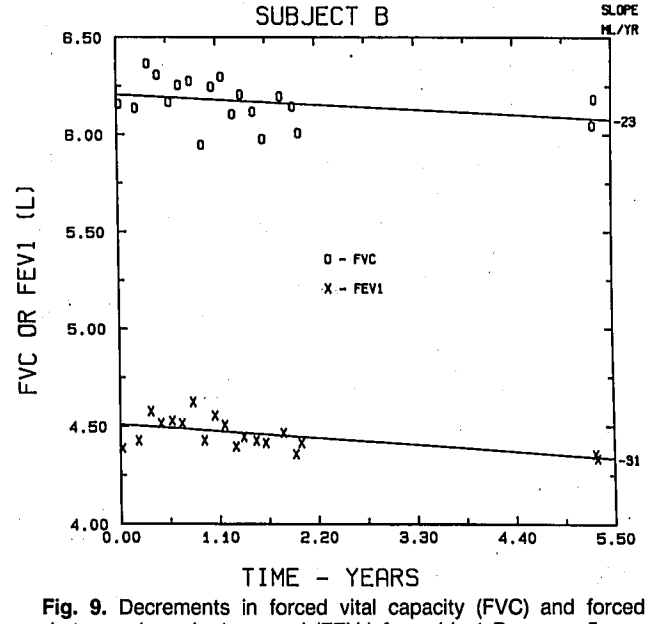

Fig. 9. Decrements in forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV_1) for subject B over a 5-year interval of time: numbers to right of figure are estimated annual decrements in FEV_1 in milliliters per year.

decrement in FEV_1 is estimated to be 31 mL/yr when this longer interval of observation is used. However, multiple studies at the 5-year interval are recommended. As can be seen in Fig. 9, the annual decrement in FVC would be significantly different if only one of the two FVC results (obtained on different days at the end of 5 years) were used.

The occurrence of potential "survey biases" as described by Glindmeyer et al[30] must also be considered in the interpretation of longitudinal results. The ATS recommends that spirometers have an accuracy of ±3% of reading. This means that if an FEV_1 of 4.00 L is observed, then its actual value should be between 3.88 and 4.12 L. Because the precision (reproducibility) of a spirometer is usually better than its accuracy, the instrumentation variability described above may be reduced by using the same spirometer for all longitudinal examinations. However, the potential for a systematic bias during a set of examinations exists because spirometer calibration may change from time to time. This systematic bias can usually be detected by comparing the mean FVCs and FEV_1s for a group of subjects studied on a given day with those of the same group of subjects studied on previous occasions. If a large difference exists between these group means, in excess of what might be expected due to aging, then a "survey bias" may have occurred.

This technique is illustrated in Fig. 10, where the simulated mean FEV_1 for the same 25 subjects is plotted as a function of time for six different studies. Notice in Fig. 10 that for most of the examinations, a slight annual decrement is observed. However, the FEV_1 at the 3-year examination is obviously elevated, possibly due to a survey bias. If only the initial and 3-year examinations had been used (dashed line), this group of subjects would have demonstrated an increase in FEV_1 over the 3-year time interval. The most likely explanation for the higher result demonstrated in Fig. 10 is a very slight spirometer calibration difference at the time of the 3-year examinations. If a survey bias is observed, then those data must be eliminated from the analysis or be corrected.

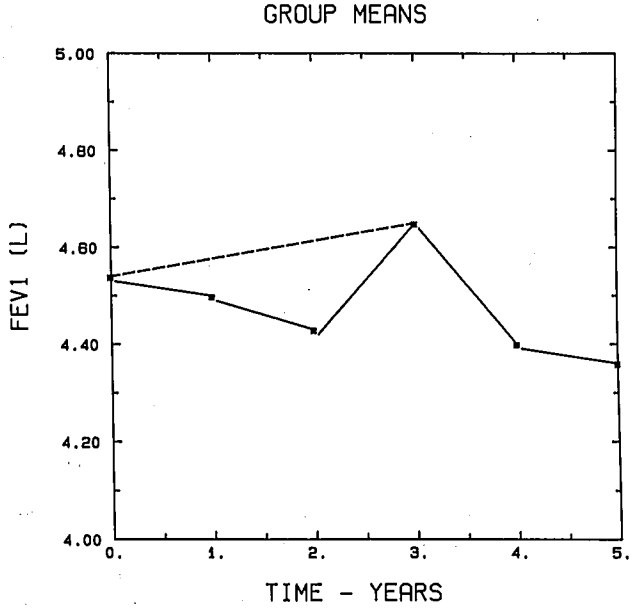

Fig. 10. Illustration of group means (same 25 subjects in each group) *v* time: the group mean observed on the occasion of the 3-year examination illustrates introduction of a "survey bias." Dashed line illustrates error in estimating annual change in forced expiratory volume in 1 second (FEV_1) due to "survey bias."

The above discussion should clearly show that for longitudinal studies to be valid and reliable, it is imperative that an effective quality assurance program be in place such as the ATS has recommended.[31] In addition, multiple initial examinations, perhaps during the first 3 weeks of employment, should be obtained to provide more stable baseline values. All future values could then be referenced to these values to compute an annual decrement. Any interpretation of follow-up examination results should consider the number of test replications and the time interval during which the data were collected.[19] In the absence of conclusive longitudinal studies of normal workers, this study is in agreement with the *NIOSH Manual of Spirometry in Occupational Medicine*[18] recommendation: a worker whose annual decline in FEV_1 or FVC is greater than 10% (after correcting for age) should receive a follow-up clinical evaluation. This recommendation is similar to the conclusion of Pennock et al,[26] who determined that significant week-to-week changes in normal subjects are 11% for FVC and 12% for FEV_1.

Predicted Normal Values

As discussed previously, the interpretation of spirometry results requires comparison with appropriately selected predicted "normal" values. Although many predicted normal values or equations are available, only a few of these values were obtained through studies using methods that meet or exceed the minimum ATS and NIOSH standards for spirometry. None of these studies include values for blacks or, specifically, active workers who have not been exposed and are asymptomatic. In addition, because NIOSH methods exceed those required by the ATS, values obtained by NIOSH investigators may differ from those obtained in other studies; therefore, the following study of normal reference values was conducted to develop values based on NIOSH predictions for use in studies by the Institute.

Methods

The study population of nonexposed, asymptomatic workers was obtained from a number of different NIOSH studies.[32,33] Nonexposed control subjects from the NIOSH Blue Collar Study, Portland Cement Study, Nontextile Cotton Study, and Health Hazard Evaluations were included in the study population. Subjects who had any significant exposure, moderate or severe respiratory symptoms of cough, phlegm, dyspnea, or wheezing, and any history of cigarette smoking were excluded from the study population. Specifically excluded were the following: all those "nonexposed" control subjects with cough or phlegm on most days for as much as 3 months during the year; those with shortness of breath while walking with other people of the same age; and those with attacks of shortness of breath with wheezing. All subjects performed at least five forced expiratory maneuvers using equipment (Ohio Medical Products model 840) and procedures which met the NIOSH and ATS minimum requirements. Because the ATS only requires three acceptable curves, with the two largest FVCs within 5% of each other, a subset of the data was used in the "three-curve" analysis. Each subject's FVC maneuvers were analyzed in the sequence in which they were performed and were compared to the minimum ATS requirements. Once the minimum ATS requirements for acceptable performance had been met, no additional curves were used in the three-curve analysis. For peak flow, FVC, and FEV_1, the largest value was used regardless of the curve(s) on which they occurred. Other flow rates were taken from the best curve (curve with largest sum of FVC and FEV_1). For comparison, all of the curves were used to derive the maximum envelope flow volume curve which was used in a separate analysis (five-curve analysis).

Prediction equations were derived using linear least squares regression analysis with multiple variables. For example, for FVC, two equations were used:

$$FVC = C_1 + C_2 \times age + C_3 \times height$$

$$FVC = C_1 + C_2 \times age + C_3 \times height + C_4 \times (age - 25) \times group$$

where group = 0 if age < 25; group = 1 if age ≥ 25.

For calculating a predicted value, the second equation can be simplified to two equations (age < 25 and age ≥ 25) of the form:

$$FVC = K_1 + K_2 \times age + K_3 \times height.$$

Results

Means and SEMs for age, height, and spirometric parameter by sex and race are shown in Table 4. All groups contained at least 126 subjects. The age distribution can be seen in Figs. 3 through 6 showing FEV_1 *v* age. Notice in Fig. 4 that relatively few black males were older than 40 years of age. Having only a few subjects in the older age group may have resulted in a less reliable prediction equation for black males.

It was apparent from the results and from the studies by Knudson[23,34] that older males may have a different slope in their pulmonary function parameters than younger males. Therefore, we performed regression analysis using both a single linear fit and two linear fits, forcing the lines to intersect at 25 years of age for male subjects and at 20 years of age for female subjects. Figures 3 and 4 show the FEV_1 results for white and black males. The corresponding age coefficients are given in Table 5. The fit of observed data with the model using the two linear curves was slightly better (addition coefficients, associated with using (two) linear fit model to the data, statistically different from zero) than the fit with a single linear model for the males. However, no corresponding improvement was observed for the females, probably because there were few female subjects less than 20 years of age in our working population.

To facilitate comparison among various groups, pre-

TABLE 4
Means and SEMs for Study Population, by Sex and Race (Using Three-Curve Analysis)

Parameter	White Males	Black Males	White Females	Black Females
No.	158	129	141	126
Age (yr)	31.7 (1.01)	27.8 (0.86)	41.6 (1.24)	34.3 (1.13)
Height (cm)	174 (0.59)	175 (0.65)	161 (0.50)	162 (0.51)
FVC* (L)	4.93 (0.07)	4.33 (0.07)	3.32 (0.05)	2.93 (0.05)
FEV_1 (L)	4.08 (0.06)	3.66 (0.06)	2.71 (0.05)	2.48 (0.04)
Peak flow (L/s)	9.62 (0.14)	9.39 (0.16)	6.27 (0.11)	6.24 (0.14)
$FEF_{50\%}$ (L/s)	5.09 (0.10)	4.87 (0.14)	3.45 (0.10)	3.45 (0.11)
$FEF_{75\%}$ (L/s)	2.10 (0.07)	1.94 (0.07)	1.32 (0.07)	1.37 (0.06)

* Abbreviations used are: FVC, forced vital capacity; FEV, forced expiratory volume in 1 second; FEF, forced midexpiratory flow rate.

TABLE 5
Summary Prediction Equations Using Two-segment Linear Fit Only With Statistically Significant Improvement in the Model (three-curve analysis)*

	Parameter	Intercept		Age		Height	SEE
		<25	≥25	<25	≥25		
White males	FVC	−7.050	−5.484	0.0343	−0.0283	0.0655	0.610
	FEV_1	−5.449	−4.451	0.0073	−0.0326	0.0554	0.493
	FEV_1/FVC%†	90.615	90.615	−0.2532	−0.2352	N.S.	5.485
White females	FVC†	−3.149	−3.149	−0.0177	−0.0177	0.0446	0.496
	FEV_1†	−1.869	−1.869	−0.0233	−0.0233	0.0345	0.381
	FEV_1/FVC%†	93.391	93.391	−0.2600	−0.2600	NS	6.154
Black males	FVC	−4.606	−2.723	0.0377	−0.0376	0.0472	0.652
	FEV_1	−2.077	−1.081	−0.0018	−0.0416	0.0343	0.532
	FEV_1/FVC%†	94.202	94.202	−0.3321	−0.3321	NS	6.011
Black females	FVC†	−3.901	−3.901	−0.0111	−0.0111	0.0446	0.448
	FEV_1†	−2.311	−2.311	−0.0191	−0.0191	0.0338	0.341
	FEV_1/FVC%†	97.380	97.380	−0.3388	−0.3388	NS	8.244

* Abbreviations used are: SEE, standard error of the estimate; FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 second.
† The intercept and age coefficients are identical in both age groups; therefore, a single linear fit was used.

dicted values were calculated using an age of 34 years (approximately equal to the mean for the entire study population) and a height of 174.5 cm for males and 161.5 cm for the females. Notice in Tables 6 (males) and 7 (females) that the FVCs and FEV_1s were approximately 15% higher for the whites, regardless of sex. The flow rates were also higher for the white subjects. In contrast, there were no differences in the FEV_1/FVC% between the races. Figure 11 shows the flow volume curves for the males using the three- and five-curve analyses, and Fig. 12 shows the flow volume curves for females. As shown in Figs. 11 and 12, the white subjects had larger flow volume curves than the black subjects. As expected, slightly larger values are obtained when more curves (five curves *v* three curves) are used in the analysis. The FVCs and FEV_1s are approximately 1% to 2% higher with five curves and the flow rates are 3% to 7% higher.

Table 8 compares prediction equations from various studies. Notice in Table 8 that the age coefficients with a single linear fit are not as negative for FVC as reported by Knudson[33,34] and Morris.[35] However, the age coefficients are very similar to those of Knudson when a segmented linear fit is used for the males. This was expected, as Knudson also used a segmented (two) linear fit. A linear fit to FVC and FEV_1 *v* age for males older than 25 years of age results in a more negative age coefficient. The predicted FVC is about the same as reported by Knudson but less than that of Morris.[35] The

TABLE 6
Predicted Values for Several Spirometric Parameters for Black Males (N = 129) and White Males (N = 158), Aged 34 Years and Height 174.5 cm

Parameter	Using Three Curves		% Difference (W-B)/W	Using Five Curves		% Difference (W-B)/W
	White	Black		White	Black	
FVC* (L)	4.90	4.16	15.1%	4.94	4.20	15.0%
FEV_1 (L)	4.06	3.49	14.0%	4.09	3.48	14.9%
FEV_1/FVC%	82.6	82.9	0.3%	82.7	83.0	0.4%

* Abbreviations used are FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 second.

TABLE 7
Predicted Values for Several Spirometric Parameters for Black Females and White Females, Aged 34 Years and Height 161.5 cm

Parameter	Using Three Curves		% Difference (W-B)/W	Using Five Curves		% Difference (W-B)/W
	White	Black		White	Black	
FVC* (L)	3.45	2.92	15.4%	3.50	2.97	15.1%
FEV_1 (L)	2.91	2.50	14.1%	2.96	2.53	14.5%
FEV_1/FVC%	84.5	85.9	1.5%	84.4	85.6	1.3%

* Abbreviations used are: FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 second.

age coefficients and predicted values for FEV_1 are similar to those of Morris and Knudson. Table 9 shows a comparison of the SE of the estimate (SEE) for various studies of white males and females. Our SEEs were similar to those of other studies.

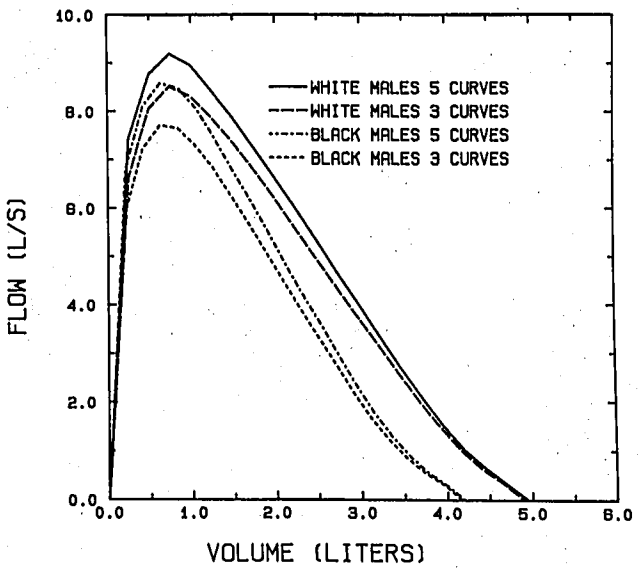

Fig. 11. Predicted flow volume curve for a white and black male (aged 34 years, height 174.5 cm) using both three- and five-curve analyses.

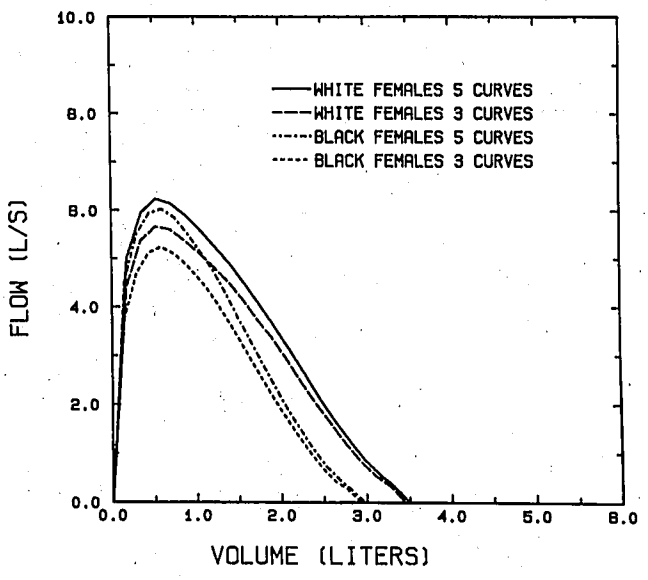

Fig. 12. Predicted flow volume curve for a white and black female (aged 34 years, height 161.5 cm) using both three- and five-curve analyses.

TABLE 8
Comparison of Several Prediction Equations, Using an Age of 34 Years; and for White Males a Height of 174.5 cm and for White Females a Height of 161.5 cm

	Study	Age Coefficient		Predicted	
		Males	Females	Males	Females
Forced vital capacity	Present*	−0.021	−0.018	4.90	3.45
	Present†	−0.021	−0.018	4.94	3.50
	Present‡	−0.028		4.98	
	Knudson[34]	−0.029	−0.022	4.90	3.45
	Knudson[23]	−0.030	−0.017	4.93	3.40
	Morris[35]	−0.025	−0.024	5.08	3.64
	Crapo[16]	−0.021	−0.022	5.09	3.60
Forced expiratory volume	Present*	−0.028	−0.023	4.06	2.91
	Present†	−0.028	−0.024	4.09	2.96
	Present‡	−0.033		4.11	
	Knudson[34]	−0.027	−0.021	3.95	2.85
	Knudson[23]	−0.029	−0.020	4.10	2.90
	Morris[35]	−0.032	−0.025	3.97	2.88
	Crapo[16]	−0.024	−0.025	4.20	3.07

* Using three best curves only and single linear fit.
† Using all five or more curves and single linear fit.
‡ Using three best curves only, and two linear fits (males, age ≥ 25).

Discussion

The age coefficients shown in Table 8 are all somewhat similar, with more negative age coefficients being observed for the older males when a segmented (two) linear fit is used. This is true in both our study and those of Knudson.[23,34] Also notice in Table 8 that Knudson's age coefficients are more negative for FVC than for FEV_1, which indicates that the $FEV_1/FVC\%$ may increase with increasing age.

The lower observed FVC, FEV_1, and flow rates for blacks in comparison with white values are consistent with other studies of normal blacks.[36] Considering the flow volume curves, blacks probably have a smaller thorax than whites of the same height.[36] If white normal reference values are used, then a correction factor of 0.85 of the white values is recommended for FVC and FEV_1, although the preferred method is to use predicted values derived from a black population. No correction of the $FEV_1/FVC\%$ is needed.

The predicted values presented in this paper, as well as those of Knudson et al[23,34] and Crapo et al[16] are recommended, depending on whose values best reflect the population. No single set of predicted values applies to all populations. We therefore agree with the recommendation of Clausen[37] to measure a small number of normal subjects (n = 10 to 20) who represent the population that will be evaluated. If more than three out of ten or six out of 20 of the presumably normal subjects are classified as abnormal using the criteria, then something may need to be changed. Either the testing procedure is at fault or the set of normal reference values chosen is inappropriate.

Additional sources of information concerning pulmonary function medical surveillance are contained in References 38 through 43.

TABLE 9
Comparison of SE of The Estimate (SEE) and Number of Subjects (N) for Various Studies of White Males and Females

Study	Forced Vital Capacity				Forced Expiratory Volume in 1 Second	
	Males	N	Females	N	Males	Females
Present*	0.619	158	0.496	141	0.500	0.381
Present†	0.624	158	0.508	141	0.487	0.385
Present‡	0.610	158		141	0.493	
Knudson[34]	0.601	128	0.519	321	0.541	0.434
Knudson[23]	0.638	86	0.493	204	0.524	0.388
Morris[35]	0.740	517	0.520	471	0.550	0.470
Crapo[16]	0.645	125	0.391	126	0.482	0.328

* Using three best curves only and single linear fit.
† Using all five or more curves and single linear fit.
‡ Using three best curves only, and two linear fits (males, age ≥ 25).

References

1. American Thoracic Society: Draft statement on guidelines as to what constitutes an adverse respiratory health effect, with special reference to air pollution. *ATS News*, Winter 1984, pp 6–8.

2. Samet JM: A historical and epidemiologic perspective on respiratory symptoms questionnaire. *Am J Epidemiol* 1978;108:435–446.

3. Ferris BG: Epidemiology standardization project: Respiratory questionnaires. *Am Rev Respir Dis* 1978;118:7–53.

4. Goldman B, Rose H, Peters JM: The occupational environmental health history. *JAMA* 1981;246:2831–2836.

5. The Occupational and Environmental Health Committee of the American Lung Association of San Diego and Imperial Counties: Taking the occupational history. *Ann Intern Med* 1983;99:641–651.

6. Kanner RE, Morris AH: *Clinical Pulmonary Function Testing:* A manual of uniform laboratory procedures for the intermountain area. Salt Lake City, Intermountain Thoracic Society, 1975.

7. American Thoracic Society: Preliminary recommendations for standard technique: Single breath carbon monoxide diffusing capacity. *ATS News*, Spring 1986, pp 6–12.

8. Ferris BG: Epidemiology standardization project. *Am Rev Respir Dis* 1978;118:7–58; 89–113.

9. Miller WF, Scacci R: Pulmonary function assessment for determination of pulmonary impairment, in Brooks SM, Lockey JE, Harber P (eds): *Clinics in Chest Medicine*, Occupational Lung Diseases II, Vol 2, No. 3. Philadelphia, WB Saunders Co, 1981, pp 327–341.

10. Crapo RO, Morris AH: Standardized single breath normal values for carbon monoxide diffusing capacity. *Am Rev Respir Dis* 1981;123:185–190.

11. American Thoracic Society Statement: Evaluation of impairment/disability secondary to respiratory disease. *Am Rev Respir Dis* 1982;126:945–951.

12. Gardner RM (chairman): ATS Statement on Standardization of Spirometry. *Am Rev Respir Dis* 1979;119:831–838.

13. *Federal Register* 1980;45(42):13695–13698.

14. Gardner RM, Hankinson JL, West BJ: Evaluating commercially available spirometers. *Am Rev Respir Dis* 1980;121:73–82.

15. Hankinson JL, Villa JO: Dynamic BTPS correction factors for spirometric data. *J Appl Physiol* 1983;55:1354–1360.

16. Crapo RO, Morris AH, Gardner RM: Reference spirometer values using techniques and equipment that meets ATS recommendations. *Am Rev Respir Dis* 1981;123:659–664.

17. *NIOSH Spirometry Workbook*, Division of Training and Manpower Development. Cincinnati, US Dept of Health and Human Services/PHS/CDC/NIOSH, 1981.

18. *NIOSH Manual of Spirometry in Occupational Medicine*, Division of Training and Manpower Development. Cincinnati, US Dept of Health and Human Services/PHS/CDC/NIOSH, 1981.

19. Berry G: Longitudinal observations: Their usefulness and limitations with special reference to the forced expiratory volume. *Bull Physio-Pathol Respir* 1974;10:643–655.

20. American Thoracic Society Statement: Surveillance for respiratory hazards in the occupational setting. *Am Rev Respir Dis* 1982;126:952–956.

21. American College of Chest Physicians Statement on Spirometry: A report of the section on pathophysiology. *Chest* 1983;83:547–550.

22. Morris AH, Kanner RE, Crapo RO, et al: *Clinical Pulmonary Function Testing: A Manual for Uniform Procedures.* Salt Lake City, Intermountain Thoracic Society, 1984.

23. Knudson RJ, Lebowitz MD, Holberg CJ, et al: Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983;127:725–734.

24. Hankinson JL, Petersen MR: Data analysis for spirometry instrumentation standards, abstracted. *Am Rev Respir Dis* 1977;115(suppl):116.

25. Cochrane GM, Prieto F, Clark TJH: Intra subject variabilities of maximum expiratory flow volume curve. *Thorax* 1977;32:171.

26. Pennock BE, Rogers RM, McCaffree DR: Changes in measured spirometric indices: What is significant? *Chest* 1981;81:97–99.

27. Rozas CJ, Goldman AL: Daily spirometric variability: Normal subjects and subjects with chronic bronchitis with and without airflow obstruction. *Arch Intern Med* 1982;142:1287–1291.

28. Burrows B: An overview of obstructive lung disease. *Med Clin North Am* 1981;65:455–471.

29. Hankinson JL, Hodous TK: Short-term prospective spirometric study of new coal miners, abstracted. *Am Rev Respir Dis* 1983;127:159.

30. Glindmeyer HW, Diem JE, Jones RN, et al: Noncomparability of longitudinally and cross-sectionally determined annual changes in spirometry. *Am Rev Respir Dis* 1982;125:544–548.

31. Proposed American Thoracic Society Recommended Guidelines: Quality Assurance in Pulmonary Function Laboratories. *ATS News*, Spring 1984.

32. Petersen MS, Castellan RM: Prevalence of chest symptoms in nonexposed blue-collar workers. *J Occup Med* 1984;26:367–374.

33. Petersen MS, Hankinson JL: Spirometry reference values for nonexposed blue-collar workers. *J Occup Med* 1985;27:644–650.

34. Knudson RJ, Slatin RC, Lebowitz MD, et al: The maximal expiratory flow volume curve. *Am Rev Respir Dis* 1976;113:587.

35. Morris JF, Koski A, Johnson LC: Spirometric standards for healthy nonsmoking adults. *Am Rev Respir Dis* 1971;103:57–67.

36. Rossiter CE, Weill H: Ethnic differences in lung function: Evidence for proportional differences. *Int J Epidemiol* 1974;3:55–61.

37. Clausen JL: *Pulmonary Function Testing Guidelines and Controversies: Equipment, Methods, and Normal Values.* New York, Academic Press, 1982.

38. Turk MH: *Occupational Medicine: Surveillance, Diagnosis, and Treatment.* New York, F and S Press, 1982.

39. Weil H: *Occupational Lung Diseases, Research Approaches and Methods.* New York, Marcel Dekker Inc, 1981.

40. Ruppel G: *Manual of Pulmonary Function Testing.* St Louis, CV Mosby Co, 1982.

41. American Thoracic Society: Evaluation of impairment/disability secondary to respiratory disease. *Am Rev Respir Dis* 1982;126:945–951.

42. American Thoracic Society: Surveillance for respiratory hazards in the occupational setting. *Am Rev Respir Dis* 1982;126:952–956.

43. American Thoracic Society: Screening for adult respiratory disease. *Am Rev Respir Dis* 1983;128:768–774.

Appendix A: Simplified Spirometer Testing Method

Although a particular model of spirometer may have been tested and found to meet the ATS recommendations, another instrument of the same model, or the same instrument retested 2 months later, may not meet the requirements. Therefore, it is imperative that users be capable of testing their own instruments and evaluating new instruments. Testing a spirometer with a human subject and with an accurate 3-L syringe can identify most of the marginal spirometers. The steps outlined below will help determine the adequacy of a spirometer.

1. The first step in evaluating a spirometer should be

a comparison of manufacturer-stated specifications with the ATS recommendations. It is important to observe that a 10-L spirometer which has an accuracy specification of ±3% of full scale or ±300 mL will *not* meet the ATS accuracy requirements of ±3% of reading.

2. Check for any leaks in the tubing or spirometer. This is particularly important for volume measuring devices.

3. Simulate a normal and an obstructed patient by injecting air from a 3-L syringe into the spirometer for approximately 2 seconds (normal) and 6 seconds (obstructed). Also observe if there is adequate recorder volume sensitivity. The FVCs in both instances should fall within the ATS accuracy recommendations.

4. Perform an FVC maneuver using yourself as the subject while taking particular care to achieve relatively low flow rates at the end of the maneuver. Notice if the spirometer prematurely terminates the maneuver, or if it continues to show an increase in volume when you approach your residual volume. This test is particularly important for flow measuring devices.

5. Check the start of test determination for any unusual sensitivity to noise. Some spirometers tend to falsely start timing for the FEV_1 when the subject is shaking the mouthpiece and tubing while straining to completely inhale. When this occurs, the FEV_1 will be zero or unusually low.

6. The recorder timing accuracy should be checked with a stopwatch by simply observing the time displacement over an appropriate time period. Three-liter syringes with attachments for measurement and computation of $FEF_{25\%-75\%}$ are now becoming available. These syringes provide the best method to test the entire system's volume and time accuracy.

7. The automatically determined FEV_1 should be compared with several hand-determined FEV_1s from the chart record using the back extrapolation method (on human FVC maneuvers). This comparison is necessary to ensure that the instrument is using a start-of-test determination method which is equivalent to the back extrapolation method.

8. Check to make certain that the manufacturer provides simple and complete calibration procedures, including instructions on how to correct for room air syringe volume testing if appropriate. These procedures should be simple enough for a technician to follow, and complete enough to ensure that the instrument is functioning within ATS recommendations.

These eight steps will adequately evaluate most spirometers currently manufactured. To use these methods, the manufacturer may need to provide the appropriate "multiplying" factors to compensate for the effects of using "cool, dry" air in the syringe rather than the human breath.

Sit Tight on Anger

Every business day brings misunderstandings, setbacks and disappointments that might well provoke anger. However, outright hostility is a luxury few careers can afford. While there are momentary satisfactions in expressing anger, there also are lasting, negative consequences. Doing nothing is anathema to many managers, but, sometimes, sitting tight is the best course of "action" to follow. The question is how.

It is often effective to delay your response until the heat of the moment has passed. Otherwise, you run the risk of shooting from the hip in a way you will soon regret, but never forget. A few hours may help you put terribly annoying trivia into perspective or marshall arguments to buttress your objections to more significant matters. . . .

—From "Don't Get Mad. Don't Get Even. Get Ahead!" by Marilyn Machlowitz in *The Wall Street Journal*, July 21, 1986.