A Test of the Practical Value of Estimating Breath Sound Intensity*
Breath Sounds Related to Measured Ventilatory Function

Neely E. Pardee, M.D., F.C.C.P.; C. J. Martin, M.D.; and Edward H. Morgan, M.D., F.C.C.P.

Each of four examiners performed standardized physical examinations on a group of patients who had just undergone tests of ventilatory function. The intensity of breath sounds heard with deep inspiration was graded on a rating scale of 0 to 4; the grades in six areas of the chest were added to give a total score, with possible values ranging from 0 to 24. Correlation of breath-sound scores with percentage of predicted forced expiratory volume in one second (FEV₁) was significant at the 1 percent level for all of the examiners. Differences between the examiners in their assessment of breath sounds were not statistically significant. Grading the loudness of breath sounds was a poor screening test for mild ventilatory abnormality, but normal breath sounds nearly excluded the possibility of severe reduction in the FEV₁. Definitely reduced breath-sound intensity was strong evidence for the presence of obstructive pulmonary disease.

A clinical observation is useful only if the feature observed is related in a definable way to a disease process or physiologic abnormality, if it can be described specifically enough to communicate the quality of what has been observed, and if different examiners can perceive and record it in similar ways. Medical writings from the time of Laennec¹ to the present refer to a reduction in the loudness of breath sounds as evidence for the presence of obstructive pulmonary disease. On the other hand, texts² and teaching guides³ in current use do not explain at what level of ventilatory abnormality the breath-sound intensity may be expected to be less than normal, nor do they discuss methods for observing and recording breath sounds. One article⁴ emphasizes problems in interpretation of physical signs, including quantitation of breath sounds, due to interobserver variability. Another⁵ reports lack of correlation of reduction in breath sounds with abnormality of pulmonary function. This report gives the results of an attempt to discover whether or not judgments of breath-sound intensity could, in fact, be helpful in the clinical evaluation of obstructive pulmonary syndromes.

*From the Section of Pulmonary Diseases, Department of Internal Medicine, the Mason Clinic, Seattle. Manuscript received November 17; revision accepted March 17.

Reprint requests: Dr. Pardee, 1100 Ninth Avenue, Seattle 98101

MATERIALS AND METHODS

A total of 183 patients referred to the pulmonary function laboratory made up the study population. There were 109 men and 74 women. Ages ranged from 19 to 84 years, with a mean of 54 years. Each patient was sent to one of four examiners for a standardized chest examination. Examiners were unaware of the results of ventilatory tests. No history was taken by the examiners; an examination was considered invalid if the examiner had prior knowledge of that patient.

The number of patients checked by each examiner varied from 27 to 58 subjects.

Patients were seated for the examinations. The examiners used only diaphragm stethoscopes. Auscultation was performed at the following six locations: bilaterally over the upper anterior portion of the chest, in the midaxillae, and at the posterior bases. Breath-sound intensity was estimated in inspiration during a response to the request that the subject take a deep breath with his mouth open. When a patient's inspiratory movements were not initially brisk and vigorous, the examiner urged him to take deeper, more rapid breaths until it was clear that the loudest sounds which could be generated by voluntary effort were being heard. A rating scale⁶ similar to that used by Nairn and Turner-Warwick⁷ was used to record breath-sound intensity. Definitions of the five stages on the scale, from 0 to 4, were as follows: 0, silent; 1, barely audible breath sounds; 2, faint, but definitely heard breath sounds; 3, expected for normal; and 4, louder than usual normal. Silence, scored 0, and the intensity expected for a normal subject, scored 3, were the main reference points for the scale. The numerical grades for the six areas sampled were summed to give a total breath sound score.

The association between breath-sound score and percentage of predicted forced expiratory volume in one second (FEV₁)⁸-¹⁰ was tested by calculation of linear correlation

VALUE OF ESTIMATING BREATH SOUND INTENSITY 341
coefficients.\textsuperscript{11} Prediction models for analyzing observer error and interobserver variation in scoring performance were based on elementary probability theory.\textsuperscript{12} Comparisons between the individual examiners made use of standard errors of differences between proportions.\textsuperscript{11}

RESULTS

The correlation of breath-sound scores with percentage of predicted FEV\textsubscript{1} was significant at the 1 percent level or better for all of the examiners (correlation coefficients, +0.57 to +0.68). The regression lines for the four correlations were similar in slope and in location on the plane of the coordinates (Fig 1). The two predictive models used to analyze variability of observer performance were based on the assumptions that the correlation between FEV\textsubscript{1} and breath sound intensity defined a theoretic “true” breath sound score for any given FEV\textsubscript{1}, and that error or variation in measurement of FEV\textsubscript{1} was negligible. The first model defined the scatter of data expected if breath sounds were graded randomly at one rating-scale stage below the true value, at the true value, or at one stage above the true value. The second model indicated the limits for two-stage errors. Seventy-seven percent of the observations fell within the one-stage model. Differences between the individual examiners were not statistically significant ($P > 0.05$). Over 99 percent of the observations were between the boundaries of the two-stage model (Fig 2).

Sorting the data into tabular form (Table 1) allowed calculation of a series of percentages, each of which represented a statement of probability about the percentage of predicted FEV\textsubscript{1} in the study population. When the breath sound score was 16 or

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Breath Sound Score & No. of Subjects & FEV\textsubscript{1} \(<80\text{\% of Predicted}\) & FEV\textsubscript{1} \(<40\text{\% of Predicted}\) \\
\hline
\textless 6 & 6 & 1.00 & 0.83 \\
7-9 & 16 & 0.94 & 0.69 \\
10-12 & 30 & 0.93 & 0.40 \\
13-15 & 28 & 0.71 & 0.14 \\
16-18 & 80 & 0.36 & 0.03 \\
19-21 & 14 & 0.43 & 0.07 \\
22-24 & 9 & 0.33 & 0.00 \\
\hline
\end{tabular}
\caption{Probability of Abnormal FEV\textsubscript{1} and of Severe Reduction of FEV\textsubscript{1} at Different Breath Sound Scores}
\end{table}
more, the probability was 0.63 that the FEV\textsubscript{1} would be 80 percent of predicted or more and was 0.86 that the FEV\textsubscript{1} would be at least 60 percent of predicted. There was less than one chance in 20 that the FEV\textsubscript{1} would be under 40 percent of predicted. Breath sound scores from 13 to 15 implied a probability of 0.71 that the FEV\textsubscript{1} would be abnormal, taking 80 percent of predicted as the lower limit of normal. The chance of severe reduction of the FEV\textsubscript{1} to less than 40 percent of predicted was one in 12. Scores of 12 or less indicated a 94 percent chance that ventilatory function was abnormal and a probability of 0.54 that the FEV\textsubscript{1} was severely abnormal. Very faint inspiratory sounds, with scores of 6 or less, were always associated with an abnormal FEV\textsubscript{1}; five of six such patients had FEV\textsubscript{1} values less than 40 percent of predicted.

**DISCUSSION**

The data demonstrated that use of a standardized procedure for examination could minimize interobserver differences in assessment of breath sounds. Employment of the rating scale with its five specifically defined stages may have helped the examiners make judgments of intensity more easily than would have been the case if the method of describing breath sounds had not been made explicit. Combining the ratings from the six sites examined to generate a total score for each patient also acted to reduce the effects of observer error. Calculated average error, relative to the range of possible scores, diminishes as the number of rating operations per observation increases. Overall variability in grading breath sounds was almost certainly overestimated by the predictive models, since FEV\textsubscript{1} measurements were assumed to be nonvariable. In fact, the FEV\textsubscript{1} is known to vary substantially from the value which ideally describes each person tested. In normal subjects the FEV\textsubscript{1} may be as much as 20 percent above or below the predicted figure. The scatter of points representing paired breath sound scores and FEV\textsubscript{1}; recordings around the regression line for the correlation must be due both to FEV\textsubscript{1} variation and to errors in breath sound grading, not to breath sound grading errors alone. This implies that for some patients, the breath sound score gave a better indication of ventilatory function than the FEV\textsubscript{1}.

The question of the usefulness of judging breath loudness remains. Despite the strong positive correlation, breath sounds do not predict ventilatory function with perfect accuracy. What does a clinician gain by systematic evaluation of breath sounds? Barning and Wolfe\textsuperscript{13} point out that the information value of a test result or clinical finding depends on the degree to which it changes the prior probability of the presence of some abnormal state. With obstructive airway disease, this prior probability is the prevalence of abnormal ventilatory function in some specified population. Mueller and associates\textsuperscript{14} using an FEV\textsubscript{1} less than 60 percent of the forced vital capacity (FVC) as the definition of airway obstruction, report that in the population of Glenwood Springs, Colo, the incidence of obstructive disease varies from 1 to 17 percent, depending on age and sex. Middle-aged and elderly men have abnormal values most often. These authors quote surveys with similar findings from Berlin, New Hampshire, and Harjavalta, Finland. Since 80 percent of the predicted FEV\textsubscript{1} falls between 60 and 66 percent of the FVC, this lower limit of normal is similar to the standard used in these population surveys. During the period of our study, 29 percent of patients referred for screening ventilatory studies had a flow-rate measurement less than 80 percent of predicted. These subjects were unselected, except by virtue of having presented themselves for evaluation at this medical center. In this context the information value of finding normal breath sounds is small. Breath sound scores of 16 or more change the prior probability of an abnormal value for FEV\textsubscript{1} very little and do not eliminate the possibility that a mild ventilatory defect is present. Moderate reduction in breath sound intensity, with scores of 13 to 17, is more significant. At this level the chance that the FEV\textsubscript{1} will be reduced is 2½ times that in the patient population here and more than four times that of the general public. Definitely reduced breath sounds, scored 12 or less, are almost certain evidence of ventilatory abnormality. Even allowing for errors in estimates of probability due to sample size, there is still only one chance in eight that a subject with uniformly faint breath sounds will have a normal FEV\textsubscript{1}.

Diagnosis in obstructive pulmonary syndromes is not complete when it has been decided whether or not airway obstruction is present. The hazard of early mortality and the chance that blood gas levels will be abnormal are greatest when ventilatory abnormalities are severe. In our earlier follow-up study\textsuperscript{15} of 811 patients from a tuberculosis sanatorium, the five-year mortality was 65 percent for subjects whose maximal voluntary ventilation (MVV) was less than 40 percent of predicted. Most deaths in this group were due to obstructive airway disease or to respiratory infections other than tuberculosis. When the MVV was between 40 and 79 percent of predicted, the five-year mortality fell to 35 percent. At 80 percent of predicted MVV and above, the risk of death approached that of the adult population at large. Hypercapnia did not occur in any patient whose MVV was above 80.
percent of predicted. Hypoxemia and carbon dioxide retention were present in a few individuals who had MVV measurements between 40 and 79 percent of predicted and were common when the MVV was less than 40 percent of predicted. Forty-three percent of this last severely compromised group had some degree of hypoxemia; 25 percent had both hypoxemia and elevation of arterial carbon dioxide tension. Diener and Burrows\textsuperscript{16} found a comparable strong relationship between ventilatory function and mortality. In their study, subjects who had an FEV\textsubscript{1}, greater than 1.25 L experienced a four-year mortality close to that expected for their ages on the basis of published federal statistics. When the FEV\textsubscript{1} was less than 0.75 L, the four-year survival was only 35 percent, essentially the same as the five-year survival in our patients with MVV values less than 40 percent of predicted. Normal breath sound loudness almost ruled out severe ventilatory abnormality in our study population, and breath sound scores from 13 to 15 were associated with a probability of only 0.08 that the FEV\textsubscript{1} would be less than 40 percent of predicted. There was a sharp increase in the threat of respiratory insufficiency when breath sounds were uniformly reduced, with breath sound scores of 12 or less.

This report focuses narrowly on a single physical sign. The results show that this observation can help a clinician evaluate the risks facing his patients. He need not settle for complete uncertainty about the degree of ventilatory abnormality present, even if he does not have access to a pulmonary function laboratory. Further, the rating-scale method provides a way for records of serial examinations to document changes in a patient’s condition. It is possible that combinations of several signs may have even better ability to predict ventilatory function, and that further investigation of traditional clinical clues may be of use in teaching physical diagnosis and in increasing the value of clinical observation.

ACKNOWLEDGMENTS: We thank Dr. Kenneth Voegle, Dr. John Lipani, and Mrs. Ann Taft for their help with the physical examinations. Thanks are also due to Mr. William Kosanke for his supervision of the pulmonary function tests and to Mr. Mikel Aickin for valuable statistical advice.

References

3. The Lung: Outline of Physical Diagnosis in Chest Disease. Seattle, University of Washington Medical School, revised 1974