

# Clinical Utility of Chest Auscultation in Common Pulmonary Diseases

PAUL E. BETFENCOURT, ELIZABETH A. DEL BONO, DONNA SPIEGELMAN, ELLEN HERTZMARK, and RAYMOND I. H. MURPHY, JR.

Faulkner Hospital, Pulmonary Department, Jamaica Plain, Lemuel Shattuck Hospital, Pulmonary Department, Boston, and Tufts University School of Medicine, Department of Community Health, Division of Biometry, Boston, Massachusetts

In this investigation we applied the techniques of lung sound mapping and time-expanded wave-form analysis to four common diseases that involve the lung: interstitial pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and pneumonia (Pn). Twenty subjects were studied in each group. We also studied 15 subjects without evidence of lung disease. Differences in timing, character, and location were observed, which allowed separation among these groups. Multiple logistic regression models were created and tested by the bootstrap method. Regression models correctly classified 68 and 79% of subjects. Area under the receiver operating curve ranged from 0.96 for IPF and CHF to 0.80 for COPD. We conclude that auscultatory differences exist among common pulmonary conditions and that statistical models based on auscultatory data perform well in predicting diagnostic categories.

**Bettencourt PE, Del Bono EA, Spiegelman D, Hertzmark E, Murphy RLH Jr. Clinical utility of chest auscultation in common pulmonary disease. *Am J Respir Crit Care Med* 1994;150:1291-7.**

Adventitious lung sounds are complex signals that reflect underlying pathophysiology. For this reason, the stethoscope has been used, since its invention by Laennec, to diagnose and follow the course of diseases such as bronchial asthma, congestive heart failure, and pneumonia. Forgaes (1) noted that in the early part of this century auscultation was a sophisticated art that helped guide diagnostic and therapeutic decisions, but that the advent of the chest roentgenogram and pulmonary function studies relegated chest auscultation to a "perfunctory ritual." This decline was due in part to the problems of observer variability and lack of objective recording inherent in auscultation. Because recent technologic advances have made it possible to document lung sounds objectively, we have been interested in applying computer technology to the diagnosis of chest disorders.

In this investigation we applied the techniques of lung sound mapping and time-expanded wave-form analysis (TEWA) to four common diseases that involve the lung: interstitial pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and pneumonia (Pn). We also studied subjects without evidence of lung disease (CntI). The purpose of our study was first to determine whether quantifiable differences existed in the lung sounds observed in common clinical conditions. In addition we developed statistical models from the data we collected as a first step toward automating the analysis of lung sounds. We then studied the performance of these models.

## METHODS

### Patient Population

Patients were referred to the study by physicians at three hospitals: two

acute care facilities and one chronic care institution. The project was approved by the Institutional Review Board of each hospital. Informed consent was obtained from each study subject. The diagnoses of IPF, COPD, CHF, and Pn were established using standard clinical criteria. After the collection of all data on all potential patients, the charts and radiographs were randomized, and the diagnosis for each patient was independently agreed upon by three pulmonary physicians. Specifically, the diagnosis of IPF was consistent with previously established criteria (2—5); for Pn, the CDC guidelines (3); for CHF, those of Milne and colleagues (4); for COPD, the ATS criteria (5). If there was disagreement on the diagnosis, the case was removed from the study. We selected the first 20 patients for whom there was complete agreement on the diagnosis. Control subjects were selected from a group of healthy volunteers. They were defined as subjects with no history or symptoms of significant lung disease and normal pulmonary function.

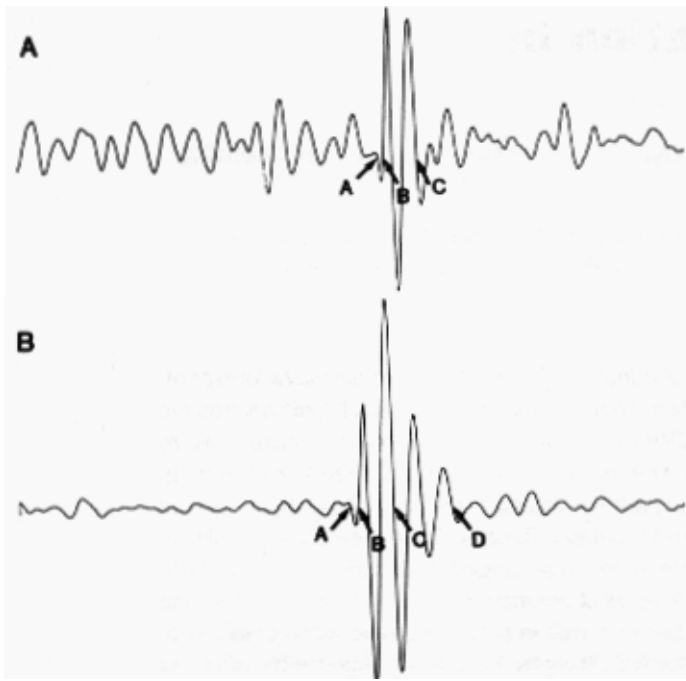
### Lung Sound Analysis

Fifty sites over each patient's chest were auscultated using a standard stethoscope to create lung sound "maps". Eight sites were auscultated anteriorly, 24 laterally, and 18 posteriorly. At each site, a trained technician classified the type of adventitious sound using a standard set of terms. With the exception of the control subjects, the technician was unaware of the subject's diagnosis. A three-point scale was used to note the "degree" of adventitious sound. For instance, 1 was used to indicate a few crackles or mild wheeze or rhonchus; 3 indicated many crackles or a loud wheeze or rhonchus. The timing of the adventitious sound was also noted as early, middle, late, or pan (throughout) inspiratory.

Tape recordings were made over sites selected for abnormal sounds for subsequent time-expanded wave-form analysis (TEWA), a computer-based method for creating visual displays of lung sounds (6). On average, two breaths at two to four sites were analyzed. A Sony Walkman Pro audio cassette recorder was used for the majority of the recordings with a Realistic electret microphone connected to the diaphragm of a Littman stethoscope chest piece by a 1-inch segment of rubber tubing. A Teac four-channel audio recorder was used for the remainder of recordings with a Sony electret ECM 30 microphone in a custom housing and air coupled to the chest wall. All recordings were band-pass-filtered at 80 to 2,000 Hz and analyzed using an A/D converter (Metabyte or Data Translation). When crackle events were noted on the display, a cursor was used

*(Received in original form January 27, 1993, and in revised form April 14, 1994)*

Correspondence and requests for reprints should be addressed to Raymond L. H. Murphy, Jr., M.D., Faulkner Hospital, Pulmonary Department, 1153 Centre Street, Jamaica Plain, MA 02130.



**Figure 1.** Parameters of crackle events. Initial deflection width (IDW) = Point A to B. Two-cycle duration (2CD) = Point A to C. Number of zero crossings (ZCS) counted from the beginning of the crackle to the point where the crackle appears indistinguishable from the baseline (Point 0 — 1B).

to mark the beginning of the event. Three parameters were then measured (Figure 1). The initial deflection width (IDW) is the time in milliseconds from the beginning of the crackle event to the first zero crossing of the baseline. The two-cycle duration (2CD) is the time in milliseconds from the beginning of the crackle event to the conclusion of two complete cycles or four zero crossings of the baseline. Finally, the total number of zero crossings in each crackle event is counted (ZCS).

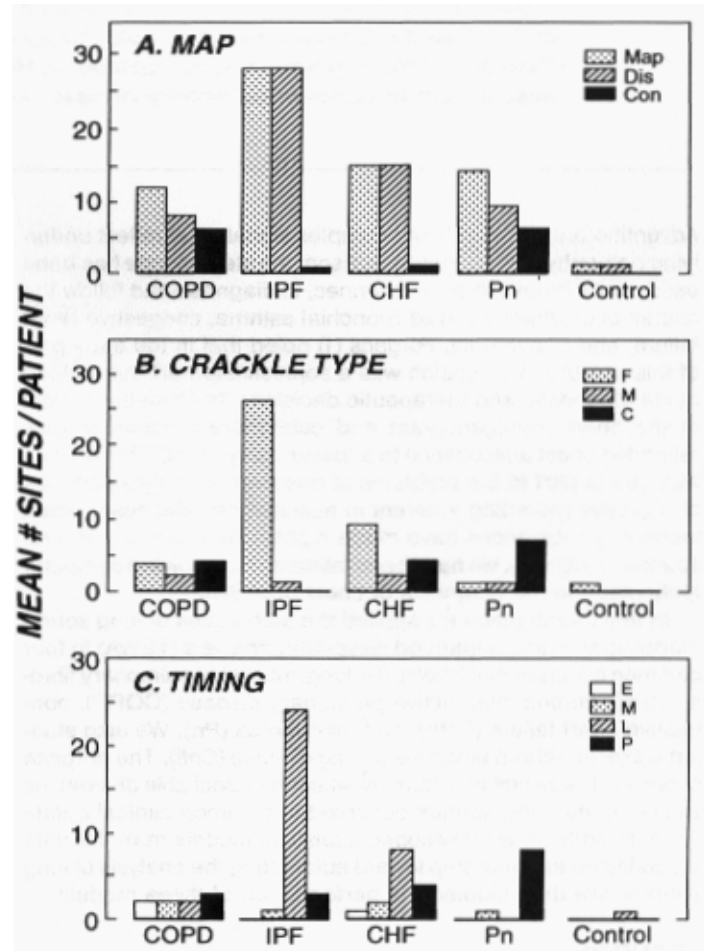
The number of sites positive for adventitious sounds was noted. Sounds were then classified as continuous (wheeze, rhonchus) and discontinuous (crackles). The discontinuous sounds were further analyzed as described in STATISTICAL ANALYSES.

### Statistical Analyses

To test for differences among the five groups, simple nonparametric analysis of various techniques (Kruskal-Wallis test) and the  $\chi^2$  test were used for continuous and categorical characteristics, respectively (7). Logistic regression models were developed to predict, separately, patient probability of being in each of the five diagnostic categories as a function of lung sound and sex and age. Two models were considered. The first (Model I) used all available variables, including TEWA data, and was applied to study participants in whom one of the four pulmonary diseases considered in this study had been diagnosed. The second model (Model II) included healthy control subjects as well as pulmonary patients with the four pulmonary diseases considered, and did not use TEWA data. Once the regression coefficients were estimated, predicted probabilities were obtained using the classification probability model given by:

$$\text{Pr (positive diagnosis for disease)} = \frac{e^{B_0 + B_1 \times \text{Variable}_1 + B_2 \times \text{Variable}_2 \dots + B_k \times \text{Variable}_k}}{1 + e^{B_0 + B_1 \times \text{Variable}_1 + B_2 \times \text{Variable}_2 \dots + B_k \times \text{Variable}_k}} \quad (7)$$

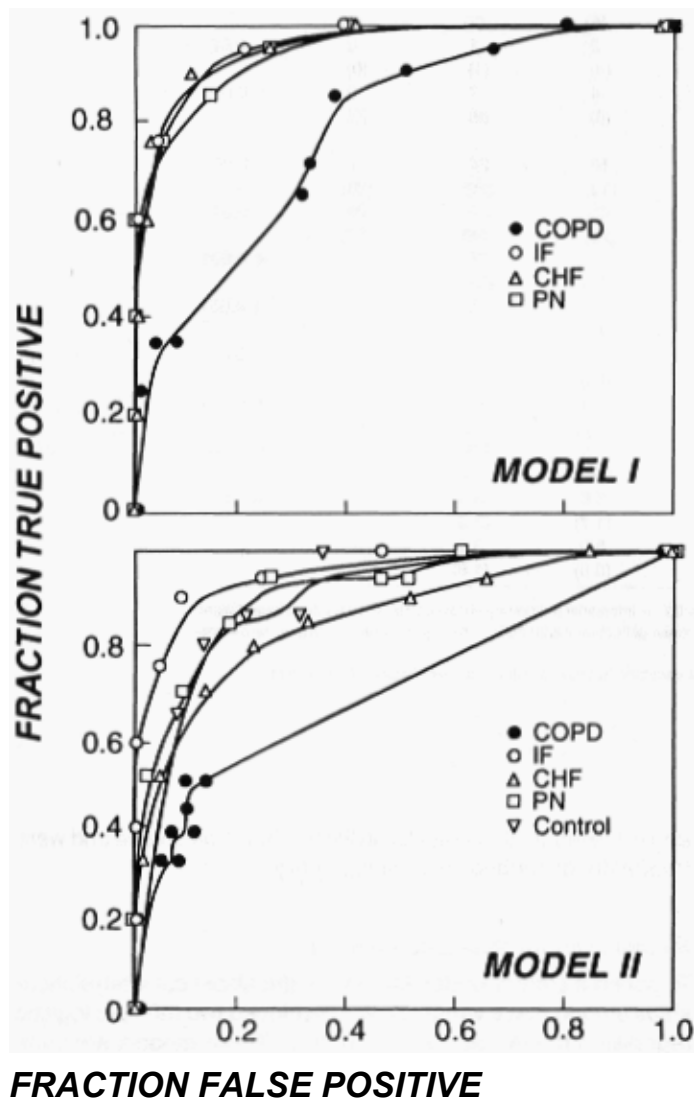
where  $B_0$  is the logistic regression intercept and for the  $k$  variables included in the model ( $\text{variable}_1, \text{variable}_2, \dots, \text{variable}_k$ ),  $B_1, B_2, \dots, B_k$  correspond to the log odds ratio of being classified in the disease group associated with a one-unit change in variables  $1 - \text{variables}(k)$ , respectively. Only statistically significant variables; i.e., those with estimated regression coefficients that were associated with  $p$  values 0.05, were included in the models. In Table 5, patients were classified into the corresponding diagnostic entity if their predicted probability for disease, as given in general form above and with coefficients as presented in Table 3, was greater than 0.50. Although this 50% cutoff value is arbitrary and may not provide the optimal classification rule in any case, predicted classifications derived from different models for the same outcome can be validly compared with this scheme. To eliminate any dependency on arbitrary cutoffs for classification, receiver-operating characteristic (ROC) curves were also used to evaluate the classificatory accuracy of each model (8, 9). Empirical ROC curves were generated for each diagnostic group using both the Model



**Figure 2.** A. Comparison among five diagnostic groups with respect to the number of sites positive out of 50 "map" sites for (1) any type of adventitious sound (map), (2) discontinuous adventitious sounds (dis), and (3) continuous adventitious sounds (con). B. Comparison among five diagnostic groups with respect to the number of sites positive out of 50 "map" sites for fine (F), medium (M), and coarse (C) crackles. C. Comparison among five diagnostic groups with respect to the number of sites positive out of 50 "map" sites for early (E), middle (M), late (L), or pan (P) (out) crackles.

I (Figure 2) and the Model II (Figure 3) modeling strategies. The cutoff values that generate the points are chosen so that the sum of the true positive rate and false positive rate is approximately constant. This criterion causes the vertical distance plus the horizontal distance between adjacent operating points to be approximately constant on linear axes, thereby providing a roughly uniform spread of operating points with points slightly closer where curvature of the ROC curve is greater. Other cutoffs could have been chosen, but these give as good a picture of the accuracy of the models as any for a subjective visual assessment as presented in the figures. For a quantitative assessment of accuracy, the nonparametric area under the ROC curve (W) is used, and results are presented in Table 4. This statistic is independent of the choice of arbitrary cutoff values and makes no distributional assumptions.

In addition, a simultaneous diagnostic prediction rule was applied in which each patient was assigned to only one of the groups. Here, the patient was considered positive for the disease category for which the patient had the highest predicted probability. To simulate the performance of the simultaneous diagnostic prediction rules among data other than the "training sample" from which the rules were derived, we used the boot-



**Figure 3.** Receiver-operating characteristic (ROC) curves for Models I and II. For each disease category, the abscissa of the curve is the fraction of patients in the disease category who are false positives. The ordinate is the fraction of patients in the disease category who have been correctly classified. Each point on the curve corresponds to a different cutoff value for classification (not shown). Note: The perfect classification rule has a ROC curve that follows the left and top borders of the graph.

strap with a "one-out" rule (10). Four hundred data sets of 80 (Model I) or 95 (Model II) subjects each were created by randomly sampling with replacement from the original data. Each "subject" in these new data sets was run through the simultaneous diagnostic prediction rules created anew with all data in the current set except that subject. Percent correctly classified and sensitivities for each of the diagnostic categories were recorded, and the median and upper and lower 25 percentiles of these were reported.

Two composite variables were created and used in all analyses. (1) The crackle "roughness" score was a weighted average of the number of crackles observed of each type where weights were assigned so that

0 = fine and 4 = coarse. A high score meant that the patient's crackles were on the average more coarse than those of a patient with a lower score. (2) The event time score was a weighted average of the number of crackles observed at each interval in the respiratory cycle, where weights were assigned so that 0 = early and 4 = late; a high score meaning that the crackles, on average, tended to be later in the respiratory cycle than those of a patient with a lower score. Paninspiratory crackles were arbitrarily assigned an event time score of 0. When number of paninspiratory crackles is included as a model covariate, this does not present any difficulty of interpretation since event time scores are effectively compared within strata of type of crackle.

## RESULTS

Ninety-five subjects were selected for study. There were 15 control subjects and 20 patients in each of the four diagnostic groups: IPF, COPD, CHF, and Pn. Testing for differences among the groups revealed that for most of the individual auscultatory acoustic characteristics listed in Table 1 there were significant differences among groups. Patients with CHF tended to be older, whereas patients with Pn and the control subjects were younger. The sexes were evenly distributed.

### Comparison of All Diagnostic Groups

Patterns, or constellations, of abnormal sounds differed among these distinct disease categories. This was most striking in the case of PR

The lung sound maps in the patients with IPF revealed many sites positive for the presence of crackles (Figure 2A). Not only were there many more sites positive for crackles but they extended higher on the chest than in patients with the other diagnoses. The crackles of IPF tended to be described as fine and occurred later in the inspiratory cycle (Figure 2B and C). When these crackles were studied by TEWA they were characterized by shorter IDWs and 2CDs (Table 1).

Patients with Pn tended to have more coarse crackles and a lower event time score, and more of their crackles were paninspiratory (Figure 2B and C). Patients with CHF tended to have crackles with larger IDWs and 2CDs (Table 1).

There were also other pattern differences among the groups. The presence of continuous sounds (wheezes, rhonchi), especially those in the upper chest and right hemithorax, distinguished patients with Pn and COPD from patients with IPF and CHF (Table 1).

### Comparison between Pairs of Diagnostic Groups

Whether differences existed between the groups can be seen in Table 1, but not where these differences occurred. In Table 2 we studied several pairs of comparisons, i.e., IPF versus CHF, IPF versus COPD and COPD versus Pn. Defining statistical significance as a p value <0.05, we found that of the 21 lung sound characteristics presented in Table 2 there were significant differences in 11 between IPF and CHF, in 15 between IPF and COPD, and in seven between COPD and Pn.

Because IPF and CHF are diagnoses with similar radiographic features we compared the characteristics that helped to separate

TABLE 1  
BASIC CHARACTERISTICS OF STUDY POPULATION

| Variable                              | COPD<br>(n = 20) | IPF<br>(n = 20) | CHF<br>(n = 20) | Pn<br>(n = 20) | Cntl<br>(n = 15) | p Value* |
|---------------------------------------|------------------|-----------------|-----------------|----------------|------------------|----------|
| Mean age                              | 62 (9)           | 68 (9)          | 78 (12)         | 55 (22)        | 49 (11)          | ≤ 0.001  |
| Sex, % female                         | 55               | 35              | 65              | 35             | 27               | 0.10     |
| Mean number of sites positive/50 for: |                  |                 |                 |                |                  |          |
| Any adventitious sound                | 12<br>(14)       | 28<br>(14)      | 15<br>(10)      | 14<br>(12)     | 1<br>(2)         | ≤ 0.001  |
| Wheezes/rhonchi                       | 6<br>(12)        | 1<br>(2)        | 1<br>(3)        | 6<br>(10)      | 0<br>(0)         | ≤ 0.001  |
| Crackles                              | 8<br>(12)        | 28<br>(13)      | 15<br>(9)       | 9<br>(10)      | 1<br>(2)         | ≤ 0.001  |
| Early inspiratory crackles            | 2<br>(6)         | 0<br>(0)        | 1<br>(3)        | 0<br>(1)       | 0<br>(0)         | 0.14     |
| Midinspiratory crackles               | 2<br>(3)         | 1<br>(3)        | 2<br>(4)        | 1<br>(3)       | 0<br>(0)         | 0.04     |
| Late inspiratory crackles             | 2<br>(4)         | 24<br>(14)      | 8<br>(6)        | 0<br>(0)       | 1<br>(2)         | ≤ 0.001  |
| Paninspiratory crackles               | 3<br>(9)         | 3<br>(8)        | 4<br>(7)        | 8<br>(9)       | 0<br>(0)         | ≤ 0.001  |
| Fine crackles                         | 3<br>(5)         | 26<br>(14)      | 9<br>(6)        | 1<br>(2)       | 1<br>(2)         | ≤ 0.001  |
| Medium crackles                       | 2<br>(4)         | 1<br>(4)        | 2<br>(4)        | 1<br>(1)       | 0<br>(0)         | 0.02     |
| Coarse crackles                       | 4<br>(9)         | 0<br>(0)        | 4<br>(8)        | 7<br>(8)       | 0<br>(0)         | ≤ 0.001  |
| Percentage of sites with:             |                  |                 |                 |                |                  |          |
| Crackles, upper chest                 | 9<br>(14)        | 32<br>(16)      | 18<br>(18)      | 24<br>(20)     | 7<br>(26)        | 0.001    |
| Crackles, right chest                 | 41<br>(42)       | 49<br>(18)      | 45<br>(22)      | 28<br>(36)     | 23<br>(37)       | 0.04     |
| Wheeze, upper chest                   | 22<br>(26)       | 8<br>(24)       | 4<br>(16)       | 25<br>(32)     | 0<br>(0)         | ≤ 0.001  |
| Wheeze, right chest                   | 34<br>(43)       | 4<br>(18)       | 1<br>(5)        | 18<br>(33)     | 0<br>(0)         | ≤ 0.001  |
| Crackle roughness score               | 1.1<br>(1.1)     | 0.2<br>(0.4)    | 1.3<br>(0.9)    | 2.2<br>(1.4)   | 0<br>(0.0)       | ≤ 0.001  |
| Event timing score                    | 2.0<br>(1.4)     | 3.0<br>(1.1)    | 2.4<br>(1.0)    | 0.5<br>(1.0)   | 1<br>(1.6)       | ≤ 0.001  |
| IDW, ms                               | 0.91<br>(.43)    | 0.65<br>(.18)   | 1.09<br>(.20)   | 0.85<br>(.40)  | —                | ≤ 0.001  |
| 2CD, ms                               | 5.4<br>(2.4)     | 4.6<br>(1.2)    | 6.6<br>(1.7)    | 5.0<br>(2.3)   | —                | ≤ 0.001  |
| ZXS                                   | 4.4<br>(2.1)     | 4.0<br>(1.0)    | 5.0<br>(0.8)    | 3.9<br>(1.8)   | —                | 0.001    |

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; IPF = interstitial pulmonary fibrosis; CHF = congestive heart failure; Pn = pneumonia; Cntl = subjects without evidence of lung disease; IDW = initial deflection width; 2CD = two-cycle duration; ZXS = zero crossings in each crackle event.

\* p value obtained from Kruskal-Wallis test for several groups means for all variables but sex; p values for sex obtained from  $\chi^2$  test. Numbers in parentheses represent standard deviation.

them. In general, patients with IPF had more sites positive in the upper portion of the chest and finer crackles occurring later in inspiration than did the crackles of CHF. TEWA reveals that the crackles of IPF have shorter IDWS, 2CDs, and fewer zero crossings than did those caused by CHF. Crackles in IPF have a lower roughness score and higher event time score.

Comparing IPF and COPD we observed that a greater number of sites for discontinuous sounds were present in the patients with IPF and a greater number of sites were positive for continuous sounds in COPD. The crackles in IPF occurred later in inspiration (also with a higher event time score), were more often described as fine, and had smaller IDWS, 2CDs, and fewer zero crossings. The crackles of COPD were observed at any time during inspiration and were more often described as coarse. The crackle roughness score in COPD was higher.

The observations regarding the comparison between COPD and Pn revealed that crackles in Pn tended to be coarse and were

frequently described as paninspiratory.

### Multiple Logistic Regression Models

To obtain a clearer understanding of the statistical interrelationships among these variables, we developed two multiple logistic regression models for each diagnosis. These models are summarized in Table 3. These statistical models included only variables that attained a significance level of  $p < 0.05$ . As noted previously, all variables listed in Table 1 were candidates for inclusion in the first model. In the second model, TEWA data were excluded so that control data could be analyzed.

ROC were constructed for each model (Table 4). The area under the ROC (W) is presented in Figure 3. This statistic has value 1 with perfect classification and value 0.5 when the classification

**TABLE 2**  
TESTS FOR DIFFERENCES BETWEEN DIAGNOSTIC GROUPS

| VARIABLE                         | p Values* for |             |            |
|----------------------------------|---------------|-------------|------------|
|                                  | IPF vs CHF    | IPF vs COPD | COPD vs Pn |
| Sex                              | 0.06          | 0.21        | 0.21       |
| Mean Age                         | 0.002         | 0.08        | 0.25       |
| Number of sites positive/50 for: |               |             |            |
| Any adventitious sounds          | 0.004         | 0.004       | 0.31       |
| Wheezes/rhonchi                  | 0.19          | 0.02        | 0.77       |
| Crackles                         | 0.002         | 0.0003      | 0.20       |
| Early inspiratory crackles       | 0.16          | 0.04        | 0.37       |
| Midinspiratory crackles          | 0.08          | 0.05        | 0.19       |
| Late inspiratory crackles        | 0.0006        | 0.0001      | 0.02       |
| Paninspiratory crackles          | 0.14          | 0.98        | 0.002      |
| Fine crackles                    | 0.0001        | 0.0001      | 0.03       |
| Medium crackles                  | 0.08          | 0.05        | 0.19       |
| Coarse crackles                  | 0.009         | 0.02        | 0.02       |
| Percentages of sites with:       |               |             |            |
| Crackles, upper chest            | 0.02          | 0.0001      | 0.02       |
| Crackles, right chest            | 0.22          | 0.30        | 0.44       |
| Wheezes, upper chest             | 0.55          | 0.03        | 0.79       |
| Wheezes, right chest             | 0.98          | 0.002       | 0.16       |
| Crackle roughness score          | 0.0001        | 0.009       | 0.007      |
| Event timing score               | 0.002         | 0.02        | 0.001      |
| IDW, ms                          | 0.0001        | 0.0003      | 0.25       |
| 2CD, ms                          | 0.0001        | 0.0002      | 0.23       |
| ZCS                              | 0.0003        | 0.006       | 0.07       |

\* For definition of abbreviations, see Table 1.

\* p value for continuous variables obtained from Wilcoxon's test, for sex obtained from  $\chi^2$  test.

system is no better than a random assignment mechanism. Model I outperforms Model II in most instances.

### Simultaneous Diagnostic Prediction Rule

We then applied the simultaneous diagnostic prediction rule, using the models developed previously, to classify each patient into the disease category for which the highest predicted probability was obtained (Table 5).

Using Model I, 90% of patients with IPF, 90% of those with Pn, 80% of those with CHF, and 55% of those with COPD were correctly classified. Using Model II, 85% of patients with IPF, 75% of those with CHF, 65% of those with Pn, and 35% of those with COPD and 87% of control subjects were correctly classified.

Model I correctly classified 79% (95% CI = 70%, 88%) of the sample, and Model II correctly classified 68% (95% CI = 59%, 77%) of the sample. Using a bootstrap procedure to simulate the performance of this system with "new" data, the results were similar: Model I correctly classified 78% (95% CI = 67%, 90%), and Model II correctly classified 68% (95% CI = 58%, 79%).

### DISCUSSION

There are two main reasons for our interest in lung sound patterns. One involves improvement in the understanding of the role of physical diagnosis in differential diagnosis and monitoring patients with chest illnesses. The second involves improvement of noninvasive diagnostic tools.

Recent advances in computer technology have made it feasible to automate the auscultation of the lung. An effort to do this, however, would require a significant investment of time and resources. Accordingly, we were interested in determining how specific the lung sound patterns were in the more common illnesses. Assuming that quantifiable pattern differences exist, the goal of computer analysis would be worth further effort.

**TABLE 3**  
VARIABLES USED IN REGRESSION MODEL I AND MODEL II

|                           | $\beta^*$ |          | p Value† |
|---------------------------|-----------|----------|----------|
|                           | Model I   | Model II |          |
| <b>COPD</b>               |           |          |          |
| Intercept                 | -0.69     |          |          |
| Discontinuous-upper chest | -4.72     |          | 0.02     |
| Continuous-right chest    | 2.11      |          | 0.02     |
| <b>IPF</b>                |           |          |          |
| Intercept                 | -2.09     |          |          |
| Discontinuous sites       | 0.19      |          | 0.0003   |
| Crackle score             | -3.76     |          | 0.003    |
| <b>CHF</b>                |           |          |          |
| Intercept                 | -32.98    |          |          |
| Age                       | 0.18      |          | 0.002    |
| Continuous-right chest    | -12.76    |          | 0.05     |
| Event time                | 1.35      |          | 0.02     |
| Initial Deflection Width  | 9.30      |          | 0.003    |
| Zero Crossing             | 1.62      |          | 0.04     |
| <b>Pn</b>                 |           |          |          |
| Intercept                 | 3.28      |          |          |
| Age                       | -0.66     |          | 0.02     |
| Discontinuous-upper chest | 7.74      |          | 0.02     |
| Event time                | -1.40     |          | 0.003    |
| Crackle score             | 1.03      |          | 0.02     |
| Discontinuous sites       | -0.12     |          | 0.04     |
| <b>Model II</b>           |           |          |          |
| <b>COPD</b>               |           |          |          |
| Intercept                 | -1.78     |          |          |
| Continuous-right chest    | 2.77      |          | 0.0008   |
| <b>IPF</b>                |           |          |          |
| Intercept                 | -2.77     |          |          |
| Discontinuous sites       | .22       |          | 0.0001   |
| Crackle score             | -3.82     |          | 0.003    |
| <b>CHF</b>                |           |          |          |
| Intercept                 | -8.96     |          |          |
| Age                       | 0.11      |          | 0.0001   |
| Continuous-upper chest    | -3.84     |          | 0.05     |
| <b>Pn</b>                 |           |          |          |
| Intercept                 | 1.20      |          |          |
| Age                       | -0.05     |          | 0.03     |
| Continuous-upper chest    | 3.29      |          | 0.05     |
| Event time                | -1.07     |          | 0.005    |
| Crackle score             | 1.10      |          | 0.0006   |
| <b>Cntl</b>               |           |          |          |
| Intercept                 | 0.11      |          |          |
| Discontinuous sites       | 0.53      |          | 0.006    |

\* For definition of abbreviations, see Table 1.

\* See Equation 1 for model definition.

† p value corresponding to the test of the null hypothesis, e.g.,  $H_0: \beta = 0$ .

Physicians have believed for some time that such patterns exist. Bronchial asthma, for example, is frequently characterized by diffuse wheezing and lobar pneumonia by localized crackles. The diagnostic values of such information is considerable. Studies such as this one, which allow us to quantify the extent to which particular lung sound patterns can be used in the detection of chest.

**TABLE 4**  
AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE (W) FOR MODELS I AND II

| Diagnostic Group | Model I |      | Model II |      |
|------------------|---------|------|----------|------|
|                  | W       | SE   | W        | SE   |
| COPD             | 0.79    | 0.07 | 0.72     | 0.07 |
| IPF              | 0.96    | 0.03 | 0.93     | 0.04 |
| CHF              | 0.96    | 0.03 | 0.87     | 0.05 |
| Pn               | 0.97    | 0.03 | 0.96     | 0.03 |
| Cntl             | -       | -    | 0.90     | 0.05 |

For definition of abbreviations, see Table 1.

TABLE 5  
CLASSIFICATION TABLES (PERCENTAGES)\*

|                 | Predicted Diagnosis |           |           |           |           |
|-----------------|---------------------|-----------|-----------|-----------|-----------|
|                 | COPD                | TPF       | CHF       | Pn        | Cntl      |
| <b>Model I</b>  |                     |           |           |           |           |
| True diagnosis  |                     |           |           |           |           |
| COPD            | <b>55</b>           | 5         | 15        | 25        |           |
| IPF             | 10                  | <b>90</b> | 0         | 0         |           |
| CHF             | 10                  | 5         | <b>80</b> | 5         |           |
| Pn              | 10                  | 0         | 0         | <b>90</b> |           |
| <b>Model II</b> |                     |           |           |           |           |
| True Diagnosis  |                     |           |           |           |           |
| COPD            | <b>35</b>           | 10        | 5         | 20        | 30        |
| IPF             | 0                   | <b>85</b> | 5         | 0         | 10        |
| CHF             | 20                  | 0         | <b>75</b> | 5         | 0         |
| Pn              | 10                  | 0         | 10        | <b>65</b> | 15        |
| Cntl            | 7                   | 0         | 7         | 0         | <b>87</b> |

\* Bold percentages indicate sensitivity of Models I and II in predicting the correct diagnosis when applied to the study sample.

illness, have received little attention. They are worth performing, however, because in addition to the above-mentioned considerations, information derived from these studies assists the clinician in situations in which more complex tests are expensive, invasive, or not readily available. In a recently reported study that examined the characteristics of crackles in a smaller group of patients Piirila and colleagues (11) were able to show that four diagnostic groups could be separated by lung sound patterns. In their study, they observed that timing and wave-form characteristics were parameters that most accurately separated their diagnostic groups (IPF, CHF, COPD, and bronchiectasis).

We believe that the following generalizations can be made regarding the conditions that we studied. First, the normal person has few, if any, adventitious sounds. We are aware that crackles are common in normal subjects after special breathing maneuvers, as has been reported by our group and others, but such maneuvers were not employed in this study (12).

Further, we can say that IPF of the severity seen in this study is a disease characterized by many sites in which discontinuous sounds are present and these sites extend to the upper zones of the lung. The crackles are identified as fine and occur late in inspiration. In this study, coarse crackles are not present. TEWA characteristics that distinguish the crackles of IPF are a short IDW, 2CD, and few zero crossings.

The crackles of Pn are almost always paninspiratory, and most sites are described as coarse as contrasted with the crackles in the other conditions. In addition, continuous (wheeze, rhonchus) adventitious sounds are relatively common in Pn.

The crackles of CHF, like those of IPF, are most often described as fine, although medium and coarse crackles are encountered. They occur at all times during inspiration, but they tend to be most numerous late in inspiration or are paninspiratory. These crackles occur predominantly at the bases. They have the longest IDW and 2CD and the greatest number of zero crossings.

COPD is characterized by the presence of continuous sounds (wheezes, rhonchi) and a few crackles that may be fine, medium, or coarse and that occur at any time during inspiration.

Our finding of a predominance of fine crackles in IPF when compared with that in the other diseases we studied is consistent with the predominantly peripheral distribution of IPF. Recently, Al Jarad and colleagues (13) demonstrated that time expanded wave-form analysis performed as well as high resolution CT in the detection of asbestosis. Pn, CHF, and COPD are more likely to affect larger airways, and, accordingly, our finding of more coarse

crackles in these disease states is not surprising. According to the more commonly accepted theories (14, 15), crackles are associated with airway/air-space opening: coarser crackles are associated with larger airways than are fine crackles.

Our multiple logistic regression models performed very well. Model I correctly classified 79% of our population, and Model II correctly classified 68% of the patients. We wished to understand whether the exclusion of the TEWA or the addition of the control subjects was the factor responsible for the poorer performance of Model I relative to Model II. We investigated this issue empirically with our data as follows. First, we excluded the TEWA data from the pool of candidate variables and redeveloped a model from the remaining available data among the same 80 subjects as used in the reported Model I results. We then compared the performance of this model with the performance of Model II among these same subjects. We found that this new model, which used demographic data and chest auscultation data alone, performed nearly as well as Model I, as reported in Table 3. This may not be surprising, as we found in Model I that TEWA was useful only for classifying CHF and Pn. It thus appears that the addition of the control group, and not the deletion of the TEWA data, is responsible for the proper performance of Model II.

The observations made by a technician using a stethoscope distinguished the disease categories more readily than did the measurements of crackle characteristics performed by TEWA. The superior performance of the observations made by the human observer may have been influenced by the fact that observations were made at 50 sites rather than at only two to four sites by TEWA. TEWA did, however, show significant differences among these groups of patients (Tables 1 and 2).

The reliance on auscultatory findings rather than on TEWA was necessary, as at the time of the study it was expensive and time-consuming to record data from 50 sites per subject and evaluate them by computer analysis. In the future, this should be easier and less expensive. Indeed, in a number of recent studies in which careful attention has been paid to criteria for auscultation, observer variability has been reported to be small (16, 17).

The rigorous method we employed wherein a technician recorded observations systematically at 50 sites demonstrated that auscultation can be a powerful diagnostic modality. The performance of the technician was evaluated by use of tape recordings and simultaneous listening by a physician.

We are aware that the best way to evaluate the performance of our models is to apply them in a prospective way to a new group of patients. However, in this study we used a statistical procedure called bootstrapping, which allowed us to simulate a scenario similar to a prospective study. Using this method the percent correctly classified by the two models did not change.

There are further limitations to our study, concerning both internal validity as well as external generalizability. To evaluate the utility of our procedure under the most homogenous circumstances, we wished to select patients in whom diagnoses were clear-cut. Because patients in whom diagnoses are clear-cut are not always typical of the general pulmonary patient population, this choice of subjects limits the external generalizability of the current study. Conversely, we cannot exclude coexisting occult bronchitis among those patients with IPF, CHF, and Pn in our study who were smokers. Thus, among these patients, the diagnoses may not be as clear-cut as we assumed, thereby limiting our ability to evaluate the utility of our procedure to distinguish between patients with the diagnostic categories considered. However, be-

cause this is a problem common in clinical practice, including the smokers in our analysis increases the external generalizability of our findings. In addition, there were age differences among the groups. Our patients with Pn were on average younger than those with CHF. This reflects the characteristics of our hospital-based population.

In one sense, it is artificial to test lung sound information in the absence of other clinical information for diagnosis. In the clinical setting, diagnoses are made by assessing all of the available information. If one considered using any of the other diagnostic modalities alone in attempting to make the separation of these four diagnostic groups, it would be clear that they would also have limitations when used in isolation. For example, pulmonary function studies are helpful in detecting the patients with obstructive lung disease, and they would be necessary to detect the restrictive component of IPF, but they would be of very little value in the diagnosis of Pn and CHF. The chest roentgenogram likely would be of limited use in the detection of COPD, but it is no doubt the single most valuable tool in the detection of the other diseases. One might argue that it is unimportant to consider lung sounds in diagnosis as one could obtain chest radiographs and pulmonary function studies and achieve adequately diagnoses in each of these categories. On the other hand, radiographs carry potential risk, are expensive, are often refused by patients, and cannot be done as frequently as auscultation. Indeed, in the intensive care unit where portable chest roentgenograms are common, basilar areas of the lung are not well visualized, and Pn in these locations may not be detected. Furthermore, although chest radio-graphs may be obtained daily, or even more frequently in seriously ill patients with CHF, it is theoretically possible to monitor lung sounds continuously.

We have included results observed from ROC curves. The area under the ROC (W) is a statistical tool that attempts to provide a common measure by which comparison can be made of the performance of different diagnostic tests (8). By focusing on the W data it becomes easier to study the performance of the models and allows comparison with other diagnostic tests. Swets (9) reported that chest roentgenograms and CT scans of the brain performed to detect any lesion have W values of about 0.98 to 0.99, Pap smears for the detection of cervical cancer have W values of about 0.82 to 0.87, and mammography for the detection of malignancy at its best is 0.91. In this study, we have observed that auscultation of the chest alone in the worst case has a W of 0.80 (slightly below Pap smears) and at its best 0.96, which compares favorably with chest radiography and CT scanning of the head. Although we do not believe that a specific diagnosis should be made only by auscultation of the chest, one can see that it is indeed a

very powerful tool. We believe that further efforts to refine and automate auscultation will be worthwhile.

## References

1. Forgacs P. Lung Sounds. London: Casel and Colier MacMillan Publishers Ltd., 1978.
2. Fulmer JD, Roberts WC, Von Galer ER, Crystal RG. Morphologic-physiologic correlates of the severity of fibrosis and degree of cellularity in idiopathic pulmonary fibrosis. *J Clin Invest* 1979;63:665—76.
3. Garner JS, Jarvis WA, Emori TG, Horan TC, Hughes JM. CDC definitions for nosocomial infections, 1988. *Am J Infect Control* 1988; 16:128—40.
4. Milne ENC, Pistolest M, Miniati M, Giuntini C. The radiologic distinction of cardiogenic and noncardiogenic edema. *MR* 1985;144:879—94.
5. Division of Lung Disease. National Institutes of Health. Report of task force on epidemiology of respiratory diseases, October 1980 (NIH publication no. 81-2019).
6. Murphy RLH, Holford SK, Knowler WC. Visual lung-sound characteristics by time-expanded wave-form analysis. *N Engl J Med* 1977;296: 968—71.
7. Rosner B. Fundamentals of biostatistics. Boston: Duxbury Press, 1986.
8. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagn Radiol* 1982;143:29—36.
9. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285—93.
10. Efron B, Tibshirani A. Statistical data analysis in the computer age. *Science* 1991;253:1176—83.
11. Piirila P, Sorijarvi ARA, Kaisia I, Rajala HM, Katila T. Crackles in patients with fibrosing alveolitis, bronchiectasis, COPD, and heart failure. *Chest* 1991;99:1076—83.
12. Workum P, Holford SK, Del Bono EA, Murphy RLH. The prevalence and character of crackles (rales) in young women without significant lung disease. *Am Rev Respir Dis* 1982;126:921—3.
13. Al Jarad N, Stickland B, Botharney G, Lock S, Logan Sinclair R, Rudd RM. Diagnosis of asbestosis by a time expanded waveform analysis, auscultation and high resolution computed tomography: a comparative study. *Thorax* 1993;48:347—53.
14. Forgacs P. Crackles and wheezes. *Lancet* 1967;2:203—5.
15. FredbergJJ, Holford SK. Discrete lung sounds: crackles (rales) as stress-relaxation quadrupoles. *J Acoust Soc Am* 1983;73:1036—46.
16. Murphy RLH, Gaensler EA, Holford SK, Del Bono EA, Epler GA. Crackles in the early detection of asbestosis. *Am Rev Respir Dis* 1984;129:375—9.
17. Workum P, Del Bono EA, Holford SK, Murphy RLH. Observer agreement, chest auscultation, and crackles in asbestos-exposed workers. *Chest* 1986;89:27—9.