Determination of Clinical Efficacy: Nuclear Medicine as Applied to Lung Scanning

Eugene L. Saenger, C. Ralph Buncher, Bonny L. Specker, and Ruth A. McDevitt

Eugene L. Saenger Radioisotope Laboratory, Department of Radiology, University of Cincinnati Medical Center, Cincinnati, Ohio

This paper describes a Society of Nuclear Medicine sponsored study of 2,023 patients which compares two methods, logistic regression (LR) and entropy minimax pattern detection (EMPD), to evaluate efficacy. Lung scans, used in determining or excluding a diagnosis of pulmonary embolism (PE), were utilized to create the data set. The LR analysis, presented here, shows that lung scan findings have significant influence on the referring physician's diagnostic thinking. Models were developed for the probability of a signout diagnosis of PE, and equal patient groups tested the validity of these regression equations. Individual models developed on each patient group yielded similar results. This analysis shows that the lung scan results affect the therapeutic management of the patients in a beneficial direction. A comparison of the sensitivity, specificity, and predictive values of EMPD and LR was done. EMPD predicts a signout diagnosis on only 41% of cases before lung scan and 71% after lung scan; LR provides a prediction of the signout diagnosis on 100% of cases. An advantage of EMPD is that it does not require prior probability estimates. However, LR uses this estimate, thus incorporating intuitive knowledge not evaluated by EMPD.

J Nucl Med 26:793–806, 1985

Widely accepted methodologies of measuring the efficacy of a medical procedure have not yet been established. One of the few large scale prospective studies to date was the American College of Radiology (*1*) investigation of the use of diagnostic radiographs in the hospital emergency room. Several studies, usually retrospective, have used a single method to assess efficacy and analyze a few hundred cases or tests (*2–6*). This current prospective study was undertaken in an effort to find one or more methods by which efficacy of clinical diagnostic procedures could be quantified. The study design and methodology was the result of the collaborative efforts of the members of The Society of Nuclear Medicine Committee on Public Health and Efficacy. It was intended to sample the community practice of medicine as best as it could be done.

The design of a comparison of two analytic techniques for determining efficacy was achieved by using logistic regression analysis as one alternative reported in this paper, and entropy minimax pattern detection (EMPD) described in the accompanying report as the other method (*7*). While the primary goal of the study was to compare methods of measuring efficacy, we also offer some comments on clinical applications of a nuclear medicine diagnostic test as it is used routinely by clinicians.

Planning and data collection were begun in 1977 under the sponsorship of The Society of Nuclear Medicine (SNM). The clinical tests chosen for analysis were diagnostic nuclear medicine procedures as requested by referring physicians in their routine medical practice. The design of the study was such that patient attributes, diagnoses and estimates of probabilities of the referring physician were obtained prior to and following the nuclear medicine study. The nuclear medicine findings were those recorded on the patients' charts, i.e., there was no reinterpretation or further analysis of these findings by nuclear medicine physicians or others.

Two definitions of efficacy utilized in the American College of Radiology study were adapted for use in this study as follows.

Received Aug. 10, 1984; revision accepted Mar. 18, 1985.

For reprints contact: Eugene L. Saenger, MD, Eugene L. Saenger Radioisotope Laboratory, Mail Location #577, University of Cincinnati Hospitals, Cincinnati, OH 45267.

Efficacy 1

A diagnostic procedure is efficacious if, and only if, it influences the physician's diagnostic thinking as measured by a change in the likelihood that a patient has the disease of interest.

Efficacy 2

A diagnostic procedure is efficacious if, and only if, the result can be shown to have an important probability of affecting the management of the patient's disease.

SELECTION OF NUCLEAR MEDICINE TEST

The test selected for study was that of pulmonary imaging (perfusion/ventilation) which is used principally for aiding or excluding the diagnosis of pulmonary embolism (PE). This test was suitable for this study since the progress of this condition and its outcome occur relatively rapidly and are subject to clinical confirmation, thus producing the necessary information in days or weeks.

In addition, the incidence of PE was known to be great enough that a sufficient number of cases could be accumulated in a reasonable length of time. The incidence of PE is estimated to be 630,000 per year in the U.S. population (*8*). There is a 92% survival rate among those treated with anticoagulants; the survival rate for untreated PE has been estimated to be 70% (*9*).

The criteria for making or excluding the diagnosis of PE utilizing lung scanning have been reviewed (*10,11*). These nuclear medicine criteria are based on a number of combinations of scintigraphic defects, correlations of abnormal lung scan patterns with x-ray findings and clinical manifestations and as such are not evaluated in this study. Instead, our interest is in the action taken by the referring physician as affected by the report of the nuclear medicine physician.

METHODS

Study design

The data base that was accumulated included cases from medical centers throughout the United States, with the distribution consisting of community hospitals, university medical centers, and government hospitals (*12*). These institutions varied in size and service facilities as well as geographic location, population base, and relationship to medical schools or teaching programs such that the sample encompassed what appears to reflect medical practice in the United States during the time frame of the data collection. With this distribution pattern, the data collected could be expected to characterize the use of lung scanning procedures as diagnostic tools in community as well as university hospitals by referring physicians engaged in clinical practices and settings (*13*).

To develop the data base, the physician members of the SNM (about 4,000) were queried as to their interest in participating. From 450 responses, 26 institutions were selected based on the above sampling criteria and the reporting physicians agreement to participate. They joined the study between June 1978 and October 1979. Subsequently, four hospitals resigned because of lack of cases for the study. The remaining 22 hospitals ranged in size from 130 to 1,200 beds. Hospitals with fewer than 100 beds were omitted since they represent less than 19% of the short-term general hospitals having nuclear medicine facilities in the US in 1977, and the number of lung scans done annually as reported by the responding hospitals was insufficient to produce cases at the rate we hoped to achieve. Twenty of the institutions were nonfederal, general, community hospitals; two were Veterans Administration hospitals; and one specialized in heart and lung disease. All were accredited by the Joint Commission for Accreditation of Hospitals. There were 20 with approved residency training programs, five associated with medical schools; 15 were affiliated with medical schools; 14 were members of the Council of Teaching Hospitals. There were 21 hospitals with general medical and surgical services. All hospitals were nonprofit. The geographic regions of the continental United States were represented with one exception, the east south-central states. Repeated attempts to enlist a hospital in this region were unsuccessful. The distribution and demographic characteristics of the participating hospitals were found to be reasonably matched to the profile of clinical practice settings.

Information was obtained about the equipment, radiopharmaceuticals and techniques utilized in performing perfusion and ventilation studies, and the sequencing of these tests with chest radiograph of patients. No modification in the usual techniques of the cooperating hospitals was requested. There were no self-initiated changes in technique during the data collection period except for one hospital that began routine ventilation studies upon delivery of new equipment.

In order to standardize the methods of estimating probabilities for the referring physicians and of collecting data, training sessions were held for the nuclear medicine physicians and the coordinator in each institution. They, in turn, trained the referring physicians, and were responsible for overseeing case collection and verifying the accuracy of the data. Instruction manuals with detailed explanations of odds and probabilities were made available to all referring physicians to assist them in filling out questionnaires.

Initially, the nuclear medicine physician in each participating hospital presented the study to the medical staff to obtain approval and cooperation. Thus, the cases analyzed in this study were obtained from refer-

ring physicians on the basis of their willingness to fill out the forms required for data acquisition. During the period of data collection consecutive studies from participating physicians were obtained. Not all consecutive ventilation-perfusion studies at a given hospital were entered because not all referring physicians who requested such studies agreed to participate.

To eliminate a possible unrecognized pattern of bias due to the lack of sequential case entry into the study, an additional random sample of 157 cases from the same time-frame but not submitted to the study was reviewed in five of the participating hospitals. This group of cases was not used in our data base, but the distribution of diagnoses did not show a pattern different than that of the cases which were analyzed. This analysis indicates that possible selection bias through failure to obtain consecutive cases was not observed.

Questionnaire

The questionnaire utilized for data collection consisted of four parts.* Part I, to be completed by the referring physician before the lung scan was done, consisted of: (1) a check-off list of symptoms, signs, laboratory findings; (2) a statement of the one most important and the one most likely diagnosis and an estimate of the odds or probability of each; (3) the one problem or reason for ordering the lung scan; (4) a management plan based on the presenting information, (5) a statement of knowledge of the patient's medical insurance; (6) the name of the referring physician; and (7) date of the test. The most important (MI) diagnosis is defined as the diagnosis with the potential to have the most impact on the immediate welfare of the patient and his prognosis. It is the diagnosis the clinician would not want to miss. The most likely (ML) diagnosis is the diagnosis with the greatest probability out of all possible diagnoses being considered. The probability of the ML diagnosis was also restricted to be greater than or the same as the probability of the MI diagnosis.

Part II of the questionnaire required the referring physician to reassess the prior odds or probability estimates given in Part I based on the results of the lung scan, i.e., to assign posterior diagnostic probabilities. The posterior estimates could be higher, lower, or the same as the prior estimates. The physician had the option of selecting new most important and most likely diagnoses with estimates of each. The referring physician also reported the post-test management plan based on the lung scan information as recorded on the patient's chart. In this way, the diagnoses, probabilities and the management plan recorded were solely those of the attending physician. The participation of the nuclear medicine physician was limited to a recorded interpretation of the lung scan.

Part III was completed by the nuclear medicine physicians. In addition to technical factors, test dates and costs, they were asked to report their scintigraphic diagnoses in descending order of likelihood but without probability or odds estimates.

Part IV of the questionnaire, the follow-up, was completed after the patient's discharge from the hospital or outpatient care facility. It consisted of: (1) discharge diagnoses; (2) date of discharge or death; (3) whether

TABLE 1
Age Distribution by Sex and Race (N = 2,023)

Age	White Male	White Female	Nonwhite Male	Nonwhite Female
<20	17	29	12	6
20–39	102	146	40	72
40–59	255	214	61	74
60–79	363	368	57	60
80+	49	78	6	14
Totals	786	835	176	226

TABLE 2
Frequency of Presenting Indications for Lung Scan for Patients with Signout Diagnosis of Pulmonary Embolus and Other Diagnoses

	Signout diagnosis			
Sex	Pulmonary embolus			All others
Male	215			747
Female	199			862
Total	414			1,609
Patient attributes & clinical assessment	%	%	Likelihood ratio	p Value based on χ^2 statistic
Dyspnea	63.3	56.3	1.12	0.01
Chest pain	61.4	54.7	1.12	0.01
Hypoxemia (po_2 > 80)	47.3	36.7	1.29	<0.01
Abnormal chest radiograph	40.6	37.0	1.10	0.18
Tachycardia	38.7	32.1	1.21	0.01
Tachypnea	38.2	31.1	1.23	0.01
Symptoms <48 hr	34.8	29.6	1.18	0.04
Rales	26.1	24.0	1.09	0.38
Immobilization	22.0	15.4	1.43	<0.01
Thrombophlebitis	21.7	10.1	2.15	<0.01
Prior pulmonary dis.	21.3	17.2	1.23	0.06
Fever	17.4	14.8	1.18	0.19
Cough	16.2	20.5	0.79	0.05
Arrhythmia	12.8	12.0	1.07	0.65
Diaphoresis	10.6	9.1	1.17	0.33
Known malignant dis.	9.2	7.3	1.26	0.19
Accentuated P_2	8.2	5.8	1.42	0.08
Confirm change	7.3	1.7	4.32	<0.01
Syncope	6.0	4.8	1.26	0.30
Cyanosis	5.8	3.5	1.67	0.04
Shock	5.8	1.6	3.59	<0.01
Oral contraceptive	1.9	2.2	0.89	0.70
Confirm no change	1.5	1.1	1.30	0.58
Pre-op	0.5	4.2	0.11	<0.01

the status was inpatient or outpatient; and (4) pulmonary angiographic findings if this procedure was performed.

Attending physicians and residents filled out the questionnaire. Medical students were excluded from participation. Before it was used in the study, the questionnaire was pretested on 63 cases which were not used in the analysis.

RESULTS

Patient description

There were 2,023 cases contributed by 997 referring physicians; 962 patients were male; 1,621 were white. One hundred eighty-six were outpatients and 1,835 were inpatients. The distribution by age, sex, and race of the study population is shown in Table 1. About half of this population was over 50 yr of age. The distribution of presenting symptoms and signs of the patients by signout diagnosis of PE compared with all other diagnoses is shown in Table 2.

All patients underwent perfusion imaging studies with technetium-99m labeled macroaggregates; 1,646 (81.4%) also had ventilation studies with xenon-133. Pulmonary angiography was performed on 102 patients (5%); 54 were negative (53%) and 48 were positive (47%).

There were 414 patients who had signout diagnoses that included PE (20.5%), with 35 deaths (8%). There were 77 deaths (4.8%) among patients (1,609) whose signout diagnosis did not include PE. Because multiple signout diagnoses could be selected there were 2,898.

Data classification

In addition to the comparison of two methods of evaluating efficacy, it was thought desirable to test the case series from the many different centers for internal consistency. Would a model developed for predicting a signout diagnosis or management from about half of the cases be an accurate predictor in the remainder? Accordingly, the total sample was split and an initial analysis was carried out to determine a predicting function. This equation was then used to predict for the remainder of the cases.

In accordance with the above design, the cases were divided into two groups based on time of receipt of the completed case reports. All of the cases entered the study within the same time period, from July 1978 to June 1980. Group I was composed of 1,065 cases for which the collection of data was completed by January 1980; the data of the 958 cases in Group II was completed by February 1981. Since the hospitals submitted their cases sporadically, the majority contributed most of their cases either to Group I or Group II (Table 3).

TABLE 3
Distribution of Cases by Hospital

Hospital	Group I # Cases	Group II # Cases
1	123	12
2	9	8
3	0	128
4	0	3
5	9	186
6	57	54
7	57	81
8	9	0
9	0	3
10	59	104
11	119	32
12	74	58
13	190	37
14	39	22
15	75	0
16	45	12
17	1	0
18	48	5
19	85	58
20	54	87
21	12	0
22	0	68
Totals	1,065	958

The two main groups were further subclassified based on whether pulmonary embolism (PE) was considered the most important (MI) and/or most likely (ML) prior diagnosis by the referring physician. These subgroups and the number of cases in each are shown in Table 4.

The two methods, logistic regression and entropy-minimax, were used to obtain predicting equations for Group I patients. The resulting equation (logistic regression) and algorithm (entropy-minimax) were then compared for their abilities to predict outcome in Group II. This analytic method was felt to be an unusually rigorous test of the ability of models derived from one data set to predict on a second data set since it involves random variation, hospital differences, and time trends in the data.

Logistic regression analysis

Logistic regression techniques were used to determine whether a statistical model could be developed to predict the probability of a signout diagnosis of PE and to determine the extent to which the lung scan influenced the probability of a signout diagnosis of PE (Efficacy definition #1). Included in these analyses were the referring physician's estimate of prior probability of the most important and most likely diagnoses, and patient attributes (Table 2). Logistic regression analysis was preferred over discriminant analysis because of its statistical robustness with both discrete and continuous data (*14*).

The patient attributes of age, sex, and race (Table 1), height and weight, the 24 symptoms and signs (Table

2), and the referring physician's prior probability estimates of the most important and most likely diagnoses were screened in a stepwise fashion to determine which variables could help predict a signout diagnosis of PE. The percentages of the statistically significant variables by signout diagnosis are given in Tables 4 and 5 for the PE/PE and PE/OTHER subgroups.

For both Groups I and II, a model was developed for each of the two subgroups of PE/PE and PE/OTHER (MI/ML) (Table 6) based on data obtained prior to the lung scan (Figures 1A and 1B). The results of the lung scan (PE or OTHER) were then added to the models and new estimates of the coefficients were obtained. The models thus developed were tested on the data from which they were derived, and then applied to data from the other group.

The models developed on the two subgroups (PE/PE and PE/OTHER) of the Group I data predicted well on the Group II data as illustrated by the sensitivities, specificities, predictive values, and accuracies shown in Tables 7 and 8. High predictive values were also observed when the models for the Group II data were applied to Group I data (not shown). A model for the OTHER/PE subgroups was not developed because of the small numbers of patients in them. The subgroup OTHER/OTHER was not used in the analysis since there was no estimate of prior probability of PE available.

Figure 1 shows the lines for the equations based on Group I, subgroup PE/PE data. The model for that subgroup included the presence or absence of thrombophlebitis and the referring physician's prior probabil-

TABLE 4
Percentage of Significant Variables by Signout Diagnosis, Subgroup PE/PE

Signout diagnosis:	Group I		Group II	
	PE	Other	PE	Other
Sex*				
Male	55.7%	51.9%	53.7%	36.0%
Female	44.3	48.1	46.3	64.0
Thrombophlebitis†				
No	72.9	88.1	79.0	87.3
Yes	27.1	11.9	21.0	12.7
Diaphoresis*				
No	11.4	11.9	91.6	82.0
Yes	88.6	88.1	8.4	18.0
Probability of most important before scan diagnosis†				
<10%	0.0	1.5	1.0	1.3
10–49	2.2	6.7	2.1	12.7
50–89	60.0	78.5	64.2	71.4
90–98	23.5	11.1	23.2	10.6
≥99	14.3	2.2	9.5	4.0
Probability of most likely before scan diagnosis*				
<10%	0.0	1.5	1.1	1.3
10–49	2.1	6.7	1.1	11.3
50–89	60.0	77.0	63.1	72.0
90–98	22.9	12.6	24.2	10.7
≥99	15.0	2.2	10.5	4.7
Nuclear medicine diagnosis*,†				
PE	84.2	13.3	75.5	8.0
Other	15.8	86.7	24.5	92.0

* Variable significant for Group II model.
† Variable significant for Group I model.

TABLE 5
Percentage of Significant Variables by Signout Diagnosis, Subgroup PE/Other

Signout diagnosis:	Group I		Group II	
	PE	Other	PE	Other
Age*				
<20	3.4%	1.4%	1.3%	3.4%
20–39	15.9	18.1	13.2	20.3
40–59	21.6	32.1	26.3	30.4
60–79	46.6	40.2	50.0	39.3
>80	12.5	8.2	9.2	6.6
Thrombophlebitis†				
No	79.6	88.8	85.5	87.9
Yes	20.4	11.2	14.5	12.1
Rales*				
No	75.0	77.7	61.8	75.8
Yes	25.0	22.3	38.2	24.2
Immobilization*				
No	77.3	86.1	90.8	82.9
Yes	22.7	13.9	9.2	17.1
Known malignant disease*				
No	90.9	95.8	86.8	93.4
Yes	9.1	4.2	13.2	6.6
Most important before scan diagnosis*,†				
<10	10.2	23.4	7.9	13.5
10–49	61.4	60.5	50.0	59.8
50–89	27.3	15.9	2.6	0.2
90–98	0.0	0.2	2.6	0.2
≥99	1.1	0.0	0.0	0.0
Nuclear medicine diagnosis*,†				
PE	77.5	3.8	65.8	5.0
Other	22.5	96.2	34.2	95.0

* Variable significant for Group II model.
† Variable significant for Group I model.

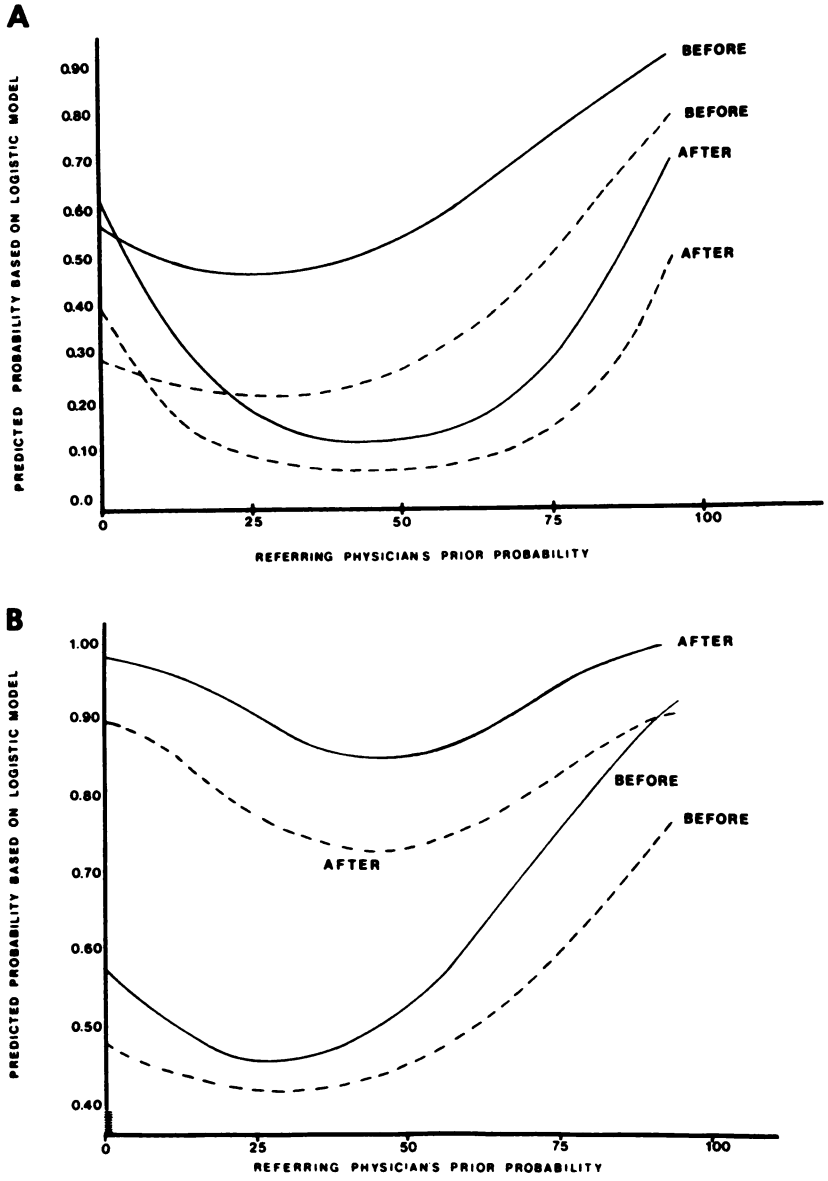

FIGURE 1
A: Prediction of probability of signout diagnosis of PE based on referring physician's prior probability of PE when lung scan is reported as *negative.* (—) With thrombophlebitis; (- - -) Without thrombophlebitis. Curves modeling before lung scan and after lung scan equations: log p(PE)/1 − p(PE) = −0.863 + 1.134 (thrombophlebitis) − 0.030 (prior probability) + 0.0005 (prior probability)^2 and log p(PE)/1 − p(PE) = −0.428 + 0.877 (thrombophlebitis) − 0.103 (prior probability) + 0.001 (prior probability)^2 + 3.678 (lung scan), respectively. B: Prediction of probability of signout diagnosis of PE based on referring physician's prior probability of PE when lung scan is reported as *positive.* Curves modeling before lung scan and after lung scan equations: log p(PE)/1 − p(PE) = −0.863 + 1.134 (thrombophlebitis) − 0.030 (prior probability) + 0.0005 (prior probability)^2 and log p(PE)/1 − p(PE) = −0.428 + 0.877 (thrombophlebitis) − 0.103 (prior probability) + 0.001 (prior probability)^2 + 3.678 (lung scan), respectively. These equations were based on Group I subgroup PE/PE data

ity. Since the lung scan could have been either negative or positive and thrombophlebitis could have been either present or absent, there were four possible combinations: (1) no thrombophlebitis, test negative; (2) no thrombophlebitis, test positive; (3) thrombophlebitis present, test negative; and (4) thrombophlebitis present, test positive. The impact of the lung scan result on the predicted probabilities of PE is seen when the curve for the equation before the test is compared with the curve for the equation after the test. When the lung scan report was negative for PE, the probability of a signout diagnosis of PE was lower than before the test at all levels of the referring physician's prior probabilities (Fig. 1). This observation is true whether thrombophlebitis is present or absent.

The models developed using only data prior to the nuclear medicine test were not able to predict satisfactorily a signout diagnosis of PE. After the addition of lung scan results, there was a definite improvement in the ability of the regression model to predict a signout diagnosis of PE. Figures 2 and 3 illustrate these findings. Before the test result is available the cases are not clearly separated into those without PE and those with PE. However, when the result of the lung scan is added, there is a clearly recognizable separation of the cases without PE and with PE. This change occurs not only within the data set from which the model was developed but also when the same model is applied to the other data set. For example, Figs. 2A and 2B illustrate the ability of the Group I models, subgroup PE/PE, to predict PE before and after the lung scan, respectively. Figures 2C and 2D illustrate how well the same models, applied to the Group II data, predict the probability of a signout diagnosis of PE both with and without the lung scan. Figures 3A, B, C, and D illustrate similar findings in the PE/OTHER subgroup. Similar results, although

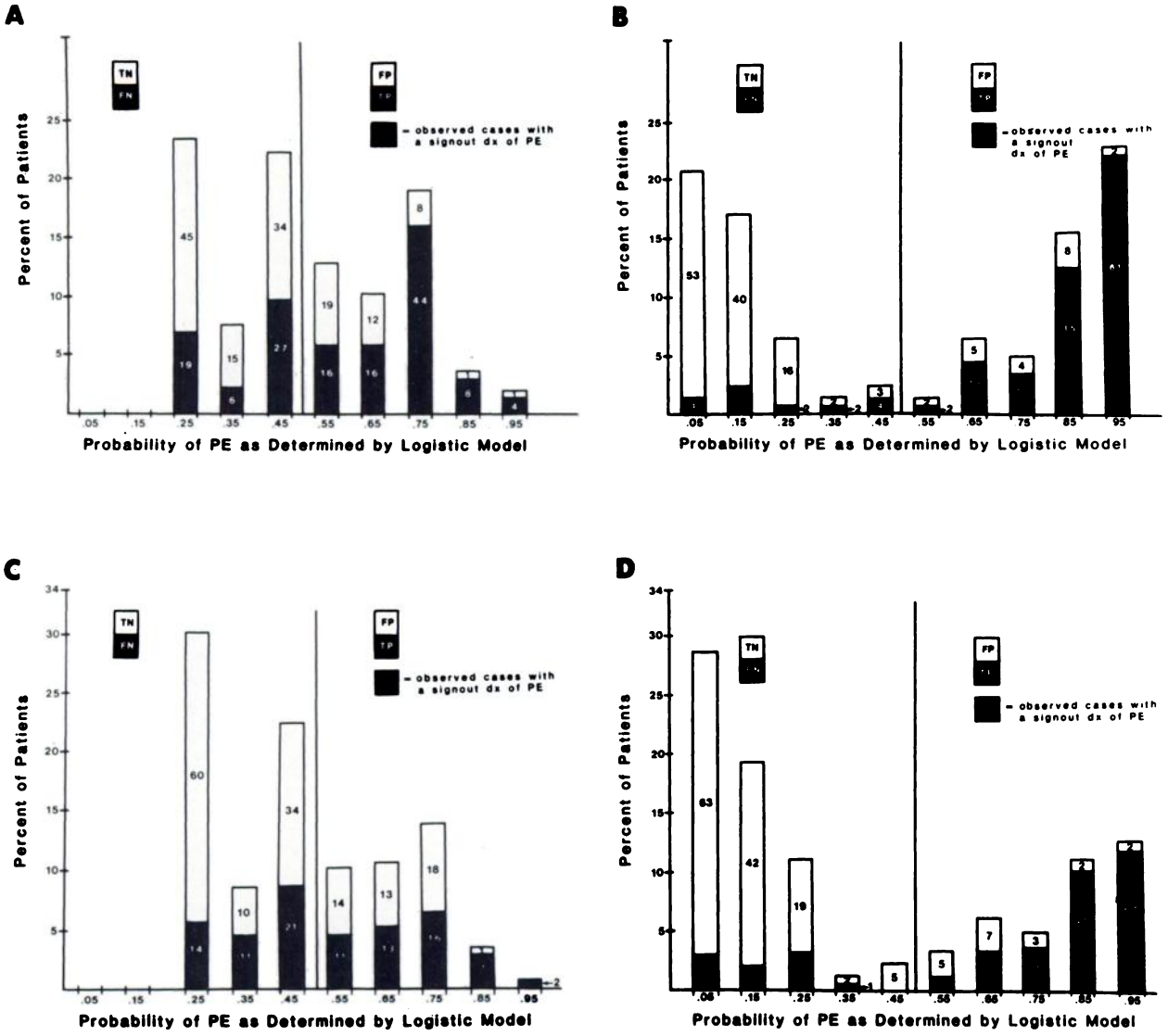

FIGURE 2
A: Group I model predicting Group I data prior to lung scan, subgroup PE/PE, n = 275. B: Group I model predicting Group I data after lung scan, subgroup PE/PE, n = 275. C: Group I model predicting Group II data prior to lung scan, n = 245. D: Group I model predicting Group II data after lung scan, n = 245

TABLE 6
Classification of Patients by Pulmonary Embolus (PE) Most Important/Most Likely (MI/ML) Prior to Lung Scan

MI/ML	Group I	Group II
PE/PE: Pulmonary embolism both MI and ML	275	245
PE/Other: Pulmonary embolism MI but *not* ML	640	514
Other/PE: Pulmonary embolism *not* MI but ML	15	7
Other/Other: Pulmonary embolism *not* MI & *not* ML	135	192
	1,065	958

not included in this report, were obtained with models that were developed on the Group II data and applied to Group I patients. Thus, the lung scan test meets the first definition of efficacy as earlier defined, i.e., it influences the physician's diagnostic thinking.

Effect of lung scan on management

In order to determine whether the lung scan had an effect on the management of the patient, the use of anticoagulant therapy (ACT) for the treatment of PE was then evaluated (Efficacy definition #2). Information on the planned management of the patient before and after performance of the lung scan was obtained

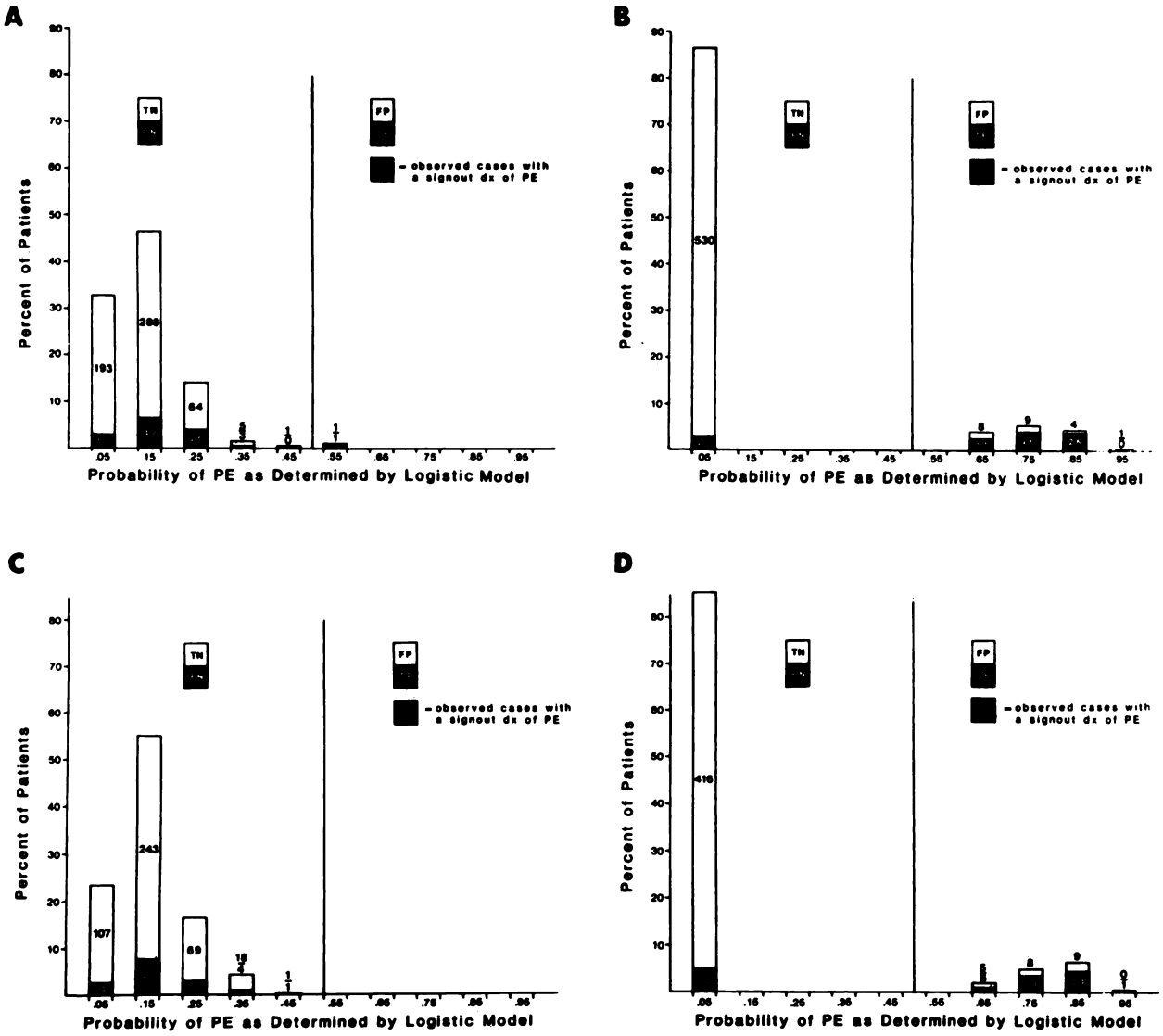

FIGURE 3
A: Group I model predicting Group I data prior to lung scan, subgroup PE/OTHER, n = 640. B: Group I model predicting Group I data after lung scan, subgroup PE/OTHER, n = 640. C: Group I model predicting Group II data prior to lung scan, subgroup PE/OTHER, n = 514. D: Group I model predicting Group I data after lung scan, subgroup PE/OTHER, n = 514

from Parts I and II of the questionnaire that had been completed by the referring physician. The numbers of patients by signout diagnosis in the eight possible pre- and postlung scan management combinations are given in the legend of Fig. 4.

For the purposes of this analysis, we assumed that patients who had a signout diagnosis of PE benefitted from the lung scan if ACT was the management plan after the lung scan results were known; patients who did not have a signout diagnosis of PE benefitted from the lung scan if ACT was not given after the results of the test were known. However, there are circumstances when ACT is not given for PE, such as a concurrent contraindication to ACT, e.g., peptic ulcer, gastrointestinal bleeding, thrombocytopenia or hematuria, and allergy to heparin. Patients with diseases other than PE, such as myocardial infarction or deep vein thrombosis, also may be treated with ACT. It is also given to some patients with cardiac valve prosthesis. Using the above definition of "benefit" which was based on the information available, the scan had a beneficial effect on the treatment of 1,770 (88%) of the 2,023 patients in the study.

Of these 1,770 patients, 407 (23%) had their therapy changed after the lung scan results were known. The remaining 1,363 patients (77%), having no change in plan of therapy before compared with after the test, benefitted from the lung scan because administering

the appropriate therapy was the ultimate goal.

Figure 4 illustrates this effect of the lung scan on the management plans of the referring physicians. The figure shows the relative number of patients in different therapy categories as thickness of a pathway, and keeps track separately of patients with PE (shaded area) and without PE (clear area). It illustrates that the number of patients receiving ACT after the lung scan is less than it was before the lung scan, in direct contradiction to Robin's contention that lung scans overdiagnose PE and increase the use of ACT (*15*). Furthermore, the "appropriate" pathways of altering therapy (B and G) show more traffic than the "inappropriate" pathways D and E. Pathways A and H are inappropriate pathways unchanged by the lung scan, and pathways C and F are appropriate pathways unchanged by the lung scan. The net effect of the lung scan is to decrease the numbers of patients on ACT, and to increase the proportion of patients with PE who are given ACT. The second definition of efficacy described previously as the probability of the test affecting the management plan for the patient has thus been met.

Based on the above assumption and calculations, it appears that referring physicians use the lung scan primarily to exclude the diagnosis of PE when clinical observations make this diagnosis an important one.

DISCUSSION

Other diagnostic tests

The rate at which pulmonary angiography is performed is reported to be between 7 and 15% of patients in the United States suspected of having PE. In this study, pulmonary angiography was performed in 5% of the 2,023 cases, a rate we consider to be a reflection of usual clinical practice at the time of data collection (*16,17*).

What is the role of angiographic corroboration of pulmonary embolism? Pulmonary angiography is usually considered to be the definitive diagnostic tool for confirmation of pulmonary embolism (*18,15*). Most reported studies have compared findings from perfusion-ventilation studies to pulmonary angiography (*2–6,16*). The number of cases in these several reports ranges from 53 to 302. They are usually retrospective, i.e., the nuclear medicine test is compared to the angiogram after both tests are completed, this comparison being the interpretation of the nuclear medicine physician compared with that of the angiographer. An exception is a recent study reported by Hull et al. (*19*). In this prospective study, they concluded that patients with segmental or larger defects on perfusion, normal ventilation, and with venous thromboembolism demonstrated by venography can safely be given anticoagulants without requiring confirmation by pulmonary angiography. In some reports the original scans and angiograms are interpreted retrospectively by one or more physicians purposely blinded to the clinical history of each patient. These reports are particularly useful for development of criteria for diagnostic interpretation of the images. These studies, however, are not aimed at measuring the impact of the report of the nuclear medicine physician on the management plan carried out by the referring physician.

Many authors, (*20,10,16,17,19,21*) report that it is impractical and expensive to perform pulmonary angiograms on a large scale and probably unnecessary except when the perfusion/ventilation study is indeterminate and/or when special considerations for a patient make it advisable. One could summarize the views of a number of authors as to the indications for pulmonary angiogram thus: (1) when doubt exists because of fail-

TABLE 7
Logistic Models Based on Group I Data, Subgroup PE/PE (Most Important/Most Likely)

Item	Group I predicting Group I		Group I predicting Group II	
	(1) without test	(2) with test	(3) without test	(4) with test
Sensitivity* (%)	63	86	52	77
Specificity† (%)	70	84	69	87
PV+‡ (%)	68	85	48	80
PV−§ (%)	64	86	69	86
Accuracy¶ (%)	66	85	62	84

* Sensitivity = $\frac{TP}{TP + FN} \times 100.$

† Specificity = $\frac{TN}{TN + FP} \times 100.$

‡ PV+ = $\frac{TP}{TP + FP} \times 100.$

§ PV− = $\frac{TN}{TN + FN} \times 100.$

¶ Accuracy = $\frac{TP + TN}{TP + FP + TN + FN} \times 100.$

TABLE 8
Logistic Models Based on Group I Data, Subgroup PE/Other (Most Important/Most Likely)

Item	Group I predicting Group I		Group I predicting Group II	
	(1) without test	(2) with test	(3) without test	(4) with test
Sensitivity (%)	1	77	.1	66
Specificity (%)	99.9	96	99.1	95
PV+ (%)	50	76	.1	69
PV− (%)	86	96	85	94
Accuracy (%)	86	93	85	91

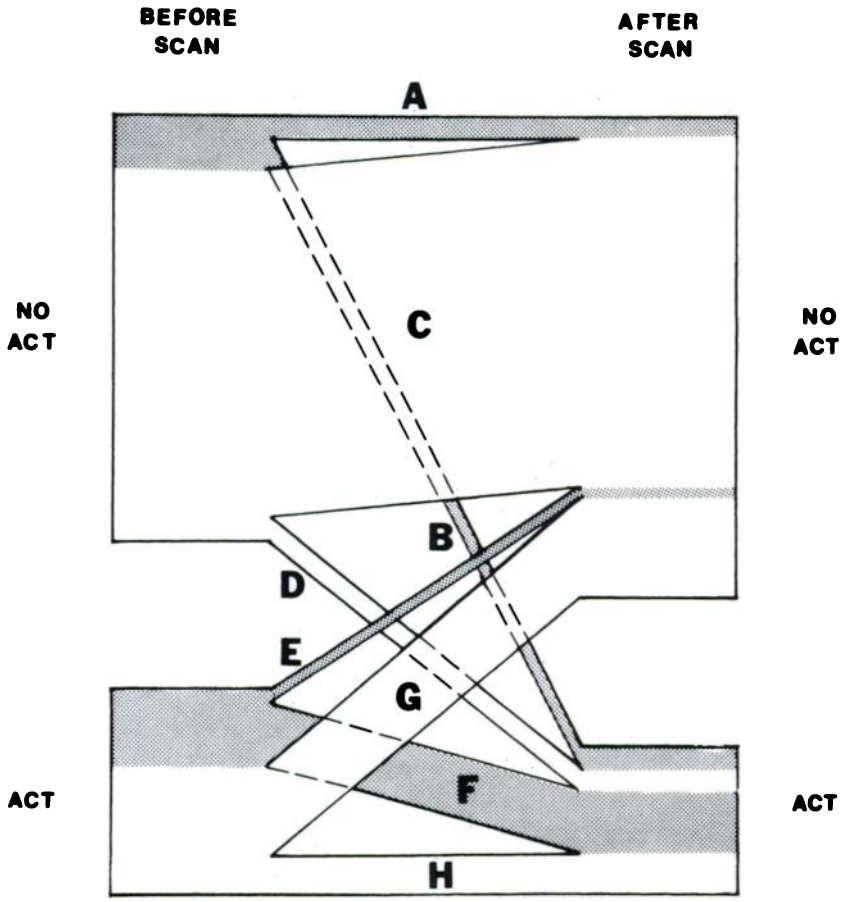

FIGURE 4
Shows the effect of lung scan on use of ACT. Shaded areas represent patients with PE. Individual groups are as follows: A, PE inappropriately kept off of ACT (64 patients); B, PE appropriately switched to ACT (108 patients); C, no PE appropriately kept off of ACT (1,143 patients); D, no PE inappropriately switched to ACT (45 patients); E, PE inappropriately switched to no ACT (25 patients); F, PE appropriately kept on ACT (217 patients); G, no PE appropriately switched to no ACT (302 patients); H, no PE inappropriately kept on ACT (119 patients)

ure of other diagnostic procedures (i.e., indeterminate lung scan findings) to affirm the presence or absence of PE thereby supporting a decision to treat or not treat; (2) the risk of anticoagulant therapy is high and therefore corroboration of the diagnosis is important enough to justify the added risk to the patient of this invasive procedure and the added cost (*22,19*); and (3) the lung scan suggests low probability but clinical suspicion is high (*17*).

Recent studies continue to recount the difficulties of making the diagnosis of PE. In a report by Goldhaber et al. (*23*), only 30% of 54 patients with major pulmonary embolisms resulting in death and found at autopsy were correctly diagnosed before death. Of those correctly diagnosed prior to death, PE was found on nine of 11 patients with lung scans (82%); pulmonary angiography identified four of five patients with PE (80%). This seems to point to underdiagnosis rather than overdiagnosis as claimed by Robin in 1977 (*15*).

The term "gold standard" is found widely throughout the literature of efficacy. It refers to some method whose certainty of correct diagnosis or healing is purportedly far better established or proven than that being investigated, at least a priori. In fact, these gold standards are often subject to the same fluctuating interpretations as the test or treatment under investigation. More often than not the final interpretation or standard is a combination of clinical and test interpretation depending on follow-up. Autopsy or surgical material is always desirable but even these well established techniques have the same problems of pattern recognition, observer differences, and biased ascertainment as the tests under investigation. The ultimate assessment remains that of the physicians whose clinical acumen including test interpretation results in an estimate of the probability of disease and the clinical management most likely to help their patients (*24*).

We should be careful to point out that the analytic methods that we discuss are not the limiting factor. That is, if a "gold standard" such as a particular diagnostic test or autopsy findings are available, then these can be used to determine outcome class membership, rather than using tests or clinical outcomes which at the time may be considered to be less definitive. However, there are times when such measures are not available. In those circumstances, it is reasonable to study the influence of a diagnostic test on discharge diagnosis or influence on subsequent therapy, even if it is widely

assumed that the test is an important determinant of the final diagnosis (*25*).

As noted above, there was no attempt either at the local hospital level or by physicians in the study advisory group to reinterpret the nuclear medicine physician's report. The evaluation of efficacy of the test was based solely on the actions of the referring physician in regard to the diagnoses being considered, and the effect of the interpretation of the test on changes both in diagnosis and therapy. This study describes current clinical practice, not "what should be." Thus, the design and conclusions of the SNM study may differ from those reported by others.

Sample size

The SNM Study has the advantage of a large population of cases from which to draw compared to many reported efficacy studies. In studies with smaller sample sizes, investigators usually test the analytic technique on the complete data set available with the assumption or hope that the method will predict as well on a new data set. Both information theory and statistical theory predict that no method of analysis will be as good on a new independent data set as it is on the original set (*26*). The basis for this statement is that the initial method pays attention not only to the strong characteristics of the data in the original set but also to idiosyncracies and special noises. A new data set obtained by or from the same setting or by different individuals or recorded by different physicians or using different techniques for determinations will produce comparable basic data characteristics but will have differences with respect to idiosyncracies and sources of nonrandom noise. Thus, when the methodology is tested on a new data set, it will predict less well because of these variations in characteristics.

Logistic regression

The statistically significant patient attributes in Group I were not the same as those in Group II (Tables 4 and 5). The inclusion or exclusion of clinically important symptoms and signs did not affect the ability of the model from either group to predict satisfactorily the results in the other group. This finding is due to the relatively limited discriminating ability of these variables when compared to the referring physicians' prior probabilities of either the most important or most likely diagnoses. If the regression analyses are repeated using only the referring physicians' probabilities the results are essentially the same. The presence or absence of the symptoms and/or signs alone could not predict the referring physicians' prior probability of PE. This observation fits with the widely observed lack of attributes of enough significance to permit correct diagnosis of PE. In the categories of diagnoses in which PE is included, both clinical and laboratory attributes yield only suggestions of this possible diagnosis.

The addition of the results of the lung scan to the regression model produced a large increase in accuracy of the model in predicting a signout diagnosis of PE. Thus, the lung scan was found to have a significant impact on the clinical judgement of the referring physician.

Entropy minimax pattern detection analysis

The entropy minimax pattern detection method of analysis, based on Christensen's method (*27*), was used to identify patterns of patient attributes capable of predicting either treatment plan or signout diagnosis, both without and then with the lung scan results. A detailed description of this analytic method and its results when applied to the data herein reported appears in a companion article (*7*). Briefly, the method attempts to identify subgroups of patients for whom diagnosis, therapy, or the lung scan test result can be predicted reliably using only patient attributes. The lung scan would provide little additional information if one could identify such a subgroup. Actually, no such subgroup of patients was found. This observation suggests that, at this time, lung scans utilized in the diagnosis of pulmonary embolism are both useful and reliable. Improvement in one's ability to predict outcomes is a measure of the test's influence on physicians' decisions regarding patient management, in this case, encouraging anticoagulant therapy for pulmonary embolism, if present, and discouraging its use when pulmonary embolism has been ruled out.

Comparison of methods

Table 9 compares the predictive values, sensitivity and specificity of the logistic regression approach and the entropy minimax pattern detection method for predicting a discharge diagnosis of PE. This comparison is

TABLE 9
Comparison of Entropy Minimax Pattern Detection to Logistic Regression Analysis: Group I Model Predicting Group II (Subgroup PE/PE n = 245)

Signout Diagnosis of PE	Sensitivity (%)	Specificity (%)	PV+ (%)	PV− (%)	%POP* (%)
Before scan					
Entropy	44	59	29	75	40
Logistic	52	69	58	69	100
After scan					
Entropy	68	76	58	83	71
Logistic	78	87	80	86	100

* Percent of subgroup of PE/PE to which method of analysis applies.

based on the Group I model predicting Group II, subgroup PE/PE. Also given are the percentages of the population for which these methods apply. Although these measures tend to be similar for both methods, the entropy minimax pattern detection method predicts on relatively small groups within the population. Since the major predicting variable in the logistic regression approach is the referring physician's prior probability of PE, it is not unusual that the entropy minimax method does not predict as well in this variable. The major advantage of the entropy minimax method is that it does not require prior probabilities, only the patients' attributes. However, there must be patient attributes, perhaps intuitive on the part of the physician, that the referring physician uses in reaching an assessment of the probability of a diagnosis of PE which have not been evaluated in the data set available to the entropy minimax method. Since the logistic regression approach utilizes this prior probability, it is quite likely that both recorded and unrecorded attributes are encompassed within the physician's probability estimates.

CONCLUSIONS

Since scientific methods of studying efficacy are relatively new and still under development, it was thought desirable to compare two different methods of evaluating efficacy. For the purpose of this study, the two methods chosen were logistic regression and entropy minimax pattern detection. They were applied to data composed of information about lung scanning as utilized by clinicians to diagnose or rule out pulmonary embolism in their routine practice of medicine.

Using these two methods, we find substantial agreement between them in the following clinical observations regarding the efficacy of lung scanning in the diagnosis of pulmonary embolism.

1. The analyses using logistic regression for the clinical attributes recorded, taken singly or in combination, indicate that no grouping of symptoms, signs and routine laboratory findings is capable of predicting that a given patient does or does not have a reasonable probability of having pulmonary embolism. The entropy minimax analysis reached similar conclusions.

2. The logistic regression model using the prior probability estimate of the referring physician in addition to the prior clinical information is also a poor predictor of the probability of a signout diagnosis of PE or no PE; adding the lung scan diagnosis significantly improves the ability of the model to predict the signout diagnosis with respect to PE (Efficacy definition #1).

3. As reflected in this study, and based on management of patients with and without PE, the lung scan is a simple and reliable test to indicate to the referring physician the likelihood of a patient having a pulmonary embolus. The diagnostic findings are reflected in the management regimens selected when one compares prior to posterior judgment of the referring physician (Efficacy definition #2).

4. The advantage of the use of logistic regression analysis is the ability to predict on the entire data set while the EMPD method predicts on relatively small subgroups of data. However, the EMPD method does not require the referring physician's prior probability, only the patient's attributes.

FOOTNOTE

* The questionnaire and the instruction manual are available from the Radioisotope Laboratory, ML 577 UC Medical Center, Cincinnati, OH 45267.

ACKNOWLEDGMENTS

This study was supported by the U.S. Department of Energy Contract No. DE-AC02-77EV04533.

The authors wish to thank Drs. Lee B. Lusted, Dennis D. Patton, and Barry A. Siegel for reviewing the manuscript, and for their helpful comments and suggestions, and are grateful to Mrs. Shira Karns for her assistance in the preparation of this manuscript, and to Ismael Kaya for computer programming and related technical assistance.

The authors also thank Dennis D. Patton for development of the graphic presentation of complex data in Figure 4.

THE SOCIETY OF NUCLEAR MEDICINE COMMITTEE ON PUBLIC HEALTH AND EFFICACY

Eugene L. Saenger
Chairman and Principal Investigator
Sheldon P. Baum
A. Bertrand Brill
Frederick W. George, III
Paul B. Hoffer
Robert W. McConnell
Barbara J. McNeil*
August Miale, Jr.
Dennis D. Patton*
E. James Potchen
F. David Rollo
William Schonbein*
Barry A. Siegel*
Edward B. Staab
Theodore J. Stahl
Henry N. Wagner, Jr.
Henry N. Wellman*

G. John Weir, Jr.
David A. Gift, Consultant
* Advisory Committee

Participating Nuclear Medicine Physicians

Faud S. Ashkar
Robert W. Burt
Robert F. Carretta
Harold L. Chandler
R. Edward Coleman
John D. Denny
Albert H. Fink
Patricia A. Germon
Gilbert J. Guiberteau
Hirsch Handmaker
Mukbil H. Hourani
John E. Kauth
Douglas R. Lilly
John F. Lindeman
Jerold M. Lowenstein
Raymond Marty
Philip Martin
Betty J. McConnell
Robert W. McConnell
Robert L. Meckelnburg
Richard W. Miller
Jose Oscar Morales
Bernard E. Oppenheim
Hee-Myung Park
Dan G. Pavel
Louis A. Perez
Thomas W. Ryerson
Vidya V. Sagar
Donald S. Schauwecker
John S. Seder
Robert A. Shor
Aslam R. Siddiqui
Theodore J. Stahl
John C. Stauffer
Dennis M. Welch
Henry N. Wellman
G. John Weir, Jr.
Ha Yong Yum

REFERENCES

1. "A Study of the Efficacy of Diagnostic Radiologic Procedures: Final Report on Diagnostic Efficacy," American College of Radiology, Lee B. Lusted, MD, Chairman, Committee on Efficacy Studies, American College of Radiology, Chicago, IL, May 1977
2. Bogren HG, Berman DS, Vismara LA, et al: Lung ventilation-perfusion scintigraphy in pulmonary embolism. *Acta Radiol [Diagn] (Stockh)* 19(6):933-944, 1978
3. Fischer KC, McNeil BJ: The indeterminate lung scan: Its characteristics and its association with pulmonary embolism. *Eur J Nucl Med* 4:49-53, 1979
4. Goodall RJR, Greenfield LJ: Clinical correlations in the diagnosis of pulmonary embolism. *Ann Surg* 191(2):219-223, 1980
5. Novelline RA, Baltarowick OH, Athanasoulis CA, et al: The clinical course of patients with suspected pulmonary embolism and a negative pulmonary arteriogram. *Radiology* 126:561-567, 1978
6. Sasahara AA, et al: Clinical and physiologic studies in pulmonary thromboembolism. *Am J Cardiol* 20:10-20, 1967
7. Gift DA, Schonbein WR, Saenger EL, et al: Application of an information-theoretic method for efficacy assessment. *J Nucl Med*: in press
8. Dalen JE, Alpert JS: Natural history of pulmonary embolism. *Prog Cardiovasc Dis* 17(4):259-270, 1975
9. Alpert JS, Smith R, Carlson J, et al: Mortality in patients treated for PE. *JAMA* 236:1477-1480, 1976
10. Biello DR, Maltar AG, McKnight RC, et al: Ventilation-perfusion studies in suspected pulmonary embolism. *Am J Roentgenol* 133:1033-1037, 1979
11. Alderson PO, Biello DR, Sachariah KG, et al: Scintigraphic detection of pulmonary embolism in patients with chronic obstructive lung disease. *Radiology* 138:661-666, 1981
12. Goetz WA: A nationally representative survey of in vivo diagnostic nuclear medicine, Medically Oriented Data System (MODS). Bureau of Medical Devices, Food and Drug Administration
13. Guidotti TL, Gries LF, Bell WR, et al: Accuracy of screening for pulmonary embolism in the emergency room. *Respiration* 37:309-317, 1979
14. Press SJ, Wilson S: Choosing between logistic regression and discriminant analysis. *J Am Stat A* 73:699-705, 1978
15. Robin E: Overdiagnosis and overtreatment of pulmonary embolism: The emperor may have no clothes. *Ann Intern Med* 87:775, 1977
16. Sostman HD, Ravin CE, Sullivan DC, et al: Use of pulmonary angiography for suspected pulmonary embolism: Influence of scintigraphic diagnosis. *Am J Roentgenol* 139:673-677, 1982
17. Braun SD, Newman GE, Ford K, et al: Ventilation-perfusion scanning and pulmonary angiography: Correlation in clinical high-probability pulmonary embolism. *Am J Roentgenol* 143:977-980, 1984
18. Menzoian JO, Williams LF: Is pulmonary angiography essential for the diagnosis of acute pulmonary embolism? *Am J Surg* 137:543-548, 1979
19. Hull RD, et al: Pulmonary angiography, ventilation lung scanning, and venography for clinically suspected pulmonary embolism with abnormal perfusion lung scan. *Ann Intern Med* 98:891-899, 1983
20. Cheely R, McCartney WH, Perry JR, et al: The role of noninvasive tests versus pulmonary angiography in the diagnosis of pulmonary embolism. *Am J Med* 70:17-22, 1981
21. Viamonte M, Koolpe H, Janowitz W, et al: Pulmonary thromboembolism—Update. *JAMA* 243:2229-2234, 1980
22. Neumann RD, Sostman HD, Gottschalk A: Current status of ventilation-perfusion imaging. *Semin Nucl Med* 10(3):198-217, 1980
23. Goldhaber SZ, Hennekens CH, Evans DA, et al: Factors associated with correct antemortem diagnosis of major pulmonary embolism. *Am J Med* 73:822-826, 1982

24. Stollerman, GH: The gold standard. Editorial. *Hosp Prac* 20:9, 1985
25. Sheps SB, Schechter MT: The assessment of diagnostic tests: A survey of current medical research. *JAMA* 252:2418-2422, 1984
26. Copas JB: Regression, prediction and shrinkage. *Jr Statis Soc B* 45(3):311-354, 1983
27. Christensen RA: *Entropy Minimax Method of Pattern Discovery and Probability Determination,* Cambridge, MA, Arthur D. Little, Inc., 1972