Lung Scan Interpretation: A Physiologic, User-Friendly Approach

William C. Klingensmith, III and Steve A. Holt

Department of Diagnostic Imaging, Saint Joseph Hospital, Denver, Colorado

A physiologically-oriented, user-friendly algorithm for interpretation of ventilation-perfusion ($\dot{V}/\dot{P}$) lung scans was compared to the widely used Biello criteria. The physiologic algorithm relies primarily on the observer's degree of certainty that a $\dot{V}/\dot{P}$ mismatch is present rather than on the size (relative to a whole segment) and number of $\dot{V}/\dot{P}$ mismatches. One hundred five patients who had undergone a ventilation study with ^{133}Xe gas, a perfusion study with ^{99m}Tc-MAA and pulmonary angiography were reviewed. Each $\dot{V}/\dot{P}$ lung scan was interpreted once with the physiologic algorithm and once with the Biello criteria by two observers. Fifty-nine percent of the 105 studies were classified as indeterminate with the physiologic algorithm versus 62% with the Biello criteria. The prevalence of pulmonary embolism in the high probability, indeterminate, low probability and normal categories for the physiologic algorithm were 60%, 26%, 7% and 0%, respectively; and for the Biello criteria were 50%, 31%, 11% and 0%, respectively. Analysis of the data with receiver-operating-characteristic curves indicated that the physiologic algorithm performed better than the Biello criteria ($p < 0.05$). In addition, the physiologic algorithm has the advantages of being more intuitive and easier to remember. We conclude that further evaluation of the physiologic approach for the interpretation of $\dot{V}/\dot{P}$ lung scans is warranted.

J Nucl Med 1992; 33:1417–1422

A number of algorithms have been proposed for the interpretation of ventilation-perfusion ($\dot{V}/\dot{P}$) lung scans in the diagnosis of pulmonary embolism (*1–4*). Of the proposed algorithms, the Biello, PIOPED and McNeil approaches have been the ones most extensively studied (*5–7*). A recent comparison of these three algorithms found no statistical differences among them, but suggested that the "Biello criteria" performed the best (*8*).

The Biello, PIOPED and McNeil algorithms all share an emphasis on the number and size of $\dot{V}/\dot{P}$ mismatches (ventilation better than perfusion) in making the interpretation of high probability for pulmonary embolism. This emphasis requires the user to memorize the details of various size categories of perfusion defects (in terms of percentages of a pulmonary segment) and the number and size combinations of $\dot{V}/\dot{P}$ mismatches that determine the interpretative categories. Many users resort to carrying a pocket card that contains the interpretive criteria because the details of the method are relatively nonintuitive.

For a number of years, we have used a more physiologic and intuitive algorithm that is easier to remember. In essence, it emphasizes the observer's certainty that a $\dot{V}/\dot{P}$ mismatch is present rather than the size and number of mismatches. The degree of certainty may be high, moderate, low or normal (essentially no probability) that a $\dot{V}/\dot{P}$ mismatch and, therefore, a pulmonary embolism is present. This approach allows a large perfusion defect or multiple segmental perfusion defects without definite mismatches due to, for technical reasons, patient movement, etc. to be classified as something other than high probability. At the same time, a relatively small perfusion defect with a clear cut mismatch can be classified as moderate or high rather than low or indeterminate.

The physiologic and Biello algorithms were compared in a retrospective series of 105 patients who had undergone ^{133}Xe gas ventilation studies (with reventilation studies as needed), ^{99m}Tc-macroaggregated albumin perfusion studies and pulmonary angiograms.

METHODS

Patient Selection

One hundred thirty consecutive pulmonary angiograms performed over an 18-mo period were used to construct an initial list of patients. Patients were excluded from this list if the pulmonary angiogram was not technically satisfactory, if they had not had a $\dot{V}/\dot{P}$ lung scan within 3 days prior to the pulmonary angiogram, or if the $\dot{V}/\dot{P}$ scan was technically unsatisfactory or incomplete.

Ventilation Study

The routine ventilation study was performed prior to the perfusion study in the posterior projection with 15–25 mCi of ^{133}Xe gas and a large field of view gamma camera. The routine study consisted of analog images in three phases. The first phase was a single breath image for 50,000–100,000 counts to evaluate ventilation. The second phase consisted of a 3-min rebreathing period followed by an equilibrium image for 300,000 counts for evaluation of airspace. The third phase consisted of a series of 30-sec images during washout of the xenon for evaluation of airway obstruction; these images were continued until activity

Received Aug. 2, 1991; revision accepted Feb. 7, 1992.

For reprints contact: Wm. C. Klingensmith III, MD, 3333 S. Bannock St., Ste 600, Englewood, CO 80110.

was no longer present, as judged from the persistence scope, for a minimum of four images.

In addition, reventilation studies following the perfusion study were obtained in projections other than the posterior projection when perfusion abnormalities were not well visualized in the posterior projection and were better visualized in other projections (*9,10*). For reventilation studies, the patient was positioned in the appropriate projection and a repeat perfusion image was obtained to document patient position and to compare it to the subsequent ventilation images. An image was then obtained of the distribution of ^{99m}Tc-MAA through the ^{133}Xe energy window for the time of the previous single breath ventilation image; this image demonstrates the distribution and amount of scatter present in the subsequent ventilation study. Following this, a complete ventilation study was acquired with a second dose of 15–25 mCi of ^{133}Xe gas.

Perfusion Study

The perfusion study was performed following the routine ventilation study using 6 mCi ^{99m}Tc-MAA. Analog images were acquired in the anterior, right and left lateral, right and left posterior oblique and posterior projections for 500,000 counts each using a large field of view gamma camera.

Pulmonary Angiography

The pulmonary angiograms were performed by five vascular radiologists experienced in pulmonary angiography. The $\dot{V}/\dot{P}$ lung scan was used to identify the regions of concern. The initial imaging sequence was performed in the frontal projection. If no pulmonary emboli were identified, at least one more injection was performed in an oblique projection. Additional oblique projections as well as magnification views and superselective injections were made at the discretion of the angiographer. In those patients in whom no pulmonary emboli were found in the suspect lung, at least one injection was made in the opposite main pulmonary artery. For all patients, the study was terminated as soon as the angiographer felt certain about the angiographic diagnosis of emboli. An intravascular filling defect was the essential criterion for an embolus.

The original interpretation was used; the studies were not reinterpreted. However, the reports were reviewed for evidence of technical inadequacies and the films were reviewed if there was a question of technical inadequacy.

Interpretation Strategy

One nuclear medicine specialist (WCK, observer A) and one diagnostic radiologist with extensive experience in nuclear medicine (SAH, observer B) interpreted each $\dot{V}/\dot{P}$ lung scan independently without knowledge of the original interpretation or the results of the pulmonary angiogram. The $\dot{V}/\dot{P}$ scan was interpreted in conjunction with the corresponding chest radiograph(s) obtained within 24 hr preceding the $\dot{V}/\dot{P}$ scan. Each observer interpreted each $\dot{V}/\dot{P}$ scan twice at one sitting, once with the physiologic algorithm and once with the Biello criteria. The Biello criteria were available for consultation in the form of a printed sheet posted next to the view boxes.

Physiologic Algorithm

The interpretation was forced into one of five categories: normal, low probability, moderate probability, high probability and indeterminate for pulmonary embolism (Table 1). In the physiologic approach, there are two key interpretive decisions: (1) the certainty of a $\dot{V}/\dot{P}$ mismatch, with ventilation preserved in face of a perfusion defect consistent with an embolic origin, and (2) the presence or absence of a radiographic density consistent with an acute pulmonary infarct. A perfusion defect consistent with an embolic origin is one that corresponds to segmental boundaries; a radiographic density consistent with an acute pulmonary infarct is a pleural-based infiltrate.

A normal interpretation requires no perfusion abnormalities and no radiographic densities consistent with an infarct. Low probability requires <10% probability of a $\dot{V}/\dot{P}$ mismatch and no radiographic density consistent with an infarct. Moderate probability requires ~10–90% probability of a $\dot{V}/\dot{P}$ mismatch and no radiographic density consistent with an infarct. High probability requires >90% probability of a $\dot{V}/\dot{P}$ mismatch. Indeterminate requires a radiographic density consistent with acute pulmonary infarct and no high probability of $\dot{V}/\dot{P}$ mismatch.

Several secondary considerations that have been shown to be helpful were also considered. One was preserved peripheral perfusion adjacent to a perfusion defect in scans demonstrating evidence of obstructive airway disease (stripe sign); this sign decreases the likelihood that the perfusion defect is embolic in origin (*11*). Another was perfusion defects that are significantly smaller than the corresponding radiographic densities, a finding that also decreases the likelihood of an embolic origin (*12*). In addition, subsegmental $\dot{V}/\dot{P}$ mismatches, even when clear cut, were interpreted as moderate probability in view of reports in the literature that pulmonary angiography documents emboli in this situation only about 40% of the time (*13*).

Biello Criteria

The Biello criteria used in this study were identical to those used in a recent comparative study of the Biello criteria and other interpretive algorithms (Table 1) (*8*). These, in turn, are the same as those published by Biello in a 1987 review article except that a single, large $\dot{V}/\dot{P}$ mismatch is interpreted as high probability rather than indeterminate (*2*).

The same secondary criteria used in the physiologic algorithm were also used in the Biello criteria. The use of a perfusion defect substantially smaller than a radiographic density to indicate low probability and a single, medium-sized $\dot{V}/\dot{P}$ mismatch to indicate intermediate probability are original components of the Biello criteria. The stripe sign is not a part of the Biello criteria, but is generally recognized as useful and was used in both interpretive algorithms.

Data Analysis

Interobserver and intraobserver agreement were evaluated with the kappa statistic (*14*). For this analysis, the results were grouped according to the implication of the $\dot{V}/\dot{Q}$ scan interpretation (prior to considering a priori clinical probability and other clinical factors): normal and low probability in general mean no pulmonary embolus, no anticoagulate; indeterminate and moderate probability in general mean a pulmonary angiogram; and high probability in general means anticoagulate. For the interobserver analysis, observer A was arbitrarily used as the reference standard; for the intraobserver analysis, the physiologic algorithm was arbitrarily used as the reference standard.

The accuracy of the two interpretive algorithms was compared using receiver-operating-characteristic (ROC) curves after the results of the two observers were averaged (*15,16*). For this analysis, the indeterminate and moderate probability categories

TABLE 1
Algorithms for the Interpretation of $\dot{V}/\dot{Q}$ Lung Scans

Category	Physiologic algorithm	Biello criteria
Normal	Normal perfusion; no radiographic density	Normal perfusion
Low probability	<10% probability of $\dot{V}/\dot{P}$ mismatch; no radiographic density consistent with infarct	Small (<25% segment) $\dot{V}/\dot{P}$ mismatch
		$\dot{V}/\dot{P}$ matches without corresponding radiographic changes
		Perfusion defect substantially smaller than radiographic density
Moderate probability	10%–90% probability of $\dot{V}/\dot{P}$ mismatch	
Indeterminate	Radiographic density consistent with infarct; no >90% probability of $\dot{V}/\dot{P}$ mismatch	Severe COPD with perfusion defects
		Single medium (25%–90% segment) $\dot{V}/\dot{P}$ mismatch
		Perfusion defect same size as radiographic density
High probability	>90% probability of $\dot{V}/\dot{P}$ mismatch	One or more large (>90% segment) $\dot{V}/\dot{P}$ mismatches
		Perfusion defect substantially larger than radiographic density
		Multiple medium (25%–90% segment) $\dot{V}/\dot{P}$ mismatches

of the physiologic algorithm were combined, but no other categories were combined. This approach gives four points on the ROC graph for each interpretive algorithm.

RESULTS

Patient Data

Of the 130 pulmonary angiograms performed over the 18-mo period, 25 were not utilized because: a lung scan was not performed (11 patients), the time between the lung scan and pulmonary angiogram exceeded 3 days (4 patients), the radiology file was unavailable (3 patients), only a perfusion scan was performed (3 patients), the indication for the lung scan and pulmonary angiogram was not pulmonary embolism (2 patients), the lung ventilation study was performed only after the perfusion study (1 patient) or the pulmonary angiogram was unsatisfactory because only digital subtraction was done with an inferior vena caval injection (1 patient). No patient was studied twice.

There were 48 males with an average age of 57.9 yr and 57 females with an average age of 55.6 yr. A chest radiograph(s) was available on the same day in 88 patients and from the preceding day in 17 patients.

Selection Bias

During the 18-mo period, a total of 1284 lung scans were performed. Since 11 of the 130 patients undergoing pulmonary angiograms did not have $\dot{V}/\dot{P}$ lung scans, the percent of patients undergoing scans who went on to pulmonary angiography was 9% (119 of 1284 patients). Thus, there was a very strong post-test (post-$\dot{V}/\dot{P}$ scan) selection bias.

Inter- and Intraobserver Variability

The two observers agreed 60% of the time when using the physiologic algorithm and 56% of the time when using the Biello criteria (Table 2). Observer A agreed with himself 84% of the time using the two different algorithms, and observer B agreed with himself 65% of the time using the two algorithms.

The kappa values for interobserver agreement for the physiologic algorithm and Biello algorithms were 0.29 and 0.20, respectively ($p < 0.003$ for both). The values for intraobserver agreement for observers A and B were 0.66 and 0.41, respectively ($p < 0.0001$ for both).

Physiologic Algorithm Versus Biello Criteria

The averaged results of the two observers for the physiologic algorithm and for the Biello criteria are shown in

TABLE 2
Interobserver and Intraobserver Agreement*

Category	Physiol algorithm Observer 1 vs. 2	Biello criteria Observer 1 vs. 2	Observer 1 Physiol-Biello	Observer 2 Physiol-Biello
Normal/Low	15/24 (63)	12/19 (63)	15/24 (63)	26/33 (79)
Intermediate	42/68 (62)	43/75 (57)	63/68 (93)	38/56 (68)
High	6/13 (46)	4/11 (36)	10/13 (77)	5/16 (31)
Total	63/105 (60)	59/105 (56)	88/105 (84)	68/105 (65)

* Data presented as agreements/total and percentage of agreements inside parentheses.

TABLE 3
Interpretive Results for the Physiologic Algorithms and Biello Criteria*

Category	Emboli/Total (%)	
	Physiologic	Biello
Normal	0/0.5 (0)	0/2 (0)
Low probability	2/28 (7)	3/28 (11)
Moderate probability	5/19 (26)	—
Indeterminate	11/43 (26)	20/65 (31)
High probability	10/14.5 (69)	5/10 (50)
Total	28/105 (27)	28/105 (27)

* The results from the two observers were averaged; the averaging results in fractional numbers.

Table 3. The physiologic algorithm resulted in 59% of the lung scans being classified as either moderate probability or indeterminate, and the Biello criteria resulted in 62% of the lung scans being classified as indeterminate. By excluding the moderate probability and indeterminate categories, the accuracy of the physiologic algorithm was 85%, while the accuracy of the Biello criteria was 80%.

Both algorithms were always correct when the interpretation was normal; the Biello criteria classified an average of two scans in this category, while the physiologic algorithm placed an average of 0.5 scans (average of two observers) in the normal category. There was a tendency, without statistical significance, for the physiologic algorithm to be more accurate in the low and high probability categories.

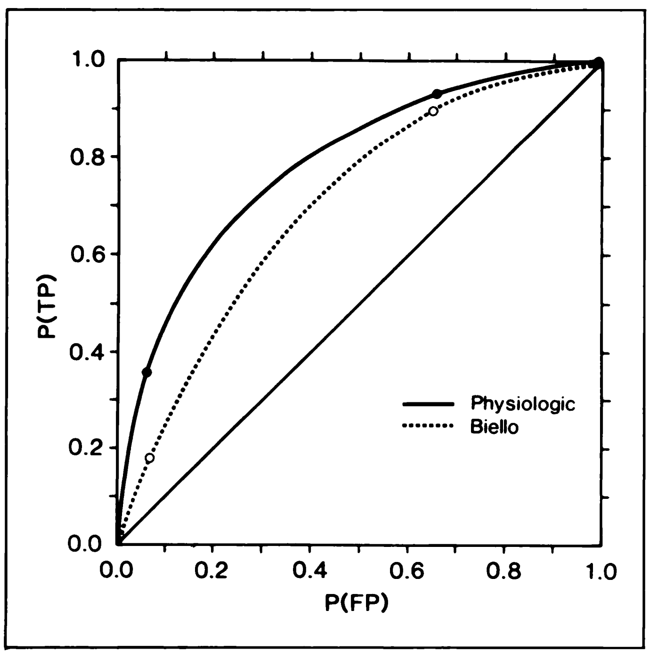

FIGURE 1. ROC curves are shown for the physiologic algorithm and the Biello criteria in the interpretation of $\dot{V}/\dot{P}$ lung scans. The fraction of false-positive results, P(FP), is shown on the x-axis and the fraction of true-positive results, P(TP), is shown on the y-axis. The areas under the ROC curves reflect the success of each algorithm in correctly categorizing the $\dot{V}/\dot{P}$ lung scans. The physiologic algorithm demonstrates a significantly greater area under the ROC curve, 0.73 versus 0.67 ($p < 0.05$).

The area under the ROC curve for the physiologic algorithm was 0.73 with a standard of error of 0.038, and for the Biello criteria it was 0.67 with a standard error of 0.062 (Fig. 1). The difference is statistically significant ($p = 0.048$).

DISCUSSION

In the present study, the physiologic algorithm demonstrated a small, but statistically significant, overall improvement in categorizing $\dot{V}/\dot{P}$ lung scans in comparison to the Biello criteria as demonstrated by ROC curve analysis. However, it is our feeling that the main advantage of the physiologic algorithm in comparison to the Biello criteria is its ease of use. The observer needs to focus on only two main considerations. First, is there a $\dot{V}/\dot{P}$ mismatch (ventilation better than perfusion) and what is the degree of certainty? Second, is there a radiographic density consistent with an acute pulmonary embolism with infarction? The answers to these two questions allow categorization of the $\dot{V}/\dot{P}$ lung scan into five areas (Table 1).

The physiologic algorithm emphasizes the pathophysiology of pulmonary embolic disease. Pulmonary emboli without infarction cause a decrease in perfusion without a concomitant decrease in ventilation. Consequently, the physiologic algorithm emphasizes the observer's certainty that a $\dot{V}/\dot{P}$ mismatch exists. The exact size and number of $\dot{V}/\dot{P}$ mismatches are secondary considerations, although, in general, the observer will be more certain that mismatches exist when they are large and/or multiple.

The Biello criteria, in contrast, emphasize the exact size and number of $\dot{V}/\dot{P}$ mismatches without considering the observer's degree of certainty that a mismatch is present. For example, the Biello criteria categorize a scan with complete absence of perfusion and normal ventilation in the right lower lobe the same as a scan with mildly or moderately decreased perfusion and normal ventilation in the right lower lobe because the $\dot{V}/\dot{P}$ mismatches are of the same size and number.

Similarly, in the case of a pulmonary embolus with infarction (and a radiographic density), the physiologic algorithm emphasizes the pathophysiology by focusing on whether or not the observer thinks that the radiographic density is consistent with an infarct, i.e., a pleural-based infiltrate. On the other hand, the Biello criteria, as usually stated, would categorize a perfusion defect the same size as the corresponding radiographic density as indeterminate even if the density was unlikely to represent an infarct, e.g., a spherical tumor. (Presumably most observers adjust for this difficulty.) In addition, we occasionally find that a normal looking $\dot{V}/\dot{P}$ scan can be seen in conjunction with a segmental density consistent with an infarct on the radiograph(s), e.g., infiltrate against the posterior surface of the heart. The physiologic algorithm would categorize this study as indeterminate, while the Biello criteria would categorize the study as normal. The present data do not include an example of an error by the Biello criteria for this reason.

For conceptual reasons, the physiologic algorithm contains two categories that indicate an intermediate probability of pulmonary embolism even though they have the same clinical implication. Thus, there is one intermediate category for each of the two main considerations in the algorithm. In other words, for instances where a $\dot{V}/\dot{P}$ mismatch is questionable, a moderate probability category is used when the observer is not certain. Also, in cases where radiographic density consistent with an acute infarct is questionable, an indeterminate category is used when one is present.

It may be argued that the physiologic algorithm is more subjective than the Biello criteria because it does not define the certainty of a $\dot{V}/\dot{P}$ mismatch in quantitative terms. However, it is our feeling that an observer will learn how to make this judgment with experience just as he learns to determine which radiographic densities are compatible with acute infarcts or whether a cardiac silhouette is enlarged. In addition, we feel that this assessment is one that has to be made in interpreting $\dot{V}/\dot{P}$ lung scans and that leaving it out of the algorithm, as the Biello criteria do, does not solve the problem. The fact that the intraobserver agreement was slightly better with the physiologic algorithm supports the contention that experienced observers do not find the subjectivity of the physiologic algorithm difficult.

The data in this retrospective study cannot be used to determine the sensitivity and specificity of the two algorithms because of the strong post-test selection bias in deciding which patients were referred for pulmonary angiography. Only 9% of patients who underwent $\dot{V}/\dot{P}$ lung scans went on to pulmonary angiography. In contrast, the percentage of patients who underwent angiography in the prospective PIOPED study was 81% (*4*). However, it is of interest that the false-negative rate for low probability interpretations in the present study for both the physiologic algorithm, 9%, and the Biello criteria, 11%, were both lower than the 16% reported in the PIOPED study. The much smaller post-test selection bias in the PIOPED study should have improved its results. In part, the lower false-negative rate in the present study might be explained by the use of reventilation imaging which would be expected to improve the accuracy of $\dot{V}/\dot{P}$ lung scans.

The level of agreement between the two observers was relatively low, 60% for the physiologic algorithm and 56% for the Biello criteria, but is similar to what others have reported for the Biello criteria and other algorithms (*5,6*). The intraobserver agreement for each observer using the two different algorithms was 84% and 65%. The relatively high agreement for one observer, 84%, may reflect the fact that each $\dot{V}/\dot{P}$ lung scan was interpreted with both algorithms at the same sitting and that knowledge of the interpretation with one algorithm influenced the interpretation with the other.

The differences between the physiologic algorithm and the Biello criteria are to some extent one of emphasis. Both methods focus on $\dot{V}/\dot{P}$ mismatches as the hallmark of pulmonary emboli, but they use different approaches to categorizing them. For instance, the Biello criteria categorize a perfusion defect that is much larger than a radiographic density as high probability for pulmonary embolism. Whereas the physiologic algorithm would likely reach the same conclusion, the reasoning would be based on the existence of a clear cut $\dot{V}/\dot{P}$ mismatch in that part of the perfusion defect that did not correspond to the radiographic density. The size of the perfusion defect relative to the radiographic density and the size of the density are secondary.

In addition, several secondary considerations are used in both the physiologic algorithm and the Biello criteria. Two of these, a perfusion defect substantially smaller than the corresponding radiographic density as an indicator of low probability of pulmonary embolism and a single medium-sized $\dot{V}/\dot{P}$ mismatch as an indicator of intermediate probability, are explicitly part of the Biello criteria (*12*). The other is the stripe sign as an indicator of a perfusion defect that is secondary to obstructive airway disease (*11*).

Other secondary signs also have been shown to be helpful, but require clinical information, previous imaging studies or follow-up $\dot{V}/\dot{P}$ lung scans. None of these were available in the present study. These other signs are: a history of prior pulmonary embolism which decreases the positive predictive value of a $\dot{V}/\dot{P}$ mismatch (*4*); prior radiographs demonstrating that a radiographic density consistent with acute infarction is actually chronic, thus decreasing the likelihood that the density represents an acute infarct; and follow-up $\dot{V}/\dot{P}$ lung scans to determine whether a small $\dot{V}/\dot{P}$ mismatch is resolved—a resolving mismatch is more likely to represent a pulmonary embolism (*17*).

A limitation of the present study is that both observers were experienced in the use of the physiologic algorithm, but inexperienced in the use of the Biello criteria. However, the fact that the Biello criteria were posted next to the view boxes and were frequently referred to should have minimized this limitation. A comparison using two sets of observers, one experienced in the physiologic algorithm and one in the Biello criteria, would be of interest.

We conclude that the physiologic algorithm appears to be at least as accurate as the Biello criteria in the interpretation of $\dot{V}/\dot{P}$ lung scans in a selected group of patients. In addition, it has the advantage of being more intuitive and easier to use. We feel that further evaluation of the physiologic algorithm is warranted.

REFERENCES

1. Biello DR, Mattar AG, McKnight RC, et al. Ventilation-perfusion studies in suspected pulmonary embolism. *Am J Roentgenol* 1979;133:1033–1037.

2. Biello DR. Radiological (scintigraphic) evaluation of patients with suspected pulmonary embolism. *JAMA* 1987;257:3257–3259.
3. McNeil BJ. Ventilation-perfusion studies and the diagnosis of pulmonary embolism. *J Nucl Med* 1980;21:319–323.
4. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. *JAMA* 1990;263:2753–2759.
5. Carter WD, Brady TM, Keyes JW, et al. Relative accuracy of two diagnostic schemes for detection of pulmonary embolism by ventilation-perfusion scintigraphy. *Radiology* 1986;145:447–451.
6. Sullivan DC, Coleman RE, Mills, et al. Lung scan interpretation: effect of different observers and different criteria. *Radiology* 1983;149:803–807.
7. Spies WG, Burstein SP, Dillehay GL, et al. Ventilation-perfusion scintigraphy in suspected pulmonary embolism: correlation with pulmonary angiography and refinement of criteria for interpretation. *Radiology* 1986;159:383–390.
8. Webber MM, Gomes AS, Roe D, et al. Comparison of Biello, McNeil, and PIOPED criteria for the diagnosis of pulmonary emboli on lung scans. *Am J Roentgenol* 1990;154:975–981.
9. Jacobstein JG. Xe-133 ventilations scanning immediately following the Tc-99m perfusion scan. *J Nucl Med* 1974;15:964–968.
10. Kipper MS, Alazraki N. The feasibility of performing Xe-133 ventilation imaging following the perfusion study. *Radiology* 1982;144:581–586.
11. Sostman HD, Gottschalk A. The stripe sign: a new sign for diagnosis of nonembolic defects on pulmonary perfusion scintigraphy. *Radiology* 1982;142:737–741, 1982.
12. Biello DR, Mattar AG, Osei-Wusu A, et al. Interpretation of indeterminate lung scintigrams. *Radiology* 1979;133:189–194.
13. Rosen JM, Palestro CJ, Markowitz D, et al. Significance of single ventilation/perfusion mismatches in krypton-81m/technetium-99m lung scintigraphy. *J Nucl Med* 1986;27:361–365.
14. Fleiss J, John CM, Landes JR. Large sample variance of kappa in the case of different sets of raters. *Psych Bull* 1979;86:974–977.
15. Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating-method data. *J Math Psych* 1969;145:447–451.
16. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves from the same cases. *Radiology* 1983;148:839–843.
17. Alderson PO, Dzebolo NN, Biello DR, et al. Serial lung scintigraphy: utility in diagnosis of pulmonary embolism. *Radiology* 1983;149:797–802.

EDITORIAL

Lung Scan Interpretation: A Physiologic, User-Friendly Approach

Let me begin this editorial by setting up the following hypothetical situation. Let us assume I was interested in finding out whether French or Italian was a better romance language for the translation of Shakespearean sonnets. In order to evaluate which language was preferable, I secured the services of a linguist. This linguist spoke English and French fluently but no Italian. However, I supplied him with an English-Italian dictionary and asked him to proceed to translate the sonnets. Fortunately, I had a computer program set up that would evaluate how successfully the meaning of the sonnets was maintained. I also asked the linguist which language was easier to use. It would hardly be surprising if my linguist did a better job translating into French, found French easier to use and got a more accurate score from the computer for his French translations. However, it would be unwise for me to conclude that French was in fact the better language for this exercise.

We are facing virtually the same situation in the article, "Lung Scan Interpretation: A Physiologic, User-Friendly Approach," published in this issue. In this instance, the authors are quite familiar with and have used their "physiologic algorithm" for years and admit that in "the present study—both observers were experienced in the use of the physiologic algorithm, but inexperienced in the use of the Biello criteria." The authors go on to state that the Biello criteria were posted next to the view boxes for easy reference (in other words the dictionary was available). Therefore, I am not surprised that the authors do better with their "physiologic" scheme than with the Biello criteria, which, until this exercise, they did not use. It is expected that their algorithm works better for them than an unfamiliar one does.

I also believe they have one significant experimental design error that could have influenced their data. They point out that they collected both sets of data at one time. In other words, they used one sitting to obtain both the "physiologic" reading and the Biello criteria reading. Unfortunately, this makes it too easy unconsciously to give the Biello reading second-class status. Just a few pedantic comparisons (e.g., a minor lung lesion like linear atelectasis with a matching perfusion defect called a Biello intermediate) would be enough to make the physiologic ROC curve look better than the Biello ROC curve. It would be better, I think, to have interpreted both studies independently of each other to avoid the possibility that they subconsciously bias their data. I hasten to point out that the authors spend considerable time discussing the retrospective nature of their series and the biases involved, and I believe they have made a very honest attempt to perform the correlation they are describing.

Let us now ask the question, is it really a good idea to have a user-friendly approach? Regardless of what type of criteria you use or even how experienced you are at reading lung scans, most people can sort lung scans into at least three categories fairly quickly. These consist of the "easy high probability" examination, the "easy low probability examination," and the "oh boy this is trouble and I'm not sure what's going on" examination. I believe that the whole purpose of having any type of "complicated criteria" at all is to extract a low

Received Mar. 24, 1992; accepted Mar. 24, 1992.

For reprints contact: Alexander Gottschalk, MD, Professor of Radiology, B-220 Clinical Center, Michigan State University, East Lansing, MI 48824-1315.