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.

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.

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 or high probability reading from the latter category.

In addition, the authors make the point that the Biello criteria emphasizes the size and number of $\dot{V}/\dot{Q}$ mismatches as distinct from the degree of certainty that a $\dot{V}/\dot{Q}$ mismatch is present or not. We must recognize that reality may be more complicated than the authors would like. It is clearly true that a partially occluding embolus can cause a partial perfusion defect that may be more difficult to assess.

Let us now turn to the addition of secondary criteria, which the authors have applied to both their physiologic algorithm and the Biello scheme. These include the use of the stripe sign as an indicator of a nonembolic perfusion defect (an excellent addition I think though I am clearly biased) (*1*). Other secondary criteria include the perfusion defect that is "substantially smaller than the radiographic density" and a recognition that a single medium size (or subsegmental) mismatch is intermediate probability (*2*, *3*). By adding the secondary criteria, the authors have made their user-friendly approach just a little less friendly. In fact, one could ask why stop at the single subsegmental lesion, why not extend your secondary criteria to include, for example, the full segment. In newly analyzed PIOPED data, there were 24 cases with a single segmental mismatched lesion. Only 11 of these (46%) had pulmonary embolism (PE) (*4*). We should wonder why the "physiologic" algorithm has no trouble with classic mismatches like this. It would seem that the reason for this is because these cases do not happen often enough in this series. In the PIOPED trial, there were 255 emboli. If only 11 have a single segmental perfusion defect, then this pattern occurred 4% of the time. In the "user-friendly" series, there are only 28 emboli. As a result, we would only expect to see a single segmental perfusion defect in one or two cases at the most. Consequently, this potentially troubling finding could make no significant impact on the data.

If we examine the two algorithms carefully, the similarities become much closer than the differences. For example, under the low probability criteria in the Biello group, the small $\dot{V}/\dot{Q}$ mismatch must be understood by and carefully *ignored* by the "physiologic approach." Because a small mismatch is often clearly a mismatch it must be given no credence by the "user-friendly technique." Assuming this to be the case, when the secondary criteria are added, the only difference for low probability is that the user-friendly technique pays no attention to the $\dot{V}/\dot{Q}$ match. If we look at the moderate/indeterminate probability group again, the only significant difference is that the user-friendly approach pays no attention to the degree of COPD with matched perfusion deficit. The authors do argue that their description of a radiographic density consistent with infarct is more meaningful than a "perfusion defect the same size as a radiographic density," but as they point out, and I agree, virtually everyone interprets the Biello criteria to mean a perfusion defect that has some relationship to a segmental or subsegmental shape. The authors do not tell us how they sort out the types of lesions that can masquerade as a pulmonary infarct on chest x-ray. My somewhat tattered chest radiology text lists 11 major categories of "homogenous opacity of recognizable segmental distribution" with over 30 different possible etiologies (*5*). Furthermore, Greenspan et al. point out that a series of excellent chest radiologists trying to pick out PE patients from a group of chest x-rays of patients with symptoms of PE whose PE status was known (i.e., PE present or no PE) had a predictive index reflecting overall accuracy of 40% (*6*). I must conclude that this could be a very unfriendly part of their user-friendly algorithm.

However, there is one major difference between the "user-friendly" and Biello schemes that should be noted. The user-friendly algorithm places all *matched* defects into the low probability category, while the Biello criteria considers a significant amount of $\dot{V}/\dot{Q}$ match to be intermediate. It is interesting that the most recent evaluation of the PIOPED data agrees with the "physiological user-friendly" algorithm in this regard. This analysis of the PIOPED data finds no evidence that $\dot{V}/\dot{Q}$ matched lesions hide PE (*4*).

A "user-friendly" approach suggests it should be easy to use, and as a result should provide good agreement between observers. This turns out to be unclear from the data the authors present. For example, in the PIOPED trial, there was a 63% agreement between observers (*7*). With the "user-friendly" technique, the two observers agreed only 60% of the time and even less using the Biello criteria. Furthermore the kappa value for interobserver agreement for the "user-friendly" system of 0.29 is considered to be just "fair" and would have to exceed 0.61 to be "substantial" (*8*). Again, this suggests that the user-friendly approach is less friendly than the authors would like it to be.

I applaud the authors for pointing out that this angiographic proven series may have selection bias. It is interesting that if these data are compared to other angiographic series they turn out to be quite different. For example, Hull (*9*) argues that his prospective series at McMaster University has high probability readings that are accurate, but low probability readings that are not. Both Biello and PIOPED argue that both the high and low probability readings are reliable. This series, using its best algorithm, has high probability readings correct only 60% of the time, but gives quite accurate readings for the low probability group. These interpretation results tend to be opposite those of the larger prospective trials and certainly suggests the possibility of some patient selection bias.

Lastly, I think we must ask ourselves if being "user-friendly" is the ultimate goal of any $\dot{V}/\dot{Q}$ scan interpretation set of rules. I think being user-friendly is nice, but being consistent is much more important. This is particularly true in a large group, or when there are people with varied experience (such as residents and senior faculty) reading the studies. There is nothing more discouraging to a referring physician than to get several sets of interpretations on the same $\dot{V}/\dot{Q}$ scan. I think it is likely that the more diagnostic instruction you have in an algorithm the more likely the observer is to find instructions that match the particular study he is viewing. I accept the authors' contention that theirs is an easy algorithm to use. However, the fact that the two observers—presumably in a position to communicate with each other quite frequently—can only agree with each other 60% of the time, while the eight member, widely scattered PIOPED Nuclear Medicine working group, with a much more complicated algorithm, agreed 63% of the time suggests that ease of use in no way equates with consistency in interpretation.

Finally, having spent much time in the past months discussing rules for the interpretation of the $\dot{V}/\dot{Q}$ scan, I am fully aware that our readers often make their own algorithms or adopt their own variations of an established algorithm. I believe that the authors of this paper have presented their data and their results fairly. I hope that the reader will be able to use both the data and the remarks in this editorial to his/her ultimate diagnostic advantage when interpreting $\dot{V}/\dot{Q}$ studies.

Alexander Gottschalk
Michigan State University
East Lansing, Michigan

REFERENCES

1. Sostman HD, Gottschalk A. The strip sign: a new sign for diagnosis of nonembolic defects on pulmonary perfusion scintigraphy. *Radiology* 1982;142:737–741.
2. Biello DR, Mattar AG, Osei-Wusu A, et al. Interpretation of indeterminate lung scintigrams. *Radiology* 1979;133:189–194.
3. 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.
4. Gottschalk A, Greenspan RH. Pulmonary embolism—refresher course delivered at the Radiological Society of North America. Chicago, IL, 1991.
5. Fraser RG, Pare JAP *Diagnosis of diseases of the chest*-second edition. Philadelphia: Saunders; 1979;2119–2132.
6. Greenspan RH, Ravin CE, Polansky SM, McLoud TC. Accuracy of the chest radiograph in diagnosed pulmonary embolism. *Invest Radiol* 1982;17:539–543.
7. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. *JAMA* 1990;263:2753–2759.
8. Feinstein AR. *Clinical epidemiology.* Philadelphia: Saunders; 1985:186.
9. Hull RD, Raskob GE. Low-probability lung scan findings: a need for change. *Ann Intern Med* 1991;114:142–143.