

The Efficacy of Lung Scans: The Emperor Had Clothes All Along

This issue of the Journal contains two ground-breaking articles, each of which describes a technique for measuring the efficacy of a diagnostic test, in this case the lung scan (1,2). Both base their analyses on the same data, a large-scale study of over 2,000 ventilation-perfusion lung scans involving 22 hospitals, and a wealth of data on signs, symptoms, pre- and post-test disease probability, and pre- and post-test management plans. The study was carried out by the Efficacy Committee of the Society of Nuclear Medicine, with support from the Department of Energy. It is the largest such study ever reported and may serve as a model for future efficacy studies.

The ventilation-perfusion lung scans (VPLS) was chosen for this study for several practical reasons.

1. It is a high-volume procedure and a large study group can be collected in a reasonable time.
2. Pulmonary embolism (PE) is an acute disease, usually requiring prompt treatment and only brief hospitalization, and signout diagnoses are made quickly.
3. The referring physician's estimate of the probability that the patient has PE, i.e., his subjective probability estimate (SPE), can be determined before and after the lung scan.
4. Management decisions are made quickly, and consist mainly of whether or not to give anticoagulant therapy (ACT). Thus, it was possible to collect full data on over 2,000 lung scans in about 2 yr.

Selection of the lung scan suffers from some obvious and practical disadvantages, however. The only realistic "gold standard" is pulmonary angiography, a procedure that incurs additional cost and a small but finite risk. Pulmonary angiography is usually reserved for patients with an indeterminate ventilation-perfusion lung scan, or with a positive VPLS but a contraindication to ACT. Pulmonary angiography was in fact performed in only 5% of the patients in this study, a figure only slightly less than the national average (3). For this reason the efficacy study made no attempt to classify the scan results as true positive, true negative, etc. (except with respect to the signout diagnosis), but only to determine whether referring physicians did in fact make use of the results of the VPLS in decisions regarding diagnosis and therapy.

The two efficacy articles examine the question of efficacy from different angles. The article by Gift et al. (2) describes the efficacy of a diagnostic test in terms of its ability to reduce "information entropy," which is a measure of the degree of uncertainty associated with a diagnostic setting. When the diagnosis is established beyond doubt there is no uncertainty, and the entropy is zero. For most clinical settings there is some uncertainty, and the entropy is some positive value. The object of a diagnostic test is to reduce the uncertainty as much as possible, i.e., get the entropy as close to zero as possible, and the degree to which the test achieves this is a measure of its diagnostic efficacy. The entropy is determined by clinical and laboratory data and the VPLS results, but not by the physician's estimate of disease probability. The entropy-minimax model finds the diagnostic algorithm yielding the lowest entropy, i.e., the one best predicting the final outcome.

The other article, by Saenger et al. (1) describes efficacy in terms of a logistic regression model. This model attempts to develop an equation that predicts the signout diagnosis (or the treatment plan) based on clinical and laboratory data, the physician's estimate of disease probability, and the VPLS results, each represented by a variable with coefficients and/or exponents. The model finds the coefficients, etc., yielding the best equation, i.e., the one best predicting the final outcome.

Each of the two sets of authors tested the robustness of their model by dividing the study population into two groups of roughly the same size, and using the model developed in one

group (“training group”) to predict the management plan and signout diagnosis in the other group (“testing group”). If the model is too “local,” i.e., if it is based too much on peculiarities and idiosyncrasies of the group that was used to develop it, then it will fail to perform well in the other group. Both efficacy models did well in both the first and the second groups, validating their performance as predictors.

The goal of the efficacy study was to answer two questions.

1. Does the result of the VPLS really alter the referring physician’s estimate of the probability that the patient has PE?

2. Does the result of the VPLS really contribute to decisions on patient management? The first question addresses diagnostic efficacy (Efficacy-D, or Efficacy-1); the second, management efficacy (Efficacy-M, or Efficacy-2). The question of outcome efficacy (Efficacy-O) could not be addressed by this study as it would have required an unobtainable control group—patients thought to have PE from whom the VPLS was withheld.

Consider the first question. A diagnostic test has value to the extent that it is able to move the referring physician away from his initial estimate of disease probability, either up or down. Suppose a physician calls you from the emergency room asking you to perform an emergency VPLS on a patient he believes may have PE. You may ask, “What do you think the chances of PE are?” The reply is, “Oh, about 60%.” You decide, “OK, we’ll go ahead.” You do the study, then report back, “The study is abnormal, and I’d say there’s about a 60% chance that your patient has PE.” The reply is, “But I already knew that—I told you that! Can’t you tell me something I didn’t already know?” A diagnostic test cannot be efficacious if it systematically fails to alter the SPE.

In the efficacy study, if the result of the VPLS had no effect on the referring physician’s subjective estimate of disease probability (SPE), then one could use data obtained prior to the VPLS to predict whether or not the patient’s signout diagnosis would be PE or not PE. The VPLS would clearly not be efficacious, as it would not alter the SPE. The efficacy study found otherwise: Without the VPLS report, it was impossible to predict the signout diagnosis (PE or not PE) using any combination of signs, symptoms, and laboratory data, but when the VPLS report was included, the signout diagnosis could be predicted with high accuracy. It can be concluded that referring physicians do in fact rely heavily on the VPLS report in establishing their diagnosis.

Of course, without a “gold standard” to verify the VPLS diagnosis it is impossible to say whether the results of the VPLS alter the SPE in the right direction. Could negative scans undercall, and positive ones overcall? A *normal* perfusion lung scan in a patient with documented PE would be so rare that for all practical purposes that possibility can be ignored. The literature documents no such cases. Positive lung scans are, of course, far less specific than we would like them to be, and concern has been expressed that false-positive lung scans lead to overdiagnosing (and overtreating) PE (4). Again, the efficacy study showed otherwise: The effect of widespread use of the VPLS is to reduce the number of patients thought to have PE, and to reduce the use of ACT.

The second question the efficacy study addressed was whether the referring physician actually uses the VPLS result in formulating a management plan. Referring physicians were asked prior to the VPLS what management plan (ACT or no ACT) they would carry out if the VPLS were not available. Following the VPLS they were asked again what their management plan was then. If the plan had been essentially the same after the VPLS as it had been before, then the VPLS would not be efficacious as it would not contribute to decisions on patient management. Again, the study showed otherwise: Referring physicians do rely heavily on the VPLS to decide on their patient management plans.

It is well worth looking more closely at this effect. Figure 4 of the article by Saenger et al. is a “traffic pattern” showing the impact of the VPLS on patient management. While the largest traffic flow (VPLS effect) is keeping non-PE patients off of ACT (channel C in the figure, representing 1,143 patients), the second-largest flow, and the largest *change* in management, is taking non-PE patients off of ACT (channel G in the figure, representing 302 patients). This change in management is much more frequent than the inappropriate switching of non-PE patients to ACT (channel D, 45 patients), and is even greater than the appropriate keeping of PE patients on ACT (channel F, 217 patients). Thus, the VPLS

brought about the appropriate management change no-PE-switched-to-no-ACT nearly seven times as often as the inappropriate change no-PE-switched-to-ACT. This finding is in direct opposition to the widely quoted assertion by Robin (4) that widespread use of lung scanning leads to overtreatment with ACT. Now we have convincing experimental proof that this is not the case: the net effect of lung scanning is to *reduce* the use of ACT, not to increase it.

In summary, the efficacy study as analyzed by the two methods leads to the same two conclusions: Physicians rely on lung scans to make diagnoses, and they rely on them to decide on treatment, which, incidentally, cuts down on the use of anticoagulants. The study results confirm and even quantitate the efficacy of lung scanning in the diagnosis and management of pulmonary embolism. The efficacy study should serve as a model for future studies of the efficacy of other diagnostic tests.

Dennis D. Patton, MD
Division of Nuclear Medicine
University of Arizona
Health Science Center
Tucson, Arizona

REFERENCES

1. Saenger EL, Buncher CR, Sprecker BL, et al: Determination of clinical efficacy: Nuclear medicine as applied to lung scanning. *J Nucl Med* 26:793-806, 1985
2. Gift DA, Schonbein WR, Saenger EL, et al: Application of an information-theoretic method for efficacy assessment. *J Nucl Med* 26:807-811, 1985
3. 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
4. Robin ED: Overdiagnosis and overtreatment of pulmonary embolism: The emperor may have no clothes. *Ann Intern Med* 87:775-781, 1978