Users' Guides to the Medical Literature

III. How to Use an Article About a Diagnostic Test

B. What Are the Results and Will They Help Me in Caring for My Patients?

Roman Jaeschke, MD, MSc; Gordon H. Guyatt, MD, MSc; David L. Sackett, MD, MSc; for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

You are back where we put you in the previous article[1] on diagnostic tests in this series on how to use the medical literature: in the library studying an article that will guide you in interpreting ventilation-perfusion (V/Q) lung scans. Using the criteria in Table 1, you have decided that the Prospective Investigation of Pulmonary Diagnosis (PIOPED) study[2] will provide you with valid information. Just then, another physician comes looking for an article to help with the interpretation of V/Q scanning. Her patient is a 28-year-old man whose acute onset of shortness of breath and vague chest pain began shortly after completing a 10-hour auto trip. He experienced several episodes of similar discomfort in the past, but none this severe, and is very apprehensive about his symptoms. After a normal physical examination, electrocardiogram and chest radiograph, and blood gas measurements that show a P_{CO_2} of 32 mm Hg and a P_{O_2} of 82 mm Hg, your colleague has ordered a V/Q scan. The results are reported as an "intermediate-probability" scan.

You tell your colleague how you used GRATEFUL MED to find an excellent article addressing the accuracy of V/Q scanning. She is pleased that you found the article valid, and you agree to combine forces in applying it to both your patients.

In the previous article on diagnostic tests, we presented an approach to deciding whether a study was valid, and the results therefore worth considering. In this installment, we explore the next steps, which involve understanding and using the results of valid studies of diagnostic tests.

WHAT ARE THE RESULTS?

Are Likelihood Ratios for the Test Results Presented or Data Necessary for Their Calculation Included?

Pretest Probability.—The starting point of any diagnostic process is the patient, presenting with a constellation of symptoms and signs. Consider the two patients who opened this exercise—the 78-year-old woman 10 days after surgery and the 28-year-old anxious man, both with shortness of breath and nonspecific chest pain. Our clinical hunches about the probability of pulmonary embolus (PE) as the explanation for these two patients' complaints, that is, their pretest probabilities, are very different: the probability in the elderly woman is high, and in the young man the probability is low. As a result, even if both have intermediate-probability V/Q scans, subsequent management is likely to differ. One might well treat the elderly woman but order additional investigations in the young man.

Two conclusions emerge from this line of reasoning. First, whatever the results of the V/Q scan, they do not tell us whether PE is present. What they do accomplish is to modify the pretest probability of PE, yielding a new posttest probability. The direction and magnitude of this change from pretest to posttest probability are determined by the test's properties, and the property that we shall focus on in this series is the likelihood ratio (LR).

The second conclusion we can draw from our two contrasting patients is that the pretest probability exerts a major influence on the diagnostic process. Each item of the history and physical examination is a diagnostic test that either increases or decreases the probability of a target disorder. Consider the young man who presented to your colleague. The fact that he presents with shortness of breath raises the possibility of PE. The fact that he has been immobile for 10 hours increases this probability, but his age, lack of antecedent disease, and normal physical examination, chest radiograph, and arterial blood gas measurements all decrease this probability. If we knew the properties of each of these pieces of information (and for some of them, we do[3,4]), we could move sequentially through them, incorporating each piece of information as we go and continuously recalculating the probability of the target disorder. Clinicians do proceed in this fashion, but because the properties of the individual items of history and physical examination usually are not available, they often must rely on clinical experience and intuition to arrive at the pretest probability that precedes ordering a diagnostic test. For some clinical problems, including the diagnosis of PE, their intuition has proved surprisingly accurate.[2]

Nevertheless, the limited information about the properties of items of history and physical examination often results in clinicians' varying widely in their estimates of pretest probabilities. There are a number of solutions to this problem. First, clinical investigators should study the history and physical exami-

From the Departments of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario.

A complete list of the members (with affiliations) of the Evidence-Based Medicine Working Group appears in the first article of this series (*JAMA*. 1993;270:2093-2095). The following members contributed to this article: Gordon Guyatt (chair), MD, MSc; Eric Bass, MD, MPH; Patrick Brill-Edwards, MD; George Browman, MD, MSc; Deborah Cook, MD, MSc; Michael Farkouh, MD; Hertzel Gerstein, MD, MSc; Brian Haynes, MD, MSc, PhD; Robert Hayward, MD, MPH; Anne Holbrook, MD, PharmD, MSc; Roman Jaeschke, MD, MSc; Elizabeth Juniper, MCSP, MSc; Hui Lee, MD, MSc; Mitchell Levine, MD, MSc; Virginia Moyer, MD, MPH; Jim Nishikawa, MD; Andrew Oxman, MD, MSc, FACPM; Ameen Patel, MD; John Philbrick, MD; W. Scott Richardson, MD; Stephane Sauve, MD, MSc; David Sackett, MD, MSc; Jack Sinclair, MD; K.S. Trout, FRCE; Peter Tugwell, MD, MSc; Sean Tunis, MD, MSc; Stephen Walter, PhD; and Mark Wilson, MD, MPH.

Reprint requests to McMaster University Health Sciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5 (Dr Guyatt).

Table 1.—Evaluating and Applying the Results of Studies of Diagnostic Tests

Are the results of the study valid?
Primary guides:
Was there an independent, blind comparison with a reference standard?
Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?
Secondary guides:
Did the results of the test being evaluated influence the decision to perform the reference standard?
Were the methods for performing the test described in sufficient detail to permit replication?
What are the results?
Are likelihood ratios for the test results presented or data necessary for their calculation provided?
Will the results help me in caring for my patients?
Will the reproducibility of the test result and its interpretation be satisfactory in my setting?
Are the results applicable to my patient?
Will the results change my management?
Will patients be better off as a result of the test?

Table 2.—Test Properties of Ventilation-Perfusion (V/Q) Scanning

	Pulmonary Embolism				
	Present		Absent		
V/Q Scan Result	No.	Proportion	No.	Proportion	Likelihood Ratio
High probability	102	102/251 = 0.406	14	14/630 = 0.022	18.3
Intermediate probability	105	105/251 = 0.418	217	217/630 = 0.344	1.2
Low probability	39	39/251 = 0.155	273	273/630 = 0.433	0.36
Normal/near normal	5	5/251 = 0.020	126	126/630 = 0.200	0.10
Total	**251**	. . .	**630**	. . .	. . .

nation to learn more about the properties of these diagnostic tests. Fortunately, such investigations are becoming common. Panzer and colleagues[5] have summarized much of the available information in the form of a medical text, and overviews on the accuracy and precision of the history and physical examination are being published concurrently with the Users' Guides in the *JAMA* series on The Rational Clinical Examination.[6] In addition, for some target disorders such as myocardial ischemia, multivariable analyses can provide physicians with ways of combining information to generate very precise pretest probabilities.[7] Second, when we don't know the properties of history and physical examination we can consult colleagues about their probability estimates; the consensus view is likely to be more accurate than our individual intuition. Finally, when we remain uncertain about the pretest probability, we can assume the highest plausible pretest probability, and the lowest possible pretest probability, and see if this changes our clinical course of action. We will illustrate how one might do this later in this discussion.

Likelihood Ratios.—The clinical usefulness of a diagnostic test is largely determined by the accuracy with which it identifies its target disorder, and the accuracy measure we shall focus on is the LR. Let's now look at Table 2, constructed from the results of the PIOPED study. There were 251 people with angiographically proven PE and 630 people whose angiograms or follow-up excluded PE. For all patients, V/Q scans were classified into four levels, from high probability to normal or near normal. How likely is a high-probability scan among people who do have PE? Table 2 shows that 102 of 251 people (or 0.406) with PE had high-probability scans. How often is the same test result, a high-probability scan, found among people who, although suspected of it, do not have PE? The answer is 14 of 630 or 0.022. The ratio of these two likelihoods is called the LR and for a high-probability scan equals 0.406 divided by 0.022 or 18.3. In other words, a high-probability lung scan is 18.3 times as likely to occur in a patient with, as opposed to a patient without, a PE. In a similar fashion, the LR can be calculated for each level of the diagnostic test result. Each calculation involves answering two questions: first, how likely it is to get a given test result (eg, a low-probability V/Q scan) among people with the target disorder (PE), and second, how likely it is to get the same test result (again, a low-probability scan) among people without the target disorder (no PE). For a low-probability V/Q scan these likelihoods are 39/251 (0.155) and 273/630 (0.433), and their ratio (the LR for a low-probability scan) is 0.36. As shown in Table 2, we can repeat these calculations for the other scan results.

What do all these numbers mean? The LRs indicate by how much a given diagnostic test result will raise or lower the pretest probability of the target disorder. An LR of 1 means that the posttest probability is exactly the same as the pretest probability. Likelihood ratios greater than 1 increase the probability that the target disorder is present, and the higher the LR the greater this increase. Conversely, LRs less than 1 decrease the probability of the target disorder, and the smaller the LR, the greater the decrease in probability and the smaller its final value.

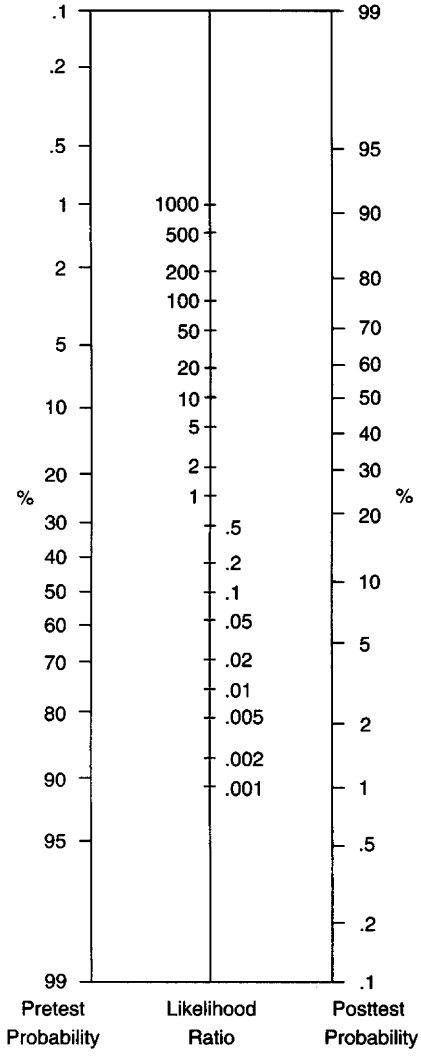
How big is a big LR, and how small is a small one? Using LRs in your day-to-day practice will lead to your own sense of their interpretation, but as a rough guide:

- Likelihood ratios greater than 10 or less than 0.1 generate large and often conclusive changes from pretest to posttest probability.
- Likelihood ratios of 5 to 10 and 0.1 to 0.2 generate moderate shifts in pretest to posttest probability.
- Likelihood ratios of 2 to 5 and 0.5 to 0.2 generate small (but sometimes important) changes in probability.
- Likelihood ratios of 1 to 2 and 0.5 to 1 alter probability to a small (and rarely important) degree.

Having determined the magnitude and significance of the LRs, how do we use them to go from pretest to posttest probability? We can't combine likelihoods directly, the way we can combine probabilities or percentages; their formal use requires converting pretest probability to odds, multiplying the result by the LR, and converting the consequent posttest odds into a posttest probability. While not too difficult,[2] this calculation can be tedious and involves the following: the equation to convert probabilities into odds is (probability/[1−probability]), which is equivalent to probability of having the target disorder/probability of not having the target disorder. A probability of 0.5 represents odds of 0.50/0.50, or 1 to 1; a probability of 0.80 represents odds of 0.80/0.20, or 4 to 1; a probability of 0.25 represents odds of 0.25/0.75, or 1 to 3, or 0.33. Once you have the pretest odds, the posttest odds can be calculated by multiplying the pretest odds by the LR. The posttest odds can be converted back into probabilities using a formula of probability = odds/(odds+1).

Fortunately, there is an easier way. A nomogram proposed by Fagan[8] (Figure) does all the conversions for us and allows us to go very simply from pretest to posttest probability. The first column of this nomogram represents the pretest probability, the second column represents the LR, and the third shows the posttest probability. You obtain the posttest probability by anchoring a ruler at the pretest probability and rotating it until it lines up with the LR for the observed test result.

Recall our elderly woman with suspected PE after abdominal surgery. Most clinicians would agree that the probability of this patient's having PE is quite high, at about 70%. This value then represents the pretest probability. Suppose that her V/Q scan was reported as high probability. Anchoring a ruler at her pre-



Nonogram for interpreting diagnostic test results. Adapted from Fagan.[8]

test probability of 70% and aligning it with the LR of 18.3 associated with a high-probability scan, her posttest probability is over 97%. What if her V/Q scan yielded a different result? If her V/Q scan result is reported as intermediate (LR, 1.2), the probability of PE hardly changes (to 74%), while a near-normal result yields a posttest probability of 19%.

We have pointed out that the pretest probability is an estimate, and that one way of dealing with the uncertainty is to examine the implications of a plausible range of pretest probabilities. Let us assume the pretest probability in this case is as low as 60%, or as high as 80%. The posttest probabilities that would follow from these different pretest probabilities appear in Table 3.

The same exercise may be repeated for our second patient, the young man with nonspecific chest pain and hyperventilation. If you consider that his presentation is compatible with a 20% probability of PE, using our nomogram the posttest probability with a high-probability scan result is 82%, an intermediate probability is 23%, and a near-normal probability is 2%. The pretest probability (with a range of possible pretest probabilites from 10% to 30%), LRs, and posttest probabilities associated with each of the four possible scan results also appear in Table 3.

Table 3.—Pretest Probabilities, Likelihood Ratios (LRs) of Ventilation-Perfusion Scan Results, and Posttest Probabilities in Two Patients With Pulmonary Embolus

Pretest Probability, % (Range)*	Scan Result (LR)	Posttest Probability, % (Range)*
78-Year-Old Woman With Sudden Onset of Dyspnea Following Abdominal Surgery		
70 (60-80)	High probability (18.3)	97 (96-99)
70 (60-80)	Intermediate probability (1.2)	74 (64-83)
70 (60-80)	Low probability (0.36)	46 (35-59)
70 (60-80)	Normal/near normal (0.1)	19 (13-29)
28-Year-Old Man With Dyspnea and Atypical Chest Pain		
20 (10-30)	High probability (18.3)	82 (67-89)
20 (10-30)	Intermediate probability (1.2)	23 (12-34)
20 (10-30)	Low probability (0.36)	8 (4-6)
20 (10-30)	Normal/near normal (0.1)	2 (1-4)

*The values in parentheses represent a plausible range of pretest probabilities. That is, while the best guess as to the pretest probability is 70%, values of 60% to 80% would also be reasonable estimates.

Readers who have followed the discussion to this point will understand the essentials of interpretation of diagnostic tests and can stop here. They should consider the next section, which deals with sensitivity and specificity, optional. We include it largely because clinicians will still encounter studies that present their results in terms of sensitivity and specificity and may wish to understand this alternative framework for summarizing the properties of diagnostic tests.

Sensitivity and Specificity.—You may have noted that our discussion of LRs ignored any talk of normal and abnormal tests. Instead, we presented four different V/Q scan interpretations, each with their own LR. This is not, however, the way the PIOPED investigators presented their results. They fell back on the older (but less useful) concepts of sensitivity and specificity.

Sensitivity is the proportion of people with the target disorder in whom the test result is positive, and specificity is the proportion of people without the target disorder in whom the test result is negative. To use these concepts we have to divide test results into normal and abnormal; in other words, create a 2×2 table. The general form of a 2×2 table, which we use to understand sensitivity and specificity, is presented in Table 4. Look again at Table 2 and observe that we could transform our 4×2 table into any of three such 2×2 tables, depending on what we call normal or abnormal (or what we call negative and positive test results). Let's assume that we call only high-probability scans abnormal (or positive). The resulting 2×2 table is presented in Table 5.

To calculate sensitivity from the data in Table 2 we look at the number of people with proven PE (251) who were diagnosed as having the target disorder on V/Q scan: 102 – sensitivity of 102/251, or approximately 41% ($a/[a+c]$). To calculate specificity we look at the number of people without the target disorder (630) who were was classified on V/Q scan as normals: 616 – specificity of 616/630, or 98% ($d/[b+d]$). We can also calculate LRs for the positive and negative test results using this cut point, 18.3 and 0.6, respectively.

Let's see how the test performs if we decide to put the threshold of positive vs negative in a different place in Table 2. For example, let's call only the normal/near-normal V/Q scan result negative. This 2×2 table (Table 6) shows the sensitivity is now 246/251, or 98% (among 251 people with PE, 246 are diagnosed on V/Q scan), but what has happened to specificity? Among 630 people without PE, only 126 have a negative test result (specificity of 20%). The corresponding LRs are 1.23 and 0.1. Note that with this cut we not only lose the diagnostic information associated with the high-probability scan result, but also interpret intermediate- and low-probability results as if they increase the likelihood of PE, when in fact they decrease the likelihood. You can generate the third 2×2 table by setting the cut point in the middle—if your sensitivity and specificity are 82% and 63%, respectively, and associated LRs of a positive and negative test 2.25 and 0.28, you have it right. (If you were to create a graph where the vertical axis will denote sensitivity [or true-positive rate] for different cutoffs and the horizontal axis will display 1–specificity [or false-posi-

tive rate] for the same cutoffs, and you connect the points generated by using cut points, you would have what is called a receiver operating characteristic [ROC curve]; an ROC curve can be used to formally compare the value of different tests by examining the area under each curve, but once one has abandoned the need for a single cut point, it has no other direct clinical application.)

You can see that in using sensitivity and specificity you have to either throw away important information or recalculate sensitivity and specificity for every cut point. We recommend the LR approach because it is simpler and more efficient.

Thus far, we have established that the results are likely true for the people who were included in the PIOPED study, and ascertained the LRs associated with different results of the test. How useful is the test likely to be in our clinical practice?

WILL THE RESULTS HELP ME IN CARING FOR MY PATIENT?

Will the Reproducibility of the Test Result and Its Interpretation Be Satisfactory in My Setting?

The value of any test depends on its ability to yield the same result when reapplied to stable patients. Poor reproducibility can result from problems with the test itself (eg, variations in reagents in radioimmunoassay kits for determining hormone levels). A second cause for different test results in stable patients arises whenever a test requires interpretation (eg, the extent of ST-segment elevation on an electrocardiogram). Ideally, an article about a diagnostic test will tell readers how reproducible the test results can be expected to be. This is especially important when expertise is required in performing or interpreting the test (and you can confirm this by recalling the clinical disagreements that arise when you and one or more colleagues examine the same electrocardiogram, ultrasound, or computed tomographic scan, even when all of you are experts).

If the reproducibility of a test in the study setting is mediocre and disagreement between observers is common, and yet the test still discriminates well between those with and without the target condition, it is very useful. Under these circumstances, it is likely that the test can be readily applied to your clinical setting. If reproducibility of a diagnostic test is very high and observer variation very low, either the test is simple and unambigious or those interpreting it are highly skilled. If the latter applies, less skilled interpreters in your own clinical setting may not do as well.

Table 4.—Comparison of the Results of Diagnostic Test With the Result of Reference Standard*

	Reference Standard	
Test Result	Disease Present	Disease Absent
Disease present	True positive (*a*)	False positive (*b*)
Disease absent	False negative (*c*)	True negative (*d*)

*Sensitivity = $a/(a+c)$.
Specificity = $d/(b+d)$.
Likelihood ratio for positive test result = $[a/(a+c)]/[b/(b+d)]$.
Likelihood ratio for negative test result = $[c/(a+c)]/[d/(b+d)]$.

In the PIOPED study, the authors not only provided a detailed description of their diagnostic criteria for V/Q scan interpretation, they also reported on the agreement between their two independent readers. Clinical disagreements over intermediate- and low-probability scans were common (25% and 30%, respectively), and they resorted to adjudication by a panel of experts.

Are the Results Applicable to My Patient?

The issue here is whether the test will have the same accuracy among your patients as was reported in the article. Test properties may change with a different mix of disease severity or a different distribution of competing conditions. When patients with the target disorder all have severe disease, LRs will move away from a value of 1 (sensitivity increases). If patients are all mildly affected, LRs move toward a value of 1 (sensitivity decreases). If patients without the target disorder have competing conditions that mimic the test results seen in patients who do have the target disorder, the LRs will move closer to 1 and the test will appear less useful. In a different clinical setting in which fewer of the nondiseased have these competing conditions, the LRs will move away from 1 and the test will appear more useful.

The phenomenon of differing test properties in different subpopulations has been most strikingly demonstrated for exercise electrocardiography in the diagnosis of coronary artery disease. For instance, the more extensive the severity of coronary artery disease, the larger are the LRs of abnormal exercise electrocardiography for angiographic narrowing of the coronary arteries.[9] Another example comes from the diagnosis of venous thromboembolism, where compression ultrasound for proximal-vein thrombosis has proved more accurate in symptomatic outpatients than in asymptomatic postoperative patients.[10]

Sometimes, a test fails in just the patients one hopes it will best serve. The LR of a negative dipstick test for the rapid diagnosis of urinary tract infection is approximately 0.2 in patients with clear symptoms and thus a high probability of urinary tract infection, but is over 0.5 in those with low probability,[11] rendering it of little help in ruling out infection in the latter, low-probability patients.

Table 5.—Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only High-Probability Scans Are Positive (Truly Abnormal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High probability	102	14
Others	149	616
Total	**251**	**630**

*Sensitivity, 41%; specificity, 98%; likelihood ratio of a high-probability test result, 18.3; likelihood ratio of other results, 0.61.

Table 6.—Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only Normal/Near-Normal Scans Are Negative (Truly Normal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High, intermediate, and low probability	246	504
Near normal/normal	5	126
Total	**251**	**630**

*Sensitivity, 98%; specificity, 20%; likelihood ratio of high, intermediate, and low probability, 1.23; likelihood ratio of near normal/normal, 0.1.

If you practice in a setting similar to that of the investigation and your patient meets all the study inclusion criteria and does not violate any of the exclusion criteria, you can be confident that the results are applicable. If not, a judgment is required. As with therapeutic interventions, you should ask whether there are compelling reasons why the results should not be applied to your patients, either because the severity of disease in your patients, or the mix of competing conditions, is so different that generalization is unwarranted. The issue of generalizability may be resolved if you can find an overview that pools the results of a number of studies.

The patients in the PIOPED study were a representative sample of patients with suspected PE from a number of large general hospitals. The results are therefore readily applicable to most clinical practices in North America. There are groups to whom we might be reluctant to generalize the results, such as critically ill patients (who were excluded from the study, and who are likely to have a different spectrum of competing conditions than other patients).

Will the Results Change My Management?

It is useful in making, learning, teaching, and communicating management decisions to link them explicitly to the probability of the target disorder. Thus, for any target disorder there are probabilities below which a clinician would dismiss a diagnosis and order no further tests (a "test" threshold). Similarly, there are probabilities above which a clinician would consider the diagnosis confirmed, and would stop testing and initiate treatment (a "treatment" threshold). When the probability of the target disorder lies between the test and treatment thresholds, further testing is mandated.[12]

Once we decide what our test and treatment thresholds are, posttest probabilities have direct treatment implications. Let us suppose that we are willing to treat those with a probability of PE of 80% or more (knowing that we will be treating 20% of our patients unnecessarily). Furthermore, let's suppose we are willing to dismiss the diagnosis of PE in those with a posttest probability of 10% or less. You may wish to apply different numbers here; the treatment and test thresholds are a matter of judgment, and differ for different conditions depending on the risks of therapy (if risky, you want to be more certain of your diagnosis) and the danger of the disease if left untreated (if the danger of missing the disease is high—such as in PE—you want your posttest probability very low before abandoning the diagnostic search). In our young man, a high-probability scan results in a posttest probability of 82% and may dictate treatment (or, at least, further investigation), an intermediate-probability scan (23% posttest probability) will dictate further testing (perhaps bilateral leg venography, serial impedance plethysmography or ultrasound, or pulmonary angiography), while a low-probability or normal scan (probabilities of <10%) will allow exclusion of the diagnosis of PE. In the elderly woman, a high-probability scan dictates treatment (97% posttest probability of PE), an intermediate result (74% posttest probability) may be compatible with either treatment or further testing (likely a pulmonary angiogram), while any other result mandates further testing.

If most patients have test results with LRs near 1, the test will not be very useful. Thus, the usefulness of a diagnostic test is strongly influenced by the proportion of patients suspected of having the target disorder whose test results have very high or very low LRs so that the test result will move their probability of disease across a test or treatment threshold. In the patients suspected of having PE in our V/Q scan example, review of Table 2 allows us to determine the proportion of patients with extreme results (either high probability with an LR of over 10, or near-normal/normal scans with an LR of 0.1). The proportion can be calculated as (102+14+5+126)/881 or 247/881=28%. Clinicians who have repeatedly been frustrated by frequent intermediate- or low-probability results in their patients with suspected PE will already know that this proportion (28%) is far from optimal. Thus, despite the high LR associated with a high-probability scan, and the low LR associated with a normal/near-normal result, V/Q scanning is of limited usefulness in patients with suspected PE.

A final comment has to do with the use of sequential tests. We have demonstrated how each item of history, or each finding on physical examination, represents a diagnostic test. We generate pretest probabilities that we modify with each new finding. In general, we can also use laboratory tests or imaging procedures in the same way. However, if two tests are very closely related, application of the second test may provide little or no information, and the sequential application of LRs will yield misleading results. For instance, once one has the results of the most powerful laboratory test for iron deficiency, serum ferritin, additional tests such as serum iron or transferrin saturation add no further information.[13]

Will Patients Be Better Off as a Result of the Test?

The ultimate criterion for the usefulness of a diagnostic test is whether it adds information beyond that otherwise available, and whether this information leads to a change in management that is ultimately beneficial to the patient.[14] The value of an accurate test will be undisputed when the target disorder, if left undiagnosed, is dangerous, the test has acceptable risks, and effective treatment exists. High probability or near-normal/normal results of a V/Q scan may well eliminate the need for further investigation and result in anticoagulants' being appropriately given or appropriately withheld (either course of action having a substantial influence on patient outcome).

In other clinical situations, tests may be accurate, and management may even change as a result of their application, but their impact on patient outcome may be far less certain. Examples include right heart catheterization for many critically ill patients, or the incremental value of magnetic resonance imaging scanning over computed tomography for a wide variety of problems.

HOW YOU CAN USE THESE GUIDES FOR CLINICAL PRACTICE AND FOR READING

By applying the principles described in this and the preceding article you will be able to assess and use information from articles about diagnostic tests. You are now equipped to decide whether an article concerning a diagnostic test represents a believable estimate of the true value of a test, what the test properties are, and the circumstances under which the test should be applied to your patients. Because LRs are now being published for an increasing number of tests,[5] the approach we have outlined has become directly applicable to the day-to-day practice of medicine.

References

1. Jaeschke R, Guyatt G, Sackett DL, for the Evidence-Based Working Group. Users' guides to the medical literatuere, III: how to use an article about a diagnostic test, A: are the results of the study valid? *JAMA.* 1994;271:389-391.
2. The PIOPED Investigators. Value of ventilation/perfusion scan in acute pulmonary embolism: results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). *JAMA.* 1990; 263:2753-2759.
3. Mayeski RJ. Pulmonary embolism. In: Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems.* Philadelphia, Pa: American College of Physicians; 1991.
4. Stein PD, Terrin NL, Hales CA, et al. Clinical, laboratory, roentgenographic, and electrocardiographic findings in patients with acute pulmonary embolism and no pre-existing cardiac or pulmonary disease. *Chest.* 1991;100:598-603.
5. Panzer RJ, Black ER, Griner PF. *Diagnostic Strategies for Common Medical Problems.* Philadelphia, Pa: American College of Physicians; 1991.
6. Sackett DL, Rennie D. The science and art of the clinical examination. *JAMA.* 1992;267:2650-2652.
7. Pozen MW, D'Agostino RB, Selker HP, et al. A predictive instrument to improve coronary care-unit admission practices in acute ischemic heart disease. *N Engl J Med.* 1984;310:1273-1278.
8. Fagan TJ. Nomogram for Bayes's theorem (C). *N Engl J Med.* 1975;293:257.
9. Hlatky MA, Pryor DB, Harrell FE. Factors affecting sensitivity and specificity of exercise electrocardiography. *Am J Med.* 1984;77:64-71.
10. Ginsberg JS, Caco CC, Brill-Edwards PA, et al. Venous thrombosis in patients who have undergone major hip or new surgery: detection with compress US and impedance plethysmography. *Radiology.* 1991;181:651-654.
11. Lachs MS, Nachamkin I, Edelstein PH, et al. Spectrum bias in the evaluation of diagnostic tests: lessons from the repid dipstick test for urinary tract infection. *Ann Intern Med.* 1992;117:135-140.
12. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* 2nd ed. Boston, Mass: Little Brown & Co Inc; 1991:145-148.
13. Guyatt GH, Oxman A, Ali M. Diagnosis of iron deficiency. *J Gen Intern Med.* 1992;7:145-153.
14. Guyatt GH, Tugwell P, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J.* 1986; 134:587-594.