APPLYING EVIDENCE IN THE DIAGNOSIS OF A PATIENT WITH SHORTNESS OF BREATH: HOW TO EVALUATE A NEW DIAGNOSTIC TEST

Case and Discussion: *Rhodes S. Adler, MD, Deborah Korenstein, MD, Cheryl Goldstein, MD, and Thomas G. McGinn, MD, MPH*

Feature Editor: *John G. Frohna, MD, MPH*

As clinicians, we routinely rely on our clinical skills in combination with information from diagnostic tests to establish accurate diagnoses, so that we may initiate appropriate treatment for our patients. As technology rapidly expands, however, we face the daunting task of evaluating the usefulness of promising new diagnostic tests when applying them to patient care.

This article uses a patient scenario as the basis for examining how to obtain and incorporate the available medical evidence when faced with a diagnostic dilemma. The evidence-based medicine (EBM) concepts presented help to illuminate the often intuitive methods routinely used in clinical practice when generating a differential diagnosis and testing diagnostic hypotheses. The article begins with the clinical assessment of an elderly patient who presents to the emergency department (ED) with dyspnea and progresses through an analysis of how to estimate pretest probabilities for the diagnoses in a differential. This is followed by a brief discussion of how to critically appraise a published study about a new diagnostic test and a demonstration of the usefulness of likelihood ratios when applying diagnostic tests to individual patients.

Assessing the Clinical Problem

Ms. Baxter is an 82-year-old nursing home resident who was sent to the ED because of increasing shortness of breath over the previous 24 hours. You are a medical resident who is working from 8 PM to 8 AM in the ED, and you begin to evaluate Ms. Baxter when she arrives. From the transfer information sent by the nursing home, you read that her medical problems include chronic obstructive pulmonary disease (COPD), coronary artery disease (CAD), and dementia. Her medications are inhaled steroids and ipratropium with a nebulized beta-agonist as needed, a long-acting nitrate, aspirin, and metoprolol 25 mg twice daily.

On examination, Ms. Baxter is alert and knows she is in the hospital. She is able to follow simple instructions but otherwise cannot give any details of her present illness. You turn your focus to her physical examination in order to determine the cause of her shortness of breath. She seems comfortable sitting at 45 degrees, but when you lower the head of the stretcher to examine her abdomen, she starts coughing. Her blood pressure is 138/90 mm Hg, pulse is 68 bpm and regular, respiratory rate is 24 breaths/min, and temperature is 98.2 ° F. Oxygen saturation is 92% on 2 L. Neck examination shows jugular venous distension (JVD) to 2 cm above her clavicle at a 45-degree angle. Her chest is moderately kyphotic. Her point of maximal impulse (PMI) is not palpable, and her heart sounds are audible but distant. There is no third heart sound (S_3). She has a systolic murmur along the left sternal border that does not radiate to the carotids or cardiac apex. Pulmonary examination reveals diminished breath sounds, faint wheezes, end-expiratory scattered rhonchi, and bibasilar rales. There is no dullness to percussion of the chest. Her liver span is normal by percussion and not palpable below the right costal margin. She has 1+ pretibial edema of her lower extremities.

You determine you need additional history and contact the nursing home. Although the nurse on duty is not familiar with Ms. Baxter, he is able to relay a few details from her chart. Her current symptoms appear to have progressed over several days even though she received all of her prescribed medicines. Her appetite was notably decreased, and when her respiratory rate was noted to be 30 breaths/min with a pulse oximetry reading of 89% on room air, EMS

Rhodes S. Adler, MD, Deborah Korenstein, MD, Cheryl Goldstein, MD, and Thomas G. McGinn, MD, MPH; all at the Division of General Internal Medicine, Mount Sinai School of Medicine, New York, NY; and John G. Frohna, MD, MPH, Departments of Internal Medicine and Pediatrics, University of Michigan Medical School, Ann Arbor, MI.

was called. The nurse also informs you that the patient has no documented allergies and has a 40 pack-year smoking history.

After examining Ms. Baxter, you think about the differential diagnosis for her shortness of breath. Your differential includes COPD exacerbation, with or without an infection, congestive heart failure (CHF) secondary to myocardial ischemia or valvular heart disease, and less likely causes such as pulmonary embolism or sepsis. Reviewing what you have learned from Ms. Baxter's nursing home records and from the findings on her physical examination, a pulmonary infection seems unlikely since her cough is nonproductive and she is afebrile. COPD exacerbation seems likely even without obvious signs of infection, and this diagnosis is supported by several findings on physical examination, including wheezing, rhonchi, decreased breath sounds, and a diffuse PMI. CHF seems likely as well, especially with the patient's history of CAD, and is supported by the findings of JVD, bibasilar rales, and peripheral edema on exam. The murmur is likely from tricuspid regurgitation and is an unlikely cause of her acute shortness of breath (3-day progression). Pulmonary embolism and sepsis both seem quite unlikely. Having narrowed the diagnostic possibilities, you conclude that the most likely cause is systolic failure secondary to acute or chronic myocardial ischemia or a COPD exacerbation.

You discuss your working diagnoses with your ED attending, who examines the patient and confirms your findings. She then asks you to estimate the likelihood of CHF in Ms. Baxter, and you estimate a 50% chance that her symptoms are due to CHF. Your attending thinks for minute. "I think it's more like a 30% chance this is CHF," she says.

Generating a Differential Diagnosis

The process of diagnosis starts as soon as we meet a patient and have gathered enough information to begin to generate a differential diagnosis. The generation of a differential diagnosis relies upon us having both general medical knowledge and knowledge of disease prevalence and patterns of clinical presentation. Key features derived from the history, in conjunction with physical examination findings, serve to either increase or decrease the likelihood of each diagnosis under consideration in a process of hypothesis testing. In addition, we often can use our previous clinical experience with similar cases to inform our differential diagnosis. Clinical experience is a powerful tool in the diagnostic process, yet it must be grounded with accurate background knowledge and clear judgment to minimize its inherent bias. In the case of Ms. Baxter, who has a history of both cardiac and pulmonary disease, the information obtained from her history and physical examination will require careful scrutiny to determine the origin of her dyspnea.

Estimating Pretest Probability

In establishing hypotheses, it is important that we accurately assess the *pretest probability* of the diagnoses we are considering, so that subsequent testing can help us to rule in or rule out those possibilities. What exactly is the pretest probability, and how do we estimate it?

> Pretest probability is the proportion of patients who have the target disorder (prevalence), as determined before the test is carried out.

The process of estimating pretest probability can be complex but is often subconscious and linked closely with the process of differential diagnosis. The simplest method is to assess the prevalence of each possible diagnosis, either in the community at large or in patients in our particular setting. This type of information can be found in textbooks or studies of differential diagnosis. For example, from Framingham data [1,2] the disease prevalence of CHF in community-dwelling patients over 75 years of age is as high as 10% and would likely be higher in elderly patients presenting to the ED with shortness of breath. We would then need to adjust this rough estimate of pretest probability using information from the history or physical examination in a manner similar to the hypothesis testing used in generating differential diagnoses.

Conceptually, each element of the history and physical examination can be viewed as an independent diagnostic test, the results of which impact the likelihood of a particular disorder. The literature includes formal studies examining the diagnostic accuracy of elements of the history and physical examination, such as the articles periodically published in the "Rational Clinical Examination" series in *JAMA*. A chapter in the *Users' Guide to the Medical Literature* published by the American Medical Association summarizes numerous diagnostic studies based on symptoms and signs [3]. In addition, an extensive list of pretest probabilities for various settings and conditions can be found in the useful book, *Diagnostic Strategies for Common Medical Problems* [4]. Finally, clinical prediction rules give accurate and rapid estimates of a patient's probability of disease. Clinical prediction rules are front-line decision aids that combine various components of the history, physical examination, and simple diagnostic tests to help accurately estimate

Table 1. Internet Collections of Clinical Prediction Rules

Name	Web Site	Platform
InfoRetriever	http://www.InfoPOEMS.com	Internet, PDA
MedRules	http://pbrain.hypermart.net/medrules.html	PDA
The Medical Algorithms Project	http://www.medal.org	MSExcel 97
Mount Sinai EBM Web site	http://med.mssm.edu/ebm	Internet, PDA
Ottawa Health Research Institute	http://www.ohri.ca/programs/clinical_epidemiology/OHDEC/clinical.asp	Internet

PDA = personal digital assistant. (Adapted with permission from McGinn T. Practice corner: using clinical prediction rules. ACP J Club 2002 Sept-Oct;137:A12.)

pretest probability [5]. Collections of clinical prediction rules that help estimate pretest probability also can be found on the Internet (**Table 1**) [6].

When data are not available to assess pretest probability, clinical experience (which incorporates findings on history and physical examination) and clinical judgment are used to estimate pretest probability. Clinical experience is part of the art of medicine and is often the reason behind varying estimates of pretest probability among physicians. Different levels of clinical experience likely explains the difference between the resident's pretest probability estimate of CHF causing Ms. Baxter's shortness of breath and the ED attending physician's pretest probability estimate. The attending, with significantly more training and clinical experience, is able to recall previous patients with similar presenting signs and symptoms and the subsequent outcomes. Combined with her clinical experience, she has estimated the likelihood of CHF in this patient to be greater than the 10% prevalence from the Framingham data, but lower than the likelihood that pulmonary disease is the cause of the patient's dyspnea.

With your 50% probability of CHF in mind, you think about your test and treatment thresholds for CHF in Ms. Baxter. You determine that you would treat her for CHF without further testing at an 80% probability of the disorder, and that you would neither treat nor do further testing for CHF with a 20% probability of the disorder. Because your 50% probability is between your two thresholds, you need more information.

You decide that you would like to consider the results of Ms. Baxter's electrocardiogram (ECG), cardiac enzyme assays, and chest x-ray. The ECG is done before your patient is sent to radiology for the chest film and shows no evidence of acute myocardial infarction. There is poor R wave progression in V_1–V_2, which probably indicates an old myocardial infarction and is consistent with her history of CAD. The results of the cardiac enzyme assays will not be ready for several hours.

When Ms. Baxter returns from radiology, you review the chest film with your attending, which reveals hyperinflated lungs with flattening of the diaphragms, no infiltrates, and a heart size at the upper limits of normal. There is no cephalization of blood flow, and the pulmonary artery shadow is prominent. As you consider these results, you think about how they have changed your original pretest probability of CHF. The ECG did not show an acute myocardial infarction, but systolic failure remains a possibility because the patient's heart is at the upper limits in size despite her hyperinflated lung fields. There is no cephalization, but you know that COPD can produce a paucity of peripheral vascular markings because of emphysematous changes. You are pleased that the chest film has confirmed your earlier impression that Ms. Baxter does not have pneumonia. You conclude that the results of the ECG and chest x-ray have not changed your 50% pretest probability of CHF.

Test and Treatment Thresholds

Once we have determined pretest probability, we must decide whether to initiate treatment or perform further diagnostic tests. Let us imagine a similar patient to Ms. Baxter, who has a history of CHF but not COPD and who presents with dyspnea after misplacing his diuretic pills for 1 week. This patient's examination is significant for JVD, an S_3 gallop, rales half way up bilaterally, and 2+ lower extremity edema. In this case, we would be confident of the diagnosis of CHF and would initiate treatment without further testing. The probability of CHF is so high that it is above a threshold where further testing becomes unnecessary, known as the *treatment threshold*. Similarly, although COPD might appear in the initial differential diagnosis for this patient, we might determine that it is so unlikely to be the cause of his dyspnea that no treatment or testing is

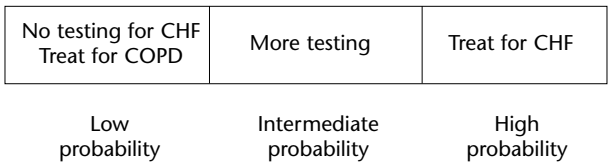

Figure 1. Test and treatment thresholds for case patient, shown at various qualitative estimates of pretest probability of disease. Decisions about where to set test and treatment thresholds are influenced by the potential severity and associated morbidity of the disease. CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease.

necessary. The probability of COPD in this patient would be below the level where testing is necessary, known as the *test threshold*. In Ms. Baxter, however, CHF seems likely but not definite. With an intermediate probability of the diagnosis, further testing is warranted. **Figure 1** illustrates the concept of test and treatment thresholds and how they influence clinical behavior.

Where we define our thresholds varies depending on the severity of the disease and the potential toxicity of the diagnostic test or the treatment. For example, in a patient with a nodular infiltrate on chest x-ray, we might begin antibiotic treatment with only a 50% probability of pneumonia, whereas we would require a 99% to 100% probability of lung cancer before initiating chemotherapy. Similarly, test thresholds vary with regard to the severity of the disorder in question and how responsive it is to treatment.

Asking a Focused Clinical Question

Overall, Ms. Baxter is not deteriorating, and the results of her first set of cardiac enzymes are negative. You decide that you will continue to monitor her until the morning, when her systolic function can be measured with echocardiography. Your attending confirms the difficulty of diagnosing Ms. Baxter's symptoms.

She then says, "I read something 2 or 3 months ago in one of the major journals about a test for B-type natriuretic peptide being able to diagnose CHF in the ED. I think the test was able to identify patients with systolic failure."

You have no knowledge of natriuretic peptide, but you wonder whether this test might rapidly and reliably establish that Ms. Baxter has CHF. You turn to the nearest computer and click on the icon for the medical school library.

Before we search for evidence about B-type natriuretic peptide (BNP), we need to clearly state our clinical question. A well-developed question will enable us to better identify evidence that will be helpful. The PICO framework (patient, intervention, comparison, outcome) is a useful approach for developing a focused clinical question [7]. The PICO components in this case might be defined as:

P = patient presenting to the ED with dyspnea

I = measurement of B-type natriuretic peptide

C = expert judgment based on usual workup

O = accuracy of CHF diagnosis

Using this framework to build a clinical question, our question in Ms. Baxter's case would be:

> In a **patient presenting to the ED with dyspnea**, is **measurement of B-type natriuretic peptide** reliable for **accurately diagnosing CHF**, compared with **expert judgment based on usual workup?**

Acquiring the Current Best Evidence

Because you are interested in a recently published article on a diagnostic test, you realize that filtered sources of evidence are most likely not useful to your search. Thus, you go to MEDLINE and enter "natriuretic peptide" AND "congestive heart failure" AND "diagnosis," limiting the search to studies published in 2002. Your search retrieves 11 items, one of which is from the *New England Journal of Medicine,* entitled "Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure" [8]. This article seems to directly answer your question, and you confirm that this is the article your attending was thinking of. You print the full-text version to read later.

When searching for evidence regarding a new diagnostic test, it is useful to keep in mind the limitations of certain evidence sources. For example, the Cochrane Database of Systematic Reviews, which is a useful filtered source of evidence, currently addresses only therapy-related issues and, therefore, would not help our search. In the near future, however, systematic reviews of diagnostic tests will be added to the Cochrane Database. Similarly, the *ACP Journal Club*, which is a good filtered source of evidence regarding new diagnostic tests, in our case would likely not be helpful because the article we are seeking was too recently published in the primary literature to have been appraised and abstracted in *ACP Journal Club*. It would be even less likely for textbook resources and related electronic references (eg,

UptoDate, *MD Consult*) to have such recent information. Hence, MEDLINE is a reasonable first option for this clinical question. It is important to keep in mind that there are far fewer articles published related to questions of diagnosis as compared to treatment. Therefore, a simple search in this situation, using two terms (eg, "natriuretic peptide" and "congestive heart failure") quickly gives us the article we need.

Appraising Evidence for Validity and Importance

Before you return to the ED the next evening, you read the article you found on MEDLINE and learn from the introduction that BNP is secreted from the ventricles in response to volume expansion and pressure overload. You then apply the criteria for critically appraising the results of a diagnostic study (**Table 2**).

The investigation of BNP was a multicenter, international study that prospectively enrolled 1586 patients who presented to EDs with acute dyspnea and who had BNP levels measured with a bedside assay. The reference standard was clinical diagnosis by two cardiologists who were blinded to the results of the peptide assay and who considered extensive clinical data. The reference standard was applied to all patients. You are satisfied with the validity of the study and decide to move on to the results.

You find that the authors do not report their results in likelihood ratios but do report the sensitivity and specificity for various levels of BNP. The authors report that a cut-off level of 100 pcg/mL best distinguished dyspnea due to CHF from other causes. A BNP level of 100 pcg/mL had a sensitivity of 90% and a specificity of 76%.

You easily calculate the positive likelihood ratio for a BNP level greater than 100 pcg/mL as:

sensitivity/(1 – specificity) = 0.90/0.24 = 3.75

You then calculate the negative likelihood ratio for a BNP value less than 100 pcg/mL as:

(1 – sensitivity)/specificity = 0.10/0.76 = 0.13

Reference Standard

We start assessing the validity of a study of a diagnostic test by asking whether the test in question was independently and blindly compared with a reference standard. To determine the accuracy of a diagnostic test, physicians must ascertain how often the test gives them the "true" answer; that is, how often it correctly diagnoses patients as either having or not having the target disorder. This "truth" of who really has the disease is often difficult to establish. We often think of a *gold standard* test as the best way to determine the diagnosis, but many diseases do not have a gold standard test for their diagnosis. In such cases, expert opinion, some combination of tests, subsequent response to therapy, or long-term follow-up can serve as appropriate reference standards. In this study of BNP, the reference standard was diagnosis by two cardiologists who reviewed all pertinent medical records, including those from the ED and from any subsequent hospitalization. The cardiologists did not have access to the BNP values (blinding). Because CHF is a clinical diagnosis without a gold standard test, this reference standard is reasonable.

Table 2. Criteria for Appraising and Applying the Results of an Individual Diagnostic Study

Are the results valid?
Was there a blind comparison between the test and an independent gold standard?
Was the test evaluated in an appropriate spectrum of patients?
Did the results of the test being evaluated influence the decision to perform the gold standard?
What are the results?
What likelihood ratios were associated with the range of possible test results?
How can I apply the results to patient care?
Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?
Are the results applicable to the patient in my practice?
Will the results change my management strategy?
Will patients be better off as a result of the test?

Adapted with permission from Guyatt G, Rennie D. User's guides to the medical literature. Chicago: AMA Press; 2002:121.

Spectrum of Patients

The next step in looking at the validity of a diagnostic test is to verify that it was evaluated in the appropriate types of patients. For example, let us imagine a study of BNP for the diagnosis of CHF in which investigators enrolled 100 patients with class IV CHF and 100 young healthy patients with upper respiratory infections. They might find that the patients with CHF all have elevated BNP levels and the patients with colds all have low levels, and the test might appear to be very useful. However, this study would give us no information on BNP levels in patients with dyspnea from other causes (eg, COPD, mild CHF, pneumonia, pulmonary embolism), in whom we would need to use the test in practice. Thus, a diagnostic test must be evaluated in patients who are similar, in terms of comorbidities and severity of illness, to the

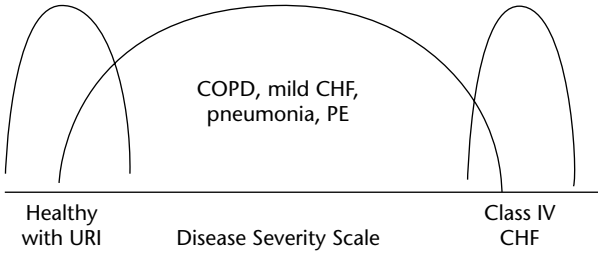

Figure 2. Spectrum of patients presenting to the emergency department with dyspnea. The line along the base of the figure represents increasing severity of disease. Our patient falls within the broad middle range of this spectrum. Thus, a diagnostic test that was evaluated in only patients at the very mild or very severe end of this spectrum of disease would demonstrate artificially elevated sensitivity or specificity and would not be helpful in patients with dyspnea from other causes. CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; PE = pulmonary embolism; URI = upper respiratory infection.

patients in whom it will be applied in practice. This property of the study sample is labeled *spectrum of patients* (**Figure 2**). In this study of BNP, patients older than age 18 years presenting to the ED with the chief complaint of dyspnea were enrolled, and patients with an obvious noncardiac etiology (eg, trauma) or with acute myocardial infarction or renal failure were excluded. We would clinically apply the BNP test to patients such as these, so the spectrum of included patients seems appropriate.

Finally, to truly assess diagnostic accuracy, the reference standard must be applied to all patients, not just patients with a positive test. In this study, the reference standard was, in fact, applied to all enrolled patients.

Likelihood Ratios

It is useful to consider the results of a diagnostic test in terms of *likelihood ratios*. The likelihood ratio for a test result expresses the ratio of the chance of finding that test result in patients with the target disorder to the chance of finding it in patients without the target disorder.

$$\text{likelihood ratio} = \frac{\text{proportion of people } \textbf{with} \text{ disease with a given test result}}{\text{proportion of people } \textbf{without} \text{ disease with that same test result}}$$

The likelihood ratio represents the ability of a test to modify the pretest probability to yield a post-test probability. The post-test probability is the chance that the target disorder is present given the result of the diagnostic test. A positive likelihood ratio is the likelihood ratio for a positive test and can be calculated as:

$$\text{sensitivity}/(1 - \text{specificity})$$

A negative likelihood ratio is the likelihood ratio for a negative test and is calculated as:

$$(1 - \text{sensitivity})/\text{specificity}$$

A likelihood ratio of exactly 1 means that the test does not alter the likelihood of the target disorder; a likelihood ratio greater than 1 increases the probability that the target disorder is present, whereas a likelihood ratio less than 1 decreases the probability of the target disorder. A nomogram allows us to use likelihood ratios to convert our pretest probability to a post-test probability (**Figure 3**). As a reference, a likelihood ratio of 1, which gives us the same post-test probability as our pretest probability, adds little to our workup. By contrast, a positive likelihood ratio of 10 or greater is considered very significant and in most circumstances will rule in a disease. On the other hand, a negative likelihood ratio of less than 0.1 is considered a significant negative value, and in most circumstances will rule out a diagnosis.

Likelihood ratios are independent of disease prevalence and allow us to interpret diagnostic test results differently in individual patients with different pretest probabilities. As we know, physicians do not always generate the same pretest probability when interpreting the same data. Using the likelihood ratios from the BNP article and the two different pretest probabilities for CHF in Ms. Baxter (ie, the resident's and the attending physician's), we can see how the same test result can generate different post-test probabilities.

Let us imagine that an assay of Ms. Baxter's BNP level was ordered in the ED and found to be greater than 100 pcg/mL. What is our new estimate of her likelihood of having CHF? Using the nomogram and drawing a line through 50% (the resident's pretest probability) and 3.75 (the positive likelihood ratio), the resident's post-test probability would be close to 80% (Figure 3). Using the nomogram and drawing a line through 30% (the attending's pretest probability) and 3.75 generates a post-test probability of approximately 60%. What if Ms. Baxter's BNP level was less than 100 pcg/mL, with a negative likelihood ratio of 0.13? Using the nomogram, the resident's post-test probability would be 12% and the attending's would be 4%.

Applying Evidence to Patient Care

Using your calculated likelihood ratios and your 50% pretest probability, you determine that a positive BNP test result in Ms. Baxter would lead you to treat her for

CHF. Out of curiosity, you calculate your attending's post-test probability and determine that if Ms. Baxter's BNP test result were positive, your attending would not treat her for CHF.

This exercise makes you realize the importance of pretest assessment and leads you to further question the attending on how she arrived at her probability. She mentions that she derived her number from her knowledge of the prevalence of CHF in the general population (10%) combined with new knowledge gained from having recently participated in a quality assurance project. As part of the project, she reviewed the prevalence of CHF in patients presenting to the ED, which she learned was approximately 20% overall. You had assumed the baseline rate of CHF in patients presenting to the ED to be much higher. After the discussion with your attending, you agree that your estimate was most likely too high. You also agree that a negative test would have led both of you to treat Ms. Baxter for COPD.

As for the follow-up on Ms. Baxter, her echocardiogram does not show left-ventricular dysfunction. She is treated for a COPD exacerbation and is able to return to the nursing home without being hospitalized.

Recapping our evidence-based approach to this case thus far, we have estimated the pretest probability of CHF in Ms. Baxter, established test and treatment thresholds, identified a test for diagnosing CHF in the ED (BNP measurement) and determined that the test is valid, and calculated likelihood ratios for possible test results. The final EBM step is determining how, if at all, we should apply the evidence we have found to our decisions regarding Ms. Baxter's care. As we can see, these decisions can be influenced significantly by our pretest assessment.

Starting at the resident's pretest probability of 50%, a positive BNP test result would move the probability of CHF to 80%—our treatment threshold—and would, indeed, impact the care of Ms. Baxter by leading to treatment for CHF without further testing. Even with a positive test result, the attending physician would not treat Ms. Baxter for CHF but would perform further diagnostic testing, because the likelihood of CHF would increase to only about 60%—below our treatment threshold. A negative test, on the other hand, would lower the probability of CHF to below our test threshold of 20% whether the pretest probability was 30% or 50% and, thus, would impact the care of the patient by leading to consideration of an alternative diagnosis, in this case COPD. In summary, it appears that the BNP test may be useful in helping to make or rule out the diagnosis of CHF in patients presenting to the ED with an intermediate probability of CHF.

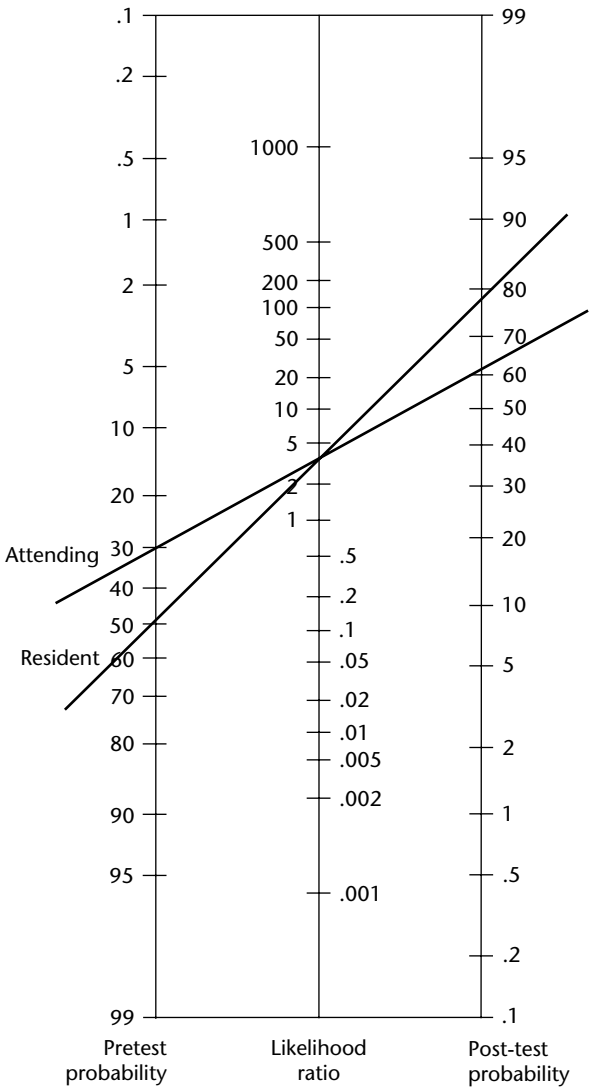

Figure 3. Likelihood ratio nomogram. Lines represent plots of a positive likelihood ratio of 3.75 for the resident and attending physician, based on their respective pretest probabilities of 50% and 30%.

Conclusion

Table 3 summarizes the five-step EBM approach as it applies to the diagnosis of Ms. Baxter's shortness of breath. In this case, we have demonstrated how to use EBM when faced with a diagnostic dilemma and the need to evaluate a new diagnostic test. The first step in making an accurate diagnosis is to integrate evidence

Table 3. Summary of the Evidence-Based Medicine Approach to Diagnosis of Shortness of Breath in Ms. Baxter

Assess	
The patient	Ms. Baxter presents to the ED with increasing shortness of breath over the previous 24 hours. She has a history of CAD, COPD, and dementia. Her physical examination, ECG, and chest x-ray are not diagnostic for either CHF or a COPD exacerbation.
The problem	Is there a test that could rapidly and reliably establish that Ms. Baxter has CHF?
Ask	In a patient presenting to the ED with dyspnea, is measurement of B-type natriuretic peptide reliable for accurately diagnosing CHF, compared with expert judgment based on usual workup?
Acquire	Electronic full-text of diagnostic study from MEDLINE (< 5 minutes to obtain)
Appraise	The diagnostic study on measuring BNP meets the validity criteria. There was an independent blind comparison of BNP measurements with a reference standard that was evaluated in an appropriate spectrum of patients who all had BNP levels measured. The likelihood ratio for a BNP level of < 100 pcg/mL or > 100 pcg/mL discriminates best between dyspnea from systolic failure and other causes of dyspnea.
Apply	For patients like Ms. Baxter, measurement of BNP levels would affect her management because a diagnosis of CHF could be made rapidly, resulting in prompt treatment for either noncardiac or cardiac shortness of breath.

from our knowledge of disease and disease prevalence with a patient's history and physical examination to formulate a differential diagnosis and estimate pretest probability. Several resources can be used to help guide our estimation of pretest probability, such as clinical prediction rules and the differential diagnosis literature. To determine whether or not a new diagnostic test is helpful to us, we must assess the test for validity and applicability to the patient at hand. If the study is valid, likelihood ratios derived from the study allow us to determine post-test probabilities, which are clinically useful if they move the probability of the target disorder either above a treatment threshold or below a test threshold.

This case highlights the impact that pretest probability assessment can have on our final diagnosis. The same positive test result can lead to different clinical decisions and actions depending on how well we assess pretest probability. As new diagnostic tests proliferate, EBM provides us with the tools to assess their power and the impact they should have on the patient care we provide.

Address correspondence to: Rhodes S. Adler, MD, Mount Sinai School of Medicine, 1470 Madison Ave., Box 1087, New York, NY 10029 (e-mail: Rhodes.Adler@mssm.edu).

References

1. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol 1993;22:6A–13A.
2. Cowie MR, Mosterd A, Wood DA, et al. The epidemiology of heart failure. Eur Heart J 1997;18:208–25.
3. Guyatt G, Rennie D. Users' guides to the medical literature. Chicago: AMA Press; 2002.
4. Black ER, Bordley DR, Tape TG, Panzer RJ, editors. Diagnostic strategies for common medical problems. 2nd edition. Philadelphia: American College of Physicians; 1999.
5. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature. XXII. How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2000;284:79–84.
6. McGinn TG. Practice corner: using clinical prediction rules. ACP J Club 2002 Sept-Oct;137:A11–2.
7. Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions [editorial]. ACP J Club 1995 Nov-Dec;123:A12.
8. Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002; 347:161–7.