Annals Clinical Decision Making: Using a Diagnostic Test

Katrina A. Armstrong, MD, MS, and Joshua P. Metlay, MD, PhD

Diagnostic tests are important tools in clinical decision making (1, 2). However, they are often misunderstood and misused (3). This article uses 2 clinical examples to review the principles of interpreting diagnostic test results and outline an approach to the appropriate selection and application of tests in internal medicine (Table 1).

Case 1

A 50-year-old man presented to urgent care in Boston in late May with 5 days of cough, anorexia, and fever after returning from a trip to India. Results of the patient's rapid influenza test were positive, and he was sent home to recover. Three days later, he returned with a persistent fever and was admitted to the medical service, where he was diagnosed with typhoid fever.

This case illustrates what can go wrong when diagnostic test results are misinterpreted. The goal of a diagnostic test is to help a clinician decide whether a patient has a certain condition. The ideal test would be perfectly accurate: If the results are positive, the patient has the condition; if the results are negative, the patient does not have the condition. Of course, few if any diagnostic tests meet these criteria. The ability of a test to determine whether a patient has a condition—that is, accuracy—is often described by sensitivity and specificity. Sensitivity is the proportion of patients with the condition who have positive test results, and specificity is the proportion of patients without the condition who have negative test results. However, the more clinically useful information is the probability that a patient with a positive test result actually has the condition, known as positive predictive value (PPV), and the probability that a patient with a negative test result does not have the condition, known as negative predictive value (NPV). Although sensitivity and specificity are characteristics of the test, PPV and NPV depend on the probability that the patient has the condition before testing, called the pretest probability. Pretest probability led to the problem with clinical decision making in this case.

How Can I Estimate the Likelihood That My Patient Actually Has the Condition Once I Know the Test Result?

There are several approaches to estimating the posttest probability that a patient has a condition (that is, the PPV or NPV) given the test results, sensitivity and specificity, and pretest probability. One approach is to input sensitivity and specificity and pretest probability into a 2 × 2 table and calculate PPV and NPV (Table 2). Another approach is to use the Bayes theorem, which uses likelihood ratios derived from sensitivity and specificity to translate pretest odds into posttest odds. Because odds are hard to use, the Bayes theorem is most often used via a nomogram (Figure 1). Although numerical estimates can be useful in some settings, it is important to remember that a positive test result in a patient with a low probability of disease does not mean that they have the disease unless specificity is approximately 100%, and a negative test result in a patient with a high probability of disease does not exclude the disease unless sensitivity is approximately 100%. In other words, beware of false-positive results when the prevalence of a condition is low and false-negative results when the prevalence is high.

How Can I Assess Pretest Probability, Given Its Importance in Interpreting the Test Result?

Ideally, pretest probability would be derived from information about the prevalence of a given diagnosis among a group of patients with a similar set of clinical signs, symptoms, and risk factors. For some clinical decisions, local, state, or national epidemiologic surveillance data about the frequency of the condition in patients similar to the person being tested are available and can be directly translated into pretest probability. For this case, the state health department provided data about the prevalence of influenza among patients with influenza-like illness by week of the year. This estimate can be adjusted on the basis of patient history or other factors. For example, the pretest probability of influenza may be lower if the patient had received a flu shot or higher if he had been exposed to a case of influenza. Sometimes this information is available in the published literature. Often, it comes from clinical experience. Recognition of cognitive biases can reduce their effect on the estimation of pretest probability from clinical experience (discussed in another paper in this series [5]).

In summary, the interpretation of diagnostic test results depends on understanding the accuracy of the test and the probability that the patient has the condition before testing. Because nearly all tests are imperfect, errors in interpretation most often arise from overestimating the effect of a test result on the posttest probability of a condition.

Case 2

A 32-year-old woman was admitted to the medical service with progressive fatigue during the past year and new shortness of breath and headaches. She took

See also:

Editorial comment 627

This article is 1 of 6 within the *Annals* Clinical Decision Making series. The series is available at Annals.org.

This article was published at Annals.org on 21 April 2020.

Table 1. Approach to Use of Diagnostic Tests

1. Develop a list of potential diagnoses, noting those that can have severe consequences if not diagnosed and treated expeditiously.
2. Estimate the likelihood of these diagnoses on the basis of the clinical scenario and the frequency of the conditions in your patient population.
3. Identify the potential diagnostic tests for the conditions above, particularly those that are most common or most serious.
4. Estimate the accuracy of the tests, most often categorizing a test result as having a large, moderate, or small effect on the probability of a condition, as well as identifying any clinically significant potential harms.
5. For a given condition, decide on a testing or treatment strategy, using a diagnostic test when it has sufficient accuracy to move a patient above or below the probability threshold at which treatment would be started and when empirical treatment or observation without treatment is too high risk.
6. Focus first on conditions consistent with the patient's presentation if missing the diagnosis would have rapid and severe adverse consequences, and then on the most likely conditions. Consider excluding less common, treatable conditions that have a high-accuracy, low-risk, low-cost test.
7. Engage patients in decisions about whether to proceed with a test, treatment, or observation, particularly when the alternatives involve clinically significant risks.

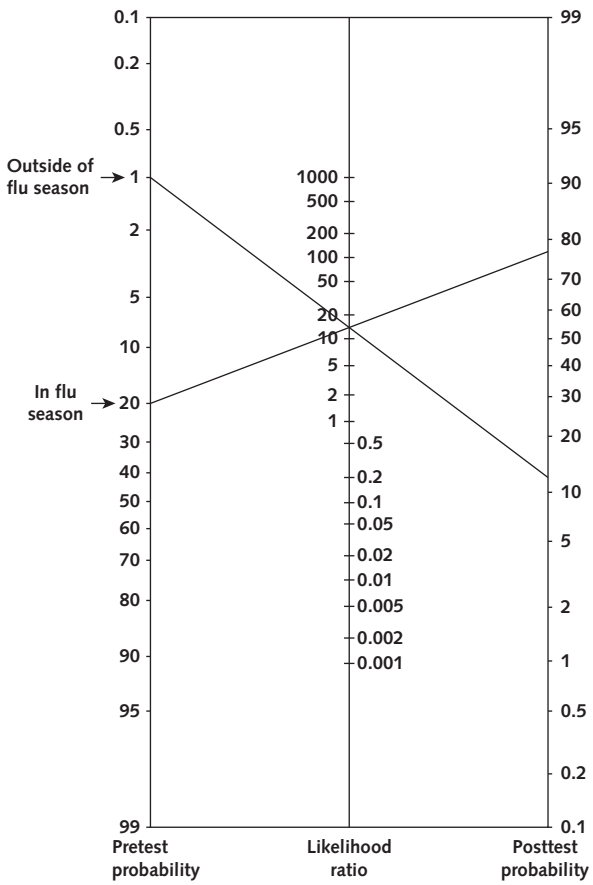
no medications and had no other medical history. Physical examination was notable for blood pressure of 145/111 mm Hg, heart rate of 116 beats/min, bibasilar crackles, and bilateral pitting edema to her shins. She had a blood urea nitrogen level of 112 mg/dL, creatinine level of 5.99 mg/dL, leukocytosis with a left shift, and mild metabolic acidosis. Total protein and albumin levels were low. Urinalysis results were abnormal, with 3+ blood and 3+ protein on dipstick, greater than 100 erythrocyte count per high-power field, 50 to 100 leukocyte count per high-power field, and a few granular casts on microscopic examination. The medical team identified acute renal failure as the primary medical problem, and they faced a long list of potential diagnostic tests.

Table 2. Posttest Probabilities of Influenza*

Clinical Scenario	Rapid Influenza Test Result, *n*		Prevalence of Disease, *n*	
	Positive	Negative		
For 1000 cases like our patient during influenza season†				
Has influenza	140	60	200	
Does not have influenza	40	760	800	
Total	180	820		–
For 1000 cases like our patient outside of influenza season‡				
Has influenza	7	3	10	
Does not have influenza	50	940	990	
Total	57	943		–

* Given that the rapid influenza test had a sensitivity of 70% and specificity of 95% and that the probability of influenza among patients with influenza-like illness is approximately 20% in influenza season but less than 1% outside of influenza season (4).
† The posttest probability that a patient with a positive test result has a condition (i.e., the positive predictive value of the test) is calculated as: Number with a positive test result who have the condition/total number with a positive test result.
Here, a positive test result indicates a 140 out of 180 or 78% probability of having influenza.
‡ A positive test result indicates a 7 out of 57 or 12% probability of having influenza.

Figure 1. Nomogram for estimating posttest probability.



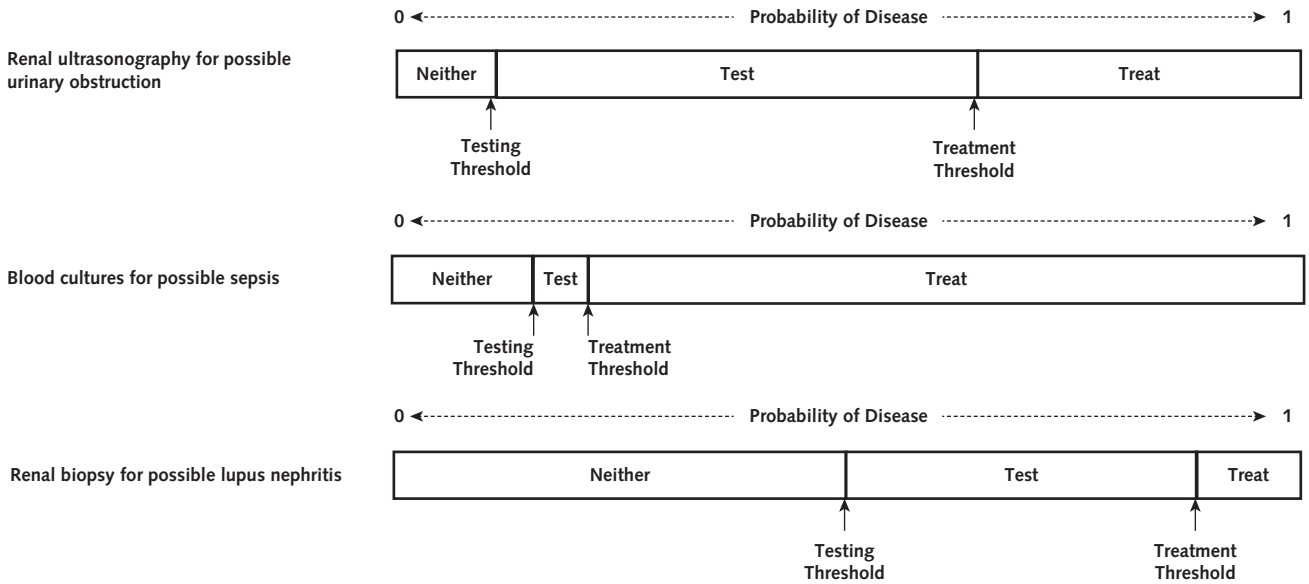
This nomogram uses the pretest probability of disease and the test's likelihood ratio (LR) to estimate the posttest probability. A straight edge is used to extend a line from the appropriate pretest probability through the appropriate LR to identify the posttest probability. For example, a rapid influenza test has a positive LR of 14, meaning that the odds of disease are increased 14-fold by a positive test result, and a negative LR of 0.32, meaning that the odds of disease are 0.32 times lower with a negative test result. The nomogram shows that a positive test result when the prevalence of influenza is 20% (that is, pretest probability) indicates that the patient has a greater than 70% chance of having influenza, whereas a positive test result when the prevalence is 1% indicates that the patient has only a 12% chance of having influenza.

Many clinical scenarios require development of a diagnostic testing approach that spans several potential diagnoses, tests, and empirical treatment strategies. The overarching goal of such an approach is to provide the necessary information for clinical decisions while limiting the harms of testing, including the risks associated with test procedures and false-positive results, as well as the cost to the patient and to society (a consideration of multiple perspectives that contribute to clinical decision making is provided in another paper in this series [6]).

How Do I Decide Which Conditions to Test For?

The first step in any diagnostic testing approach is to create a list of potential diagnoses. An accurate,

Figure 2. Test and treatment thresholds.



The decision about whether to pursue a diagnostic test as well as initiation of treatment requires consideration of the risks of each relative to the clinical suspicion for the disease and its prognosis if untreated. Renal ultrasonography for possible urinary obstruction: minimal-risk, high-accuracy test; poor disease prognosis if untreated; low-risk, high-effectiveness invasive therapy (*top*). Blood cultures for possible sepsis: minimal-risk, low-accuracy test at presentation; very poor disease prognosis if untreated; low-risk, high-effectiveness, noninvasive therapy (*middle*). Renal biopsy for possible lupus nephritis: moderate-risk, high-accuracy test; poor disease prognosis if untreated; high-risk, moderate-effectiveness therapy (*bottom*).

comprehensive list is often compiled from a combination of pathophysiologic understanding, published literature, and clinical experience. For example, in this case, the team recognized that causes of acute renal failure included pre-, intra-, and postrenal conditions and generated a long list of potential diagnoses under each category. They then reviewed the literature to ensure that the list was complete and to estimate the probability of each diagnosis. They found that a study of 748 patients with acute renal failure admitted to a medical service in Madrid in 1991 reported that 45% of patients had acute tubular necrosis, 21% had hypoperfusion, 13% had diabetes or hypertensive nephropathy, 10% had obstruction, 2% had interstitial nephritis, 2% had vasculitis, and 7% had other conditions (7). Finally, given that advances in electronic medical record–based and computer-based search and analysis strategies are increasingly providing access to real-time information about confirmed diagnoses in patients with similar presentations at the same institution or local region, the team also conducted an electronic medical record query of discharge diagnoses of patients admitted with acute renal failure to the medical services at their hospital. The search suggested that diagnoses of sepsis, congestive heart failure, and other forms of glomerulonephritis, such as IgA nephropathy and lupus nephritis, should be added to the differential diagnosis, although information on the relative likelihood of these diagnoses was not available.

How Do I Learn About the Tests for the Conditions I Am Concerned About and the Complications that a Test May Cause?

Finding information on diagnostic tests can be even more challenging than finding the probabilities of different diagnoses. The number of diagnostic tests has proliferated during the past decades, with a 2012 estimate of more than 40 000 laboratory tests alone (3). Furthermore, many laboratory and imaging tests have increased in complexity and have different protocols depending on what information is most important.

Many clinicians gain experience with a range of diagnostic tests commonly used in their practices for certain frequent diagnoses (for example, chest radiography to diagnose pneumonia), and turn to additional information sources when diagnosing an unusual condition (for example, granulomatosis with polyangiitis) or evaluating a less commonly used test (for example, accuracy of serum procalcitonin for pneumonia). Other sources include clinical guidelines, online evidence reviews, American College of Radiology appropriate use criteria (8), and published studies, as well as clinical laboratory and radiology physicians and staff who may have the most up-to-date information and know the performance of tests and testing approaches at their facility. Subspecialists may also have more recent information on tests in their field. However, because they see a different spectrum of patients than the general medical service, they may have different pretest probability estimates and testing recommendations. It can

often be enough to understand whether a negative test result has a small, moderate, or large effect on the probability that the patient does not have the condition and similarly whether a positive test result has a small, moderate, or large effect on the probability that the patient does have the condition. In addition, it is important to get a sense of any potential direct harms from the test as well as the potential for false-positive results, which lead to unnecessary follow-up interventions.

For this patient, the team identified several potential tests, including renal ultrasonography, autoimmune serologic testing, blood cultures, and renal biopsy, that should be considered given the potential diagnoses.

How Can I Decide Which of These Tests Are Most Useful for This Patient?

Deciding whether to order a test depends on pretest probability of the disease, accuracy and potential harms of the test, effectiveness and potential harms of the therapy, and severity of the condition (9) (Figure 2). In general, a test that is highly accurate can be useful even at low and high pretest probabilities because of the low rate of false-positive and false-negative results; a test that is less accurate may be useful mostly for intermediate probabilities. Of course, the usefulness of a test also depends on its risk. Minimal-risk tests are useful across a range of scenarios, whereas high-risk tests are reserved for scenarios where the value of the information is critically important for the clinical decision (for example, to move a patient above or below the threshold for receiving a potentially harmful treatment). The cost of the test can also influence its usefulness, with low-cost tests preferred over high-cost tests when characteristics and risks are equivalent. Finally, treatment options have an important effect on the use of a diagnostic test. In the extreme setting, if a completely effective and risk- and cost-free treatment was available, there would be little value to diagnostic testing, at least in terms of treatment decision making. In general, the riskier or less effective the treatment, the less likely empirical therapy will be started and the more useful a diagnostic test can be, even though the accuracy of the test is unchanged.

How Do I Choose Which Tests to Order First?

In medicine, we often consider the most serious, life-threatening diagnostic possibilities first when approaching a clinical scenario. Thus, when deciding on a diagnostic testing strategy, diagnoses that are most severe and potentially treatable should be considered

Table 3. Summary of Diagnostic Testing Decisions

Diagnosis	Test	Test Characteristics	Treatment	Treatment Characteristics	Decision
Urinary obstruction	Ultrasound	High accuracy, minimal risk	Foley catheter placement	High effectiveness, low risk	Pretest probability of obstruction was low (<1%), but given accurate, minimal-risk test and the invasive nature and potential discomfort of catheter placement, an ultrasound was obtained.
Sepsis	Blood culture	Low accuracy (until 48 h), minimal risk	Antibiotics	High effectiveness, low risk	Pretest probability of sepsis was low (<1%), but given high mortality; lack of a test with accurate, rapid turnaround; and low-risk, high-effectiveness therapy, empirical antibiotic therapy was started. When the sensitivity of blood cultures increased at 48 h, antibiotics were withdrawn.
Lupus nephritis	Antinuclear antibody test and anti–double-stranded DNA serologic tests	Moderate accuracy, minimal risk	Immunosuppression	Moderate effectiveness, high risk	The pretest probability of lupus nephritis was moderate (about 20%). The high risk of treatment and low risk of testing argued against empirical therapy and for serologic testing, even though the accuracy of serologic testing is only moderate.
Lupus nephritis	Kidney biopsy	High accuracy, moderate risk	Immunosuppression	Moderate effectiveness, high risk	The pretest probability of lupus nephritis was high (>90%) after the antinuclear antibody test and anti–double-stranded DNA results. Given the risks of the test and the treatment, the patient was presented with both options and chose to undergo kidney biopsy.

first. Of course, in any scenario, some diagnoses are sufficiently unlikely that it does not make sense to prioritize testing for them, even if they are life-threatening. For example, although aortic dissection and acute myelomonocytic leukemia can both cause renal failure and are life-threatening if untreated, they are sufficiently unlikely given this patient's presentation that testing for those conditions was not pursued. However, a serious and possible diagnosis for this patient was sepsis, which has a 20% mortality rate if not rapidly treated (10). Diagnostic tests for sepsis are complicated because rapid-turnaround tests, such as procalcitonin, have low sensitivity and specificity, and blood cultures only achieve reasonable sensitivity (90% to 95%) and specificity (98%) after 48 hours (11, 12). Thus, given the risk associated with for untreated sepsis and the effectiveness and relatively low risk of empirical antibiotic treatment, the patient was administered antibiotics despite a low pretest probability of sepsis (Figure 2). When culture results were negative at 48 hours, antibiotics were withdrawn.

The next step in the diagnostic strategy should focus on the most likely conditions. Because glomerulonephritis, including systemic lupus erythematosus (SLE), was considered the most likely diagnosis on the basis of the patient's history and laboratory results, the team conducted autoimmune serologic testing (given relatively high accuracy and low risk). The antinuclear antibody test result was positive (1:5120). As with many tests, antinuclear antibody test results can be categorized using thresholds, creating estimates of sensitivity and specificity. When the initial test has relatively low specificity, such as a low antinuclear antibody titer, the accuracy of a diagnostic testing strategy can be increased through cascade testing, in which patients who have an initial positive test result have subsequent, more specific testing, such as an anti–double-stranded DNA test. Often, cascade testing strategies are delivered through reflex testing algorithms managed by the clinical laboratory (13, 14). In this case, anti–double-stranded DNA testing was done. The result was positive at a titer of 1:160, which has a sensitivity of 57% and a specificity of 97% for SLE (positive likelihood ratio >16) according to a published review (15). Given the high test specificity and relatively high pretest probability (approximately 20%), the team estimated the posttest probability of SLE at 80% (for example, using the nomogram in Figure 1).

In addition to focusing on the most likely diagnoses, it is often helpful to consider testing for conditions that may be less likely but are treatable and have highly accurate, low-risk diagnostic tests. For example, in this case, although the probability of urinary obstruction was thought to be low, renal ultrasonography (the diagnostic test for urinary obstruction) is highly accurate, is largely without risk for clinically significant harm, and has relatively low cost. The treatment of urinary obstruction (placing a urinary catheter), although highly effective and relatively low risk, is invasive and uncomfortable. Thus, while waiting for the serologic test results, the team obtained a renal ultrasound that was negative for obstruction (Figure 2).

Should I Order a Test With Clinically Significant Risks or Start Empirical Treatment?

The team then needed to decide whether to do a renal biopsy or begin treatment of lupus nephritis, both of which have clinically significant risks and potential benefits (Figure 2). In this setting, when further testing could lead to harm and clinical judgment supports either testing or treatment, engaging the patient in the decision becomes particularly critical. The team discussed both options and their risks with the patient (communicating risk and engaging patients in shared decision making is discussed in another paper in this series [16]). The patient chose to have biopsy, primarily because of concerns about starting potentially risky therapy—given that an 80% probability of SLE left an approximately 20% probability of a different diagnosis—and the information that the biopsy could provide about prognosis. The biopsy showed diffuse proliferative crescentic and membranous lupus nephritis. Table 3 summarizes the diagnostic testing decisions involved in arriving at this diagnosis.

CONCLUSION

The appropriate use of diagnostic tests in clinical decision making is a key skill of an internist. Although selection and interpretation of diagnostic tests are second nature for many practicing clinicians, understanding the principles that underlie these skills is important for reducing diagnostic error and improving clinical decision making.

From Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts (K.A.A., J.P.M.).

Disclosures: Authors have disclosed no conflicts of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M19-1940.

Corresponding Author: Katrina A. Armstrong, MD, Massachusetts General Hospital, Gray-Bigelow 7, 55 Fruit Street, Boston, MA 02114; e-mail, karmstrong6@mgh.harvard.edu.

Current author addresses and author contributions are available at Annals.org.

Ann Intern Med. 2020;172:604-609. doi:10.7326/M19-1940

References

1. **Griner PF, Mayewski RJ, Mushlin AI, et al.** Selection and interpretation of diagnostic tests and procedures: principles and applications. Ann Intern Med. 1981;94:557-92. [PMID: 6452080]
2. **Jaeschke R, Guyatt GH, Sackett DL.** 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? The Evidence-Based Medicine Working Group. JAMA. 1994;271:703-7. [PMID: 8309035]
3. **Zhi M, Ding EL, Theisen-Toupal J, et al.** The landscape of inappropriate laboratory testing: a 15-year meta-analysis. PLoS One. 2013;8:e78962. [PMID: 24260139] doi:10.1371/journal.pone.0078962

4. **Centers for Disease Control and Prevention.** Rapid diagnostic testing for influenza: information for clinical laboratory directors. Accessed at www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm on 6 August 2019.
5. **Restrepo D, Armstrong KA, Metlay JP.** *Annals* clinical decision making: avoiding cognitive errors in clinical decision making. Ann Intern Med. 21 April 2020. [Epub ahead of print]. doi:10.7326/M19-3692
6. **Metlay JP, Armstrong KA.** *Annals* clinical decision making: incorporating perspective into clinical decisions. Ann Intern Med. 21 April 2020. [Epub ahead of print]. doi:10.7326/M19-3469
7. **Liaño F, Pascual J.** Epidemiology of acute renal failure: a prospective, multicenter, community-based study. Madrid Acute Renal Failure Study Group. Kidney Int. 1996;50:811-8. [PMID: 8872955]
8. **American College of Radiology.** ACR Appropriateness Criteria. Accessed at www.acr.org/Clinical-Resources/ACR-Appropriateness-Criteria on 6 August 2019.
9. **Pauker SG, Kassirer JP.** The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109-17. [PMID: 7366635]
10. **Rhee C, Dantes R, Epstein L, et al; CDC Prevention Epicenter Program.** Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. JAMA. 2017;318:1241-9. [PMID: 28903154] doi:10.1001/jama.2017.13836
11. **Andriolo BN, Andriolo RB, Salomão R, et al.** Effectiveness and safety of procalcitonin evaluation for reducing mortality in adults with sepsis, severe sepsis or septic shock. Cochrane Database Syst Rev. 2017;1:CD010959. [PMID: 28099689] doi:10.1002/14651858.CD010959.pub2
12. **Lamy B, Dargère S, Arendrup MC, et al.** How to optimize the use of blood cultures for the diagnosis of bloodstream infections? A state-of-the art. Front Microbiol. 2016;7:697. [PMID: 27242721] doi:10.3389/fmicb.2016.00697
13. **Srivastava R, Bartlett WA, Kennedy IM, et al.** Reflex and reflective testing: efficiency and effectiveness of adding on laboratory tests. Ann Clin Biochem. 2010;47:223-7. [PMID: 20392754] doi:10.1258/acb.2010.009282
14. **Craig WY, Ledue TB.** The antinuclear antibody assay: developing criteria for reflexive anti-dsDNA antibody testing in a laboratory setting. Clin Chem Lab Med. 2011;49:1205-11. [PMID: 21612543] doi:10.1515/CCLM.2011.613
15. **Kavanaugh AF, Solomon DH; American College of Rheumatology Ad Hoc Committee on Immunologic Testing Guidelines.** Guidelines for immunologic laboratory testing in the rheumatic diseases: anti-DNA antibody tests. Arthritis Rheum. 2002;47:546-55. [PMID: 12382306]
16. **Armstrong KA, Metlay JP.** *Annals* clinical decision making: communicating risk and engaging patients in shared decision making. Ann Intern Med. 21 April 2020. [Epub ahead of print]. doi:10.7326/M19-3495

Current Author Addresses: Dr. Armstrong: Massachusetts General Hospital, Gray-Bigelow 7, 55 Fruit Street, Boston, MA 02114.
Dr. Metlay: Division of General Internal Medicine, Massachusetts General Hospital, 100 Cambridge Street, 16th Floor, Boston, MA 02114.

Author Contributions: Conception and design: K.A. Armstrong, J.P. Metlay.
Analysis and interpretation of the data: J.P. Metlay.
Drafting of the article: K.A. Armstrong.
Critical revision of the article for important intellectual content: K.A. Armstrong, J.P. Metlay.
Final approval of the article: K.A. Armstrong, J.P. Metlay.
Administrative, technical, or logistic support: K.A. Armstrong, J.P. Metlay.