Conceptual Framework for Evaluating Laboratory Tests: Case-Finding in Ambulatory Patients

Marc D. Silverstein1 and Benoit J. Boland2

In response to concerns about inappropriate testing and healthcare costs, clinicians and laboratorians should work together to evaluate laboratory tests. Studies of routine laboratory tests ordered for screening and case-finding in the preoperative setting, at hospital admission, and in the ambulatory setting have not provided sufficient evidence of the benefits of case-finding when evaluated against commonly used criteria for screening procedures. We adapt a conceptual framework developed from radiology for evaluating imaging technologies to evaluate laboratory tests. The framework addresses diagnostic efficacy (whether the test correctly identifies abnormalities), diagnostic effectiveness (whether the test changes the physician's diagnoses), therapeutic efficacy (whether the test changes patient management), and therapeutic effectiveness (whether the test changes patients' outcomes). We propose that clinicians and laboratorians use this framework to evaluate the diagnostic efficacy, diagnostic effectiveness, therapeutic efficacy, and therapeutic effectiveness of laboratory tests in the ambulatory setting and in other settings.

Indexing Terms: laboratory management/screening

Laboratory tests are an important part of the process of care of ambulatory adult patients. Traditionally, clinical chemists have been responsible for ensuring that diagnostic laboratories perform tests accurately and report the results to the requesting physician as soon as possible. Clinicians have been responsible for ordering the appropriate tests and for managing the patient once the results are available. We are now moving beyond an era of cost containment to the "third revolution" in medical care, the era of accountability (1). The growing vertical and horizontal integration of physicians, ambulatory clinics, hospitals, and associated laboratories into healthcare systems provides a new opportunity to bring the expertise and perspective of the separate disciplines to the evaluation of laboratory tests ordered in the ambulatory setting.

Laboratory tests may be ordered in the ambulatory setting to evaluate presenting problems, to monitor known abnormalities, or to detect unsuspected disease (case-finding). As laboratory tests and other medical procedures face increasing scrutiny regarding their appropriate use and cost, attention has turned to the use of tests for case-finding as an area where medical care costs may be reduced. Accordingly, it is timely to reevaluate the effectiveness and costs of laboratory tests for case-finding in ambulatory care.

In this review we will present the perspectives of physicians, clinical epidemiologists, and health services researchers on the topic of laboratory tests ordered for case-finding in ambulatory patients. Four topics will be addressed: (a) presentation of uniform definitions and a conceptual framework for evaluating diagnostic tests; (b) the appropriate use of tests; (c) current evidence about the performance of routine laboratory tests in preoperative evaluation, at hospital admission, and in the ambulatory setting; and (d) a modification of the conceptual framework for evaluating tests applied to the evaluation of ambulatory test ordering.

Definitions

It is necessary to clarify terms used in evaluating laboratory tests. Confusion may result from the use of similar terms by laboratorians, clinicians, and health services researchers to describe different indications and characteristics of diagnostic tests. The definitions in Table 1 were developed by the Canadian Task Force on the Periodic Health Examination and are useful in designing a conceptual framework for evaluating laboratory tests (2). Screening is the application of tests and procedures to large unselected populations; people identified to have a high risk are then advised to consult their own physician for further evaluation. Case-finding is the application of a test or procedure by a physician or healthcare worker to a specific patient to detect an unsuspected condition that is unrelated to the problem for which the patient is being seen; the healthcare worker who ordered the test is responsible for further evaluation. The terms screening and case-finding are often used interchangeably, but are not identical concepts. Case-finding more accurately describes test ordering in the ambulatory setting. Screening and case-finding are both aimed at classifying people into two groups: one with a high probability of a fatal or disabling condition and the other with a low probability of a fatal or disabling condition.

The basic criteria for screening formalized by the World Health Organization (WHO) (3–5) also hold for case-finding. When the following criteria are met, a test can be judged to be efficacious for screening or case-finding:

1. Screening tests should have a significant effect on the patient's quality or quantity of life.
2. Acceptable treatment should be available.

3. The condition that screening detects should have an asymptomatic period.

4. Treatment in the asymptomatic period should be superior to treatment when the patient becomes symptomatic.

5. Acceptable tests should be available.

6. The incidence of the condition should be sufficient to justify the cost of the screening tests.

Efficacy and effectiveness are useful concepts to allow investigators and clinicians to understand the different levels of evidence needed to demonstrate when a test should be used in a research setting or adopted in a practice setting (Table 1). Efficacy describes the performance of tests under ideal conditions and in clinical trials. In contrast, effectiveness refers to the performance of a test in the usual practice setting. Efficacy and effectiveness are important concepts, but two additional concepts, efficiency and equity, are needed in the evaluation of tests. Inevitably patients, insurers, governmental organizations, and society determine the resources that are available for healthcare, and physicians and laboratorians must decide how to use the available resources for the most effective tests in the most efficient way possible. It may not always be possible for clinicians to use all the tests that are known to be effective for a specific indication for all patients. Efficiency is concerned with the optimal use of resources. In the US today we have >35 million uninsured persons, and one of the key issues in the current debate on healthcare reform is how to assure access to care. Equity can be defined as equal access to health services.

A Framework for Evaluating Diagnostic Tests

A conceptual framework for evaluating the efficacy of diagnostic tests has been proposed by Dennis Fryback, an investigator with training in industrial engineering and decision sciences, and Jack Thornbury, a radiologist at the University of Wisconsin (6). This framework originally grew from work on diagnostic imaging but is easily adapted to laboratory tests. The Fryback and Thornbury framework for diagnostic efficacy addresses six hierarchical levels: technical efficacy, diagnostic accuracy efficacy, diagnostic thinking efficacy, management efficacy, patient outcome efficacy, and societal efficacy. The first level describes the technical aspects of the imaging technology or the laboratory tests: Basically, does the imaging technology provide a good picture? Does the laboratory test measure the substance it purports to measure? The second level, diagnostic accuracy efficacy, describes yield, sensitivity, specificity, and positive and negative predictive values of a laboratory test in the clinical setting. Next, diagnostic thinking efficacy addresses the issue of whether the diagnostic test changes the thinking of the clinician. Fourth, therapeutic efficacy, addresses whether a test result changes the clinical management of the patient, whereas patient outcome efficacy, the fifth level, describes the outcomes that patients actually experience. Finally, societal efficacy addresses issues such as the health of the population and the costs of tests. The Fryback and Thornbury framework can be applied to any type of test ordered for any indication. After review of relevant issues and available data in the evaluation of routine laboratory tests, we will propose a modification of the framework to also take into account the intended management plan, not just the management initiated.

Current Issues in Evaluating Laboratory Tests

In the evaluation of routine laboratory tests in ambulatory patients, clinicians and laboratorians must address whether tests are ordered appropriately as well as the issue of the costs of laboratory tests. Physicians generally report that laboratory tests are helpful and result in a change in diagnosis, therapy, and prognosis, or in their understanding of the patient’s disease (7). Informally, however, physicians may mention inappropriate reasons for ordering tests (8). Periodically the medical literature contains reports of inappropriate test ordering (9, 10), reflecting the concern that in clinical practice many laboratory tests are obtained for reasons that would not withstand careful critical review.

When medical care costs are evaluated, the analysis usually focuses on the high-cost decisions: e.g., the decision to admit a patient to the hospital, the decision to admit to intensive care, the decision to admit to a nursing home, the decision to operate, and the decision to use life-sustaining therapy. These inpatient decisions are all characterized by important clinical events in the care of individual patients. These decisions are often referred to as “big-ticket” items. Medical care costs, however, result not only from the use of high-cost procedures, but also from the high volume use of low-cost tests. These “little-ticket” items often escape attention because they are less visible in the course of care of an individual patient. In addition, because of their high volume of use, they are accepted as being customary and
routine. Nevertheless, in the aggregate, a large volume of low-cost laboratory testing results in high medical care costs (11). For example, the Health Care Financing Administration reported that 16278672 blood cell counts at charges totaling $150,480,379 and 7751248 automated multichannel tests totaling $135,974,256 were obtained for Medicare patients in 1987 (12).

The focus on the cost of laboratory tests has resulted in attention to the appropriate use of laboratory tests, possibly motivated by concern that some portion of the increase in healthcare costs results from inappropriate and hence avoidable use of laboratory tests. Two major issues that must be considered in test ordering are analogous to the false-positive and the false-negative test results, or the alpha or type I error and the beta or type II error in statistics. The first issue is the ordering of tests that are not indicated. Several important and influential studies of frequently ordered medical and surgical procedures—e.g., upper gastrointestinal endoscopy (13), carotid endarterectomy (14), and coronary angiography (15)—suggest that, when judged against explicit predefined criteria, large numbers of these medical procedures are not indicated. Overutilization would contribute to increased costs and unnecessarily subject patients to costly and risky procedures. Second is the failure to order an indicated test, manifested by underutilization. Although underutilization may appear to be less important than overutilization, underutilization may result in failure to detect disease and ultimately result in poor outcomes with associated increased costs. Underutilization has been an understudied area.

Current Data on Routine Laboratory Tests

The value of performing routine laboratory tests at admission to hospital and before surgery has been studied extensively. The value of routine laboratory tests in the ambulatory setting, however, and in different populations needs further study. We will summarize the main conclusions of selected studies of routine hospital admission tests and preoperative tests that are relevant to the ambulatory setting.

Routine preoperative tests. A study of 1010 patients undergoing elective cholecystectomy, who were free of active disease other than cholelithiasis, evaluated the following preoperative laboratory tests: complete blood count, sodium, potassium, creatinine, urea, and urinalysis in almost all patients, and chest radiography, electrocardiogram, serum glucose, prothrombin time, and partial thromboplastin time. Overall, results for 225 (5%) of the 5003 routine tests were abnormal, but only 104 (2%) were considered potentially important. Action was taken in 17 patients, and only 4 patients (0.4%) were judged to have benefited from the tests (16).

Another study evaluated 2000 elective surgery patients who received a six-factor automated biochemical panel—glucose, complete blood count, differential cell count, platelet count, prothrombin time, and partial thromboplastin time—as routine preoperative procedure. Subsets of patients who received these preoperative tests were selected for further analyses. Of the 2785 selected tests, 957 (34%) had been ordered for a specific indication and 1828 (66%) were routinely ordered. Results for 86 (9%) of the 957 tests ordered for a specific indication were abnormal, but only 10 (0.5%) from the 1826 routine tests were abnormal, of which only 4 (0.2%) were judged to indicate a surgically significant result (17).

A Mayo Clinic study evaluated routine preoperative test results from 3782 healthy adults, ASA Class 1, who were undergoing elective surgical procedures (18). Overall, 160 patients (4.2%) had at least one abnormal result. Eighteen percent of abnormalities could have been predicted from the patient's medical history and physical examination. Twenty-nine percent of patients were further evaluated before surgery, and in only 10 patients (0.26%) was management changed. No surgical procedure was delayed and, most importantly, no adverse effects were related to preoperative laboratory abnormalities.

Overall, the studies of preoperative testing in healthy patients suggest the following: There is high use and low yield of routine laboratory tests; abnormalities are often not noted in the medical record and not subsequently evaluated; the management of patients is often not changed in response to abnormal results; and abnormal laboratory tests are poor predictors of adverse postoperative outcomes. Several years ago, a Mayo Clinic committee recommended new guidelines for the preoperative evaluation of healthy patients, which were implemented throughout the Mayo Clinic multispecialty group practice. The old and new guidelines for preoperative tests in the otherwise healthy adult are shown in Table 2. In contrast to the old guidelines, the new guidelines stratify patients into risk groups based on (a) the age-specific prevalence of underlying conditions that may adversely affect surgical outcomes, (b) the planned preoperative care, and (c) patient-specific

<table>
<thead>
<tr>
<th>Test</th>
<th>Old guidelines</th>
<th>New guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Patients</td>
<td>Age &lt;40</td>
<td></td>
</tr>
<tr>
<td>Creatinine</td>
<td>No tests</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>Age 40–60</td>
<td></td>
</tr>
<tr>
<td>Complete blood count</td>
<td>Creatinine</td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>Glucose</td>
<td></td>
</tr>
<tr>
<td>Potassium</td>
<td>Electrocardiogram</td>
<td></td>
</tr>
<tr>
<td>Chloride</td>
<td>Age &gt;60</td>
<td></td>
</tr>
<tr>
<td>Bicarbonate</td>
<td>Creatinine</td>
<td></td>
</tr>
<tr>
<td>Urea</td>
<td>Glucose</td>
<td></td>
</tr>
<tr>
<td>Aspartate aminotransferase</td>
<td>Complete blood count</td>
<td></td>
</tr>
<tr>
<td>Chest x-ray</td>
<td>Electrocardiogram</td>
<td></td>
</tr>
<tr>
<td>Age ≥40</td>
<td>Chest radiograph</td>
<td></td>
</tr>
<tr>
<td>Above tests, plus electrocardiogram</td>
<td></td>
<td>Other groups</td>
</tr>
<tr>
<td>African Americans</td>
<td>Potassium for patients receiving diuretics or a bowel preparation</td>
<td></td>
</tr>
<tr>
<td>Above tests, plus sickling test</td>
<td>Chest x-ray for patients with cardiac or pulmonary disease</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Mayo Clinic preoperative test-ordering guidelines.
risk factors and diagnoses. These guidelines have decreased routine preoperative testing in the Mayo practice, especially in the younger patients for whom no routine tests are required.

**Routine admission tests.** Studies of routine laboratory tests performed at hospital admission have also not demonstrated substantial improvement in outcomes associated with routine testing. Older, uncontrolled trials report that the yield of new diagnoses resulting from routine biochemical panels was 4–10% (19, 20). Subsequent controlled trials reported that routine admission tests have no significant effect on length of stay or hospital outcomes (21–23). One study reported a 5% increase in costs (22).

The potential benefit of routine biochemical and hematological testing at admission was assessed in a prospective study of 1000 patients admitted to a 575-bed hospital in Vancouver (24). Eighty-three new diagnoses were made in 77 patients, but critical review showed that no tests were unequivocally beneficial to the patient. The authors report that 30 of the new diagnoses might have benefited patients had the abnormalities been followed up diligently, 39 findings were of no lasting significance, and 14 diagnoses of mild asymptomatic diabetes were made.

Another comprehensive evaluation of routine admission testing evaluated urinalysis, hematocrit, leukocyte count, platelet count, a six-item biochemistry panel, prothrombin time, partial thromboplastin time, chest x-ray, and electrocardiogram for 301 patients at the University of California Irvine Medical Center (25). Results for 12% of the 1658 routine tests were abnormal, 5% led to additional testing, but only 0.5% led to a change in treatment. The authors concluded that there is little justification for ordering routine tests solely because of hospital admission.

Thus, even in the setting of hospitalized patients, where the prevalence of disease may be higher than in the ambulatory setting, routine testing has low yield, identifies abnormalities that are often not further evaluated, and has no impact on the process of care. Moreover, routine admission testing does not reduce total testing or length of stay and may indeed increase costs.

**Laboratory tests in the ambulatory setting.** Laboratory tests are frequently ordered in the comprehensive evaluation of ambulatory patients. Indeed, many patients seek an ambulatory periodic health examination ("a yearly check-up") for "screening" with routine laboratory tests. What unsuspected conditions should we expect to identify through the use of our routine tests for case-finding in the ambulatory setting? A complete blood count might detect anemia. A biochemistry panel might identify diabetes mellitus, hyperlipidemia, hyperuricemia, hyperparathyroidism, or evidence of liver, bone, or kidney disease. Urinalysis may identify urinary tract infection, diabetes, kidney disease, or genitourinary malignancy. A thyroid panel might identify hypothyroidism or, less frequently, Graves disease or Hashimoto thyroiditis.

The prevalence of these conditions in asymptomatic patients is very low, except for hyperlipidemia and diabetes mellitus. The yield of screening laboratory tests cannot be higher than the prevalence of the target conditions in the patients being evaluated. More information on the prevalence of these conditions among patients who have an initial case-finding test is needed to evaluate whether these conditions should be the target of case-finding. Further, if case-finding is indicated, information on the incidence of the conditions is needed to determine whether the test should be repeated periodically, and if so, to indicate the optimal interval for case-finding.

Only a few studies have addressed the value of screening laboratory tests in the ambulatory setting. Perhaps the best known is the Kaiser Permanente study of multiphasic testing (26). The Kaiser Permanente study was a prospective randomized trial of periodic health check-ups in which 5156 Kaiser Foundation health plan members, ages 35–54, were advised to have annual multiphasic screening, including automated multichannel laboratory tests. The long-term follow-up of this group at 16 years, compared with 5557 health plan members who otherwise received usual care but no annual screening, reported no differences in disability or total mortality, and no benefits attributed to any findings of the routine multichannel laboratory tests.

Similar disappointing results of routine laboratory tests were noted in another trial of multiphasic screening. A clinical controlled trial of multiphasic screening in 574 families in Salt Lake City in the 1970s, designed to identify differences in utilization of health services, morbidity, and health status, found that routine tests did not change morbidity or health status. The only significant difference was increased hospitalization among screened subjects (27).

Recently, the value of laboratory screening tests in the ambulatory setting has been studied in a university clinic in Basel, Switzerland. In a prospective study of an unselected cohort of 493 newly evaluated ambulatory patients, a 23-test biochemistry panel, including a lipid profile, was routinely obtained for each subject (28). The mean age of these patients was 41 years; 43% were female, and only 8% were older than 60. Results for 1162 (11%) of 10 190 routine tests were abnormal; however, the overall yield was only 32 new diagnoses, of which 25 were treated: hyperlipidemia (22 cases), alcoholic liver disease (2), and hypocalcemia (1). Thus only 25 of 493 patients (4%) benefited from case-finding, as judged by the initiation of some form of treatment, mainly through screening for hyperlipidemia.

The value of four tests from the complete blood count (hemoglobin, mean erythrocyte volume, leukocyte count, and platelet count) for case-finding was also analyzed in an unselected cohort of 595 adults from the same clinic (29). Anomalies were detected in 6% of routine tests. The overall yield was five diagnoses without treatment and only three new diagnoses with treatment, all microcytic anemia from benign conditions.

In summary, these two recent studies from Basel found that only 0.2% of the routine tests resulted in a
change in patient management. In this setting (mostly a young male population), the value of the biochemistry panel was mainly related to the yield of the lipid tests; the complete blood count had no value.

A retrospective descriptive review of 100 patients who had a comprehensive general medical examination at the Mayo Clinic in 1991 was recently undertaken to help plan a prospective study of routine ambulatory test ordering (Boland et al., ms. submitted for publication). Fifty patients from the local community and 50 patients from the referral practice were studied. The biochemistry profile, lipids, complete blood count, thyroid hormone concentrations, and urinalysis were evaluated as routine case-finding tests. The proportion of patients for whom these routine tests were used for case-finding ranged from 83% (thyroid tests) to 98% (complete blood counts). The proportion of routine tests giving anomalous results ranged from 8% (thyroid tests) to 19% (biochemistry panel). Altogether, a new diagnosis associated with treatment was found in 14% of the patients. The study identified six patients with hyperlipidemia, two patients with diabetes mellitus or glucose intolerance, two patients with drug-associated electrolyte abnormalities, one patient with hypothyroidism, and one patient with microcytic anemia. Serum lipids had the highest therapeutic yield, glucose was intermediate, and the lowest-yield tests were serum electrolytes, enzymes, creatinine, protein, complete blood count, and "sensitive" thyrotropin (TSH). These findings are consistent with a thoughtful review of the value of routine biochemical panels by Cebul and Beck (30).

The yield of case-finding tests in terms of new diagnoses and new treatments is not an optimal measure of the value of tests, but it does provide an obvious starting point for such analyses. Published studies do not provide sufficient data to justify the routine use of these ambulatory tests as screening tests or case-finding according to the WHO criteria. Indeed, the American College of Physicians (31), the Canadian Task Force on the Periodic Health Examination (2, 32), and the US Preventive Services Task Force (33) agree that hemoglobin, glucose, and thyroid testing are not routinely indicated for case-finding (Table 3). None of the three expert panels recommends a biochemistry profile as a screening or case-finding test in the ambulatory setting. All three panels recommend that cholesterol be ordered for case-finding at 5-year intervals (31–35). These expert panels used clinical judgment of panel members and the available medical literature in making their recommendations. There is a need for further study of case-finding in the ambulatory setting.

Revised Framework for Evaluating Ambulatory Laboratory Tests

Studies that have evaluated laboratory tests have not used a systematic framework to address efficacy and effectiveness. The framework proposed by Fryback and Thornbury (6) is a general framework that can be applied to imaging tests, laboratory tests, or other diagnostic procedures, and to tests ordered for evaluation of symptomatic patients or for case-finding. Studies of routine laboratory tests for case-finding in the ambulatory setting should address efficacy, effectiveness, and efficiency and assess whether patient access to these tests is equitable. We recommend that clinicians and laboratorians work together and adopt a common conceptual framework to evaluate routine laboratory tests in the ambulatory setting. The framework proposed by Fryback and Thornbury can be adapted easily for this purpose.

The successive six levels of efficacy proposed by Fryback and Thornbury can be sequentially evaluated when a new laboratory test is developed—with technical efficacy evaluation first, and societal efficacy last. In evaluating established laboratory tests that have been available for decades, we suggest that the areas of patient outcome and societal efficacy be the main targets for evaluation.

The Fryback and Thornbury framework uses only the term efficacy. We believe it would be helpful to clarify the difference between efficacy and effectiveness in the framework. Further, the framework should take into account both diagnostic thinking and intended management plan in addition to the diagnoses made and the management initiated. Therefore, we propose a framework with four levels—diagnostic efficacy, diagnostic effectiveness, therapeutic efficacy, and therapeutic effectiveness—with two components each (Table 4).

In this modified framework, the diagnostic efficacy of laboratory tests addresses whether tests detect abnormalities. This includes both technical efficacy in providing reproducible and valid results, and clinical efficacy as described by positive and negative predictive value. Both components are included because the performance of tests depends not only on the technical accuracy of the laboratory, but also on the spectrum of disease in the patients who are tested. Diagnostic effectiveness of laboratory tests addresses whether the tests result in new diagnoses. This includes the actual impact of the tests in changing the physician’s thinking about the diagnosis (impact on the differential diagnosis), as well as the actual change in the diagnoses recorded in the medical record, and most importantly, the actual diagnoses communicated to the patient.

Table 3. Expert panel recommendations (rec.) for routine laboratory tests in ambulatory care.

<table>
<thead>
<tr>
<th>Test</th>
<th>American College of Physicians</th>
<th>Canadian Task Force</th>
<th>US Preventive Services Task Force</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid tests*</td>
<td>Not rec.</td>
<td>Not rec.</td>
<td>Not rec.</td>
</tr>
<tr>
<td>Biochemistry group</td>
<td>No rec.</td>
<td>No rec.</td>
<td>No rec.</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Ages 18+, every 5 years</td>
<td>Men, ages 30–59, every 5 years</td>
<td>Ages 18+, every 5 years</td>
</tr>
</tbody>
</table>

* Includes sensitive TSH or thyroxine.

Recommendations abstracted from Hayward et al. (35).

CLINICAL CHEMISTRY, Vol. 40, No. 8, 1994 1625
Efficacy and effectiveness usually describe therapeutic interventions. When these concepts are used to evaluate tests used for diagnosis, case-finding, or screening, the evaluation includes whether the tests lead to decisions to initiate a therapeutic intervention and whether the interventions change outcomes. **Therapeutic efficacy** of laboratory tests addresses their actual impact in changing the physician's intended management plan, as well as the actual therapies initiated, not initiated, or discontinued. These therapies may include medical or pharmacological interventions, surgical interventions or procedures, behavioral interventions, counseling, and patient education. **Therapeutic effectiveness** of laboratory tests includes both patient and societal outcomes. Patient outcomes include death, disability, function, satisfaction, and cost of care. Societal outcomes include analyees of the impact of diagnostic testing on the health and function of members of the community and on societal costs, i.e., both the economic costs of providing services and the noneconomic costs, including lost productivity.

The framework in Table 4 is a revision of the important innovative work by Fryback and Thornbury. Our framework uses both efficacy and effectiveness to describe different levels, simplifies the number of levels, and explicitly addresses the impact of tests in changing the intended and actual diagnoses and in changing the intended and actual management. Further, it provides a ready mapping to Donabedian's conceptual framework for quality in healthcare, which includes structure (diagnostic efficacy), process (diagnostic effectiveness and therapeutic efficacy), and outcomes (therapeutic effectiveness) (36).

In conclusion, medicine is currently experiencing a paradigm shift away from the principle of "do no harm," under which we have the technology and resources to order tests provided we are not harming our patients, to a new paradigm exhorting us to "do what works," contain costs, and measure outcomes. As physicians and laboranitians we must evaluate our current practices before external forces oblige us to change our practice and then evaluate the impact of these changes. We need to change our clinical and laboratory practices on the basis of evidence for the effectiveness of laboratory tests in the clinical setting. To respond to this era of change, we must study the performance of laboratory tests in the ambulatory setting, continually monitor patient outcomes, and adapt our practice to the needs of society.

**Table 4. Framework for evaluating tests.**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Measure of test impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic efficacy: test results</td>
<td>Sensitivity &amp; specificity</td>
</tr>
<tr>
<td>Technical efficacy</td>
<td>Positive &amp; negative predictive value</td>
</tr>
<tr>
<td>Clinical efficacy</td>
<td>Yield</td>
</tr>
<tr>
<td>Diagnostic effectiveness: new diagnoses</td>
<td>Prevalence of disease (pretest probability)</td>
</tr>
<tr>
<td>Diagnostic thinking</td>
<td>Positive predictive value (posttest probability)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Diagnostic yield</td>
</tr>
<tr>
<td>Therapeutic efficacy: new therapies</td>
<td>Physician's intended management is usually implicit and not often studied</td>
</tr>
<tr>
<td>Therapeutic plan</td>
<td>Proportion of tests resulting in a change in management</td>
</tr>
<tr>
<td>Actual therapy</td>
<td>Therapeutic yield</td>
</tr>
<tr>
<td>Therapeutic effectiveness: outcomes</td>
<td>Morbidity, mortality, functional outcomes, costs</td>
</tr>
<tr>
<td>Patient outcomes</td>
<td>Cost per diagnosis, treatment, or outcome</td>
</tr>
<tr>
<td>Societal outcomes</td>
<td>Cost per year of life</td>
</tr>
<tr>
<td></td>
<td>Cost per quality-adjusted life (QALY)</td>
</tr>
</tbody>
</table>

References

13. Kahn KL, Kosecoff J, Chassin MR, Solomon DH, Brook RH.