The Value of Laboratory Screening and Diagnostic Tests for Prevention and Health Care Improvement

Prepared for:
American Clinical Laboratory Association and Advanced Medical Technology Association (AdvaMed)

Prepared by:
The Lewin Group, Inc.

September 2009
The Value of Laboratory Screening and Diagnostic Tests for Prevention and Health Care Improvement

Table of Contents

Overview 1

I. Introduction 2
   A. Increasing Value is a National Priority .................................................................2
   B. Value of Laboratory Medicine to Clinical Care .....................................................3
   C. Differentiating Screening and Diagnostic Tests .....................................................6
      Relationship of Screening and Diagnostic Tests to Prevention .........................6

II. Validity and Utility of Screening and Diagnostic Tests 7
   A. Analytic Validity ......................................................................................................8
   B. Clinical Validity ......................................................................................................8
   C. Clinical Utility ........................................................................................................11
      1. Comparative Effectiveness Research .................................................................12
      2. Challenges of Demonstrating Causal Effect of Testing on Outcomes ............14
   D. Economic Outcomes and Impact .......................................................................15

III. Analytic Frameworks to Appraise Evidence of Clinical Value 18
   A. Organizations that Conduct Evidence Appraisals ...............................................19
   B. Direct and Indirect Evidence of Test Impact on Outcomes ................................21
   C. Evidence Hierarchies Used for Screening and Diagnostic Testing ..................21
   D. Challenge of Relying on RCTs ............................................................................24
   E. Challenges Associated with Implementation of Screening in Clinical Practice ...25
      1. High Value Screening Tests Underused by Clinicians ....................................25
      2. Insufficient Patient Knowledge of High Value Tests ........................................26
      3. Lag Times and Varying Evidence Expectations ................................................27

IV. Effect of Value Assessments on Reimbursement Policy 28
   A. Coverage Decisions ...............................................................................................29
   B. Payment Systems ..................................................................................................30

V. Key Findings and Conclusions 31

References 35

Appendix: Case Studies A-1
   Case Study I: MRSA .................................................................................................A-3
   Case Study II: Diabetes ............................................................................................A-17
   Case Study III: KRAS ..............................................................................................A-33
   Case Study IV: HPV .................................................................................................A-47
Overview

The contributions of clinical laboratory screening and diagnostic tests\(^a\) to health care quality and outcomes are substantial. These contributions were described in an earlier report from The Lewin Group, *The Value of Diagnostics Innovation, Adoption, and Diffusion in Health Care* (2005). This report updates key elements of that study, providing a current overview of the important role of laboratory screening and diagnostic tests in our health care system, today’s means of assessing value, and four case studies documenting value of specific tests to patient care.

Our overarching finding is that: *Innovation, demonstrated clinical benefit, and appropriate use of laboratory screening and diagnostic tests are essential for achieving the goals of health system reform. Clinical laboratory testing is integral to evidence-based improvements in health care quality, patient outcomes, efficiency, and accountability.* This finding is substantiated, in part, by the four case studies accompanying this report:

- Rapid methicillin-resistant *Staphylococcus aureus* (MRSA) testing for identifying health care-acquired infections
- Hemoglobin A1c (HbA1c) testing for screening and diagnosis of prediabetes and diabetes
- *KRAS*\(^b\) gene mutation testing for targeted treatment of colorectal cancer
- Human papillomavirus DNA (HPV) testing to screen and diagnose cervical cancer

Though each of these case studies represents a unique configuration of health problems, at-risk patient populations, and testing technology, some common capabilities of laboratory testing emerge. Examples of these common elements include ongoing innovation, evidence of clinical benefit, early detection and treatment to control disease, better targeting and accuracy that enable more effective management and decision-making, more efficient care, and cost-saving opportunities. These case studies underscore the role of laboratory medicine in augmenting the evidence base for health care, decision-support for clinicians and patients, prevention and wellness, better patient outcomes, and better value.

Many of the benefits of laboratory testing are not being realized in the current system. The main body of the report reviews the current processes for assessing the value of laboratory screening and diagnostic tests and key policy issues that limit the realization of optimal value. Among the hurdles to appropriate use of these services are: insufficient provider awareness regarding when to use tests, challenges in generating evidence regarding the clinical utility of tests for particular patient subgroups and indications, inconsistencies among clinical practice guidelines regarding appropriate use of tests, inconsistent or inadequate coverage and payment policies, and need to provide additional evidence of the favorable economic impact of laboratory testing.

---

\(^a\) Depending on the patients’ needs and the level of technical and resource capabilities required to perform the test, clinical laboratory screening and diagnostic tests may be conducted in laboratories (including independent reference laboratories, hospital and academic medical center laboratories, physician office laboratories, nursing home laboratories, and others) or at the point-of-care.

\(^b\) *v-Ki-ras2 Kirsten rat sarcoma viral oncogene*
I. Introduction

Although the U.S. leads the world in per capita and total health care spending and is the world’s greatest source of health care innovation, it ranks poorly among industrialized nations in two gross indicators of health care quality: infant mortality and life expectancy.\(^1\) As documented in recent landmark studies, only 55% of adults and 47% of children in 12 metropolitan areas across the U.S. received recommended care, contributing to the significant shortfalls in overall care quality, patient safety, and health outcomes.\(^2,3\) There are similar deficits in adherence to recommendations for screening and follow-up care—only 62% of recommended laboratory and radiology tests were provided for preventive, acute, and chronic care. Aside from their implications for patient health, these shortfalls contribute to the sizable unnecessary spending in health care in the U.S., amounting to as much as $0.30 or more of every health care dollar.\(^4\) In the current health care environment, the needs and opportunities to improve health care quality and efficiency are great and immediate.

A. Increasing Value is a National Priority

According to the Congressional Budget Office (CBO), continued escalation of health care costs is the main long-term threat to the federal budget and the nation’s overall financial well-being.\(^5\) Increasing value in the health system and reducing unnecessary spending has become a national priority. Since publication of the Institute of Medicine (IOM) report, *Crossing the Quality Chasm: A New Health Care System for the 21st Century* (2000), stakeholders have been directing reform efforts toward strategies that increase the value of our health care services and produce better patient outcomes per dollar spent.\(^6\) The intent of this approach is to align how value is created for patients across services and time, and to target medical conditions over the cycle of care (episodes of care) and differing levels of care. Health care system stakeholders (including payers, providers, patients, accreditation and quality improvement organizations, health services researchers, policymakers, industry) agree that increasing value to patients can be achieved by improving quality through the types of goals shown in Box 1.\(^7\) Laboratory screening and diagnostic tests are essential for pursuing many of these goals.

<table>
<thead>
<tr>
<th>Health System Goals for Increasing Quality and Value</th>
<th>Box 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevention</td>
<td>• Less invasive treatment methods</td>
</tr>
<tr>
<td>Early detection</td>
<td>• Faster recovery</td>
</tr>
<tr>
<td>Right diagnosis</td>
<td>• More complete recovery</td>
</tr>
<tr>
<td>Early treatment</td>
<td>• Less disability</td>
</tr>
<tr>
<td>Right treatment to the right patients</td>
<td>• Fewer relapses or acute episodes</td>
</tr>
<tr>
<td>Treatment earlier in causal chain of disease</td>
<td>• Slower disease progression</td>
</tr>
<tr>
<td>Fewer delays in the care delivery process</td>
<td>• Less need for long-term care(^6)</td>
</tr>
</tbody>
</table>
B. Value of Laboratory Medicine to Clinical Care

The contributions of clinical laboratory screening and diagnostic tests to the health system goals (listed in Box 1) are substantial. As an essential component of high quality of care, laboratory tests are used for much more than diagnosis of disease in symptomatic individuals. Laboratory testing is an integral part of many medical decisions, providing clinicians with often pivotal information necessary for prevention, diagnosis, treatment, and management of disease. Despite the extensive role of laboratory medicine in informing medical decision-making, in 2007, spending on Part B laboratory services was $6.8 billion or just 1.6% of total Medicare expenditures and 2.3% of national health care spending. Significant contributions of laboratory medicine remain untapped.

The value of laboratory medicine is realized through its many roles in patient care. These include screening of asymptomatic individuals to identify risk for developing disease, detecting disease at the earliest stages before symptoms occur, selecting safe and effective treatments, planning disease management strategies, estimating treatment response throughout the course of care, identifying threats to patient safety and public health, such as hospital-acquired infections (HAI), protecting the blood supply and transplant recipients from harmful pathogens, and drugs of abuse testing to support clinical care and assure public safety. These aspects of value can be expressed along a continuum of care such as is listed in Box 2.

Laboratory medicine also is important to clinical guidelines. As described in our 2005 report, a search of clinical practice guidelines across 23 main condition/disease categories found that 37% focused on or involved laboratory tests. Increasingly, the objective, scientific data produced by clinical laboratory tests is used to measure provider performance (individual and organizational) as well as to implement value-based purchasing that aims to optimize use of health care resources and decrease short-, medium-, and long-term costs of care. Additionally, the recent major advances in science and technology, such as those associated with molecular-level and genetic testing, are leading to changes in clinical practice. Genetic tests are now available for more than 1,700 diseases, up from about 1,250 in 2005. New testing techniques tend to be more sensitive and specific, allowing clinicians to detect, diagnose, and manage disease more effectively. For example, technologies that rely on DNA, RNA, and protein composition allow evaluation of disease states at the molecular level, supporting earlier detection and a more personalized approach to patient care.

---

Box 2
Value of Laboratory Tests in Clinical Decision-making
- Screen for disease
- Screen to determine risk for developing disease
- ‘Rule in’ of a diagnosis
- ‘Rule out’ of a diagnosis
- Start an intervention
- Adjust an intervention
- Stop an intervention
- Assess efficacy of an intervention
- Assess compliance with an intervention
- Assess prognosis

---

c Total Medicare expenditures were $433 billion in 2007.
d Total laboratory industry revenues were estimated at $51.7 billion in 2007; this is 2.3% of the $2.24 trillion in 2007 national health care expenditures estimated by the Centers for Medicare and Medicaid Services.
Other innovative rapid testing techniques, such as those for microbiology, make testing for HAIs more cost-effective and less time-consuming, supporting real-time decision-making and operational efficiency. Innovations in miniaturization have expanded the point-of-care testing (POCT) menu, enabling laboratory testing at the hospital bedside, physician’s office, other clinical settings, and, in some cases, patient self-testing at home. These new technologies are demonstrating their value to improved patient outcomes and quality of life, fewer side effects of treatment, and decreased costs of care.

Examples of specific laboratory screening and diagnostic tests that can contribute to health system value are presented in Table 1.

Table 1. Examples of Screening/Diagnostic Tests and Services that Support the Health System Goals for Improving Value and Quality

<table>
<thead>
<tr>
<th>Health System Goal*</th>
<th>Examples of Screening and Diagnostics Tests</th>
</tr>
</thead>
</table>
| Prevention of disease | • HbA1c testing to screen for prediabetes to identify at-risk patients and prompt interventions that can prevent onset of type 2 diabetes  
• Factor V Leiden genetic variation to identify increased risk of blood clots.  
• Malignant hyperthermia genetic test to identify individuals who will react to general anesthesia (see GeneTests for reference)  
• Familial adenomatous polyposis genetic test to identify individuals who will develop colon cancer and allow prophylactic colectomy or other surgical procedure |
| Right diagnosis | • Cardiac enzyme marker tests, e.g., for troponin, myoglobin, creatinine phosphokinase (CPK), which are released after a heart attack and identify heart damage  
• Fragile X syndrome test for a form of inherited mental retardation and developmental delay, to determine appropriate management and risk of familial recurrence. |
| Early detection and treatment | • BRCA1 and BRCA2 test to identify increased risk of breast cancer and ovarian cancer in the absence of tumor  
• Prenatal and newborn screening for inherited disorders, to enable initiation of treatment and to reduce adverse effects  
• (Intra-amniotic infection/inflammation) IAI test, a non-invasive cervical-vaginal fluid (CVF) proteomic biomarker test, to diagnose intra-amniotic infection in preterm labor, a condition that is 80-90% asymptomatic |
| Right treatment to the right patients | • HER-2/neu (human epidermal growth factor receptor 2, also known as ErbB-2, ERBB2) protein testing for patients with breast cancer identifies those who will benefit from targeted treatment with trastuzumab (Herceptin)  
• KRAS gene mutation testing for patients with metastatic colorectal cancer distinguishes between patients who are most unlikely to benefit from, and those who are more likely to benefit from, the drug cetuximab  
• BCR/ABL oncogene testing for patients with chronic myelogenous leukemia who will benefit from treatment with imatinib (Gleevec) |
| Fewer mistakes and repeats in treatment | • HIV viral load test to determine disease progression and whether the drug is working  
• Emphysema gene test to identify likelihood of liver disease in emphysema patients without biopsy and allows early intervention |
| Fewer delays in the care delivery process | • Rapid molecular MRSA testing for meCA gene markers to identify, within two hours, patients with antibiotic-resistant S. aureus infections to guide drug selection and timely implementation of control measures  
• Point-of-care tests  
• Developmental delay/mental retardation tests, such as chromosome analysis, array comparative genomic hybridization (aCGH), biochemical genetic tests to help determine cause and appropriate therapies |
<table>
<thead>
<tr>
<th>Health System Goal*</th>
<th>Examples of Screening and Diagnostics Tests</th>
</tr>
</thead>
</table>
| **Less invasive treatment methods** | • Gene expression profiling of RNA isolated from peripheral blood mononuclear cells for noninvasive identification of heart transplant recipients with low probability of organ rejection  
• Hereditary hemochromatosis gene test to identify individuals who have the disease and decrease need for liver biopsy  
• Cytochrome 2C19 gene test for clopidogrel to identify heart attack or stroke patients who may not respond to therapy so that other therapies can be prescribed |
| **Faster and more complete recovery** | • Breast cancer gene marker assays to assess the aggressiveness of the tumor, help tailor treatment and predict risk of recurrence  
• Hereditary hemochromatosis gene test to identify individuals at risk for excess iron absorption and potential accompanying morbidity; test can replace liver biopsies in some patients |
| **Less disability** | • Estimated glomerular filtration rate for early detection of chronic kidney disease and monitoring of kidney function, reducing disability  
• Lead screening for early detection of elevated blood levels in children reduces risk of central nervous system damage |
| **Fewer relapses or acute episodes** | • Sickle cell anemia testing to reduce risk of sickle cell anemia crisis by early detection  
• Cytochrome P450 testing to identify individual rate of drug metabolism thus reducing risk of adverse drug reactions  
• International Normalized Ratio (INR)/prothrombin time testing to monitor patients who are on warfarin therapy to prevent blood clots or severe internal bleeding |
| **Slower disease progression** | • C-reactive protein testing to detect individuals at high risk for serious cardiovascular events and candidates for statin therapy  
• Prostate specific antigen testing to screen at-risk men and monitor increases in year-to-year PSA values, which can identify prostate cancer that needs specific clinical intervention as opposed to watchful waiting  
• Cytochrome P450 2D6 testing to identify breast cancer patients who may not respond to tamoxifen therapy so that other therapies such as aromatase inhibitors can be prescribed  
• HIV genotyping to determine which antiviral drug combination is best |
| **Less need for long-term care** | • Antinuclear antibody test to detect lupus erythematosus, rheumatoid arthritis and other autoimmune disorders and help to determine appropriately targeted therapy  
• Maternal serum screening to identify pregnancies that could benefit from prenatal testing for spina bifida, Down syndrome and certain other birth defects |

*Health system goals adapted from Porter (2007)

**Sources**

C. Differentiating Screening and Diagnostic Tests

Clinical laboratory tests serve two overarching functions—screening and diagnosis:

- **Laboratory tests used for screening** assess the likelihood of the presence of a disease or condition in apparently healthy or asymptomatic individuals who are at sufficient risk for a condition to benefit from further investigation or preventive action. Screening may be targeted to a broad group (e.g., newborn screening) or subgroups at increased risk due to some combination of age, sex, family or personal history, race/ethnicity, disease state, or other factors.

- **Laboratory tests used for diagnosis** are performed to determine the presence or absence of a specific disease or condition in symptomatic individuals. Diagnostic tests also may be used for prognosis, enabling selection of clinical care alternatives and treatments, and for monitoring treatment effectiveness and guiding treatment modifications.

The regulatory policies applied to screening and diagnostic tests are similar; either may take one of two pathways to gain access to the market in the U.S.: clearance and approval by the Food and Drug Administration (FDA) or introduction under the Clinical Laboratory Improvement Amendments (CLIA). The pathway chosen typically depends on whether a test is to be marketed by a manufacturer as a product (such as in the form of a test kit) or as a laboratory developed test for use solely by the developing laboratory.

There are some key differences in the way that other policies and requirements related to clinical practice, clinical guideline development, and reimbursement (coverage and payment decisions), are applied to screening and diagnostic tests. These differences are discussed in subsequent sections of this report.

**Relationship of Screening and Diagnostic Tests to Prevention**

Preventive medicine has had an integral role in the growing movement toward explicit, evidence-based practice guidelines. Even so, the health system remains largely oriented to treating disease than preventing it. As such, the health and economic benefits that could accrue from wellness, prevention, and screening are far from being realized.

Three main types of prevention are primary, secondary, and tertiary. Screening and diagnostic tests contribute significantly to all three. Primary prevention is aimed at preventing the onset of disease and typically involves managing risk factors in healthy people that may lead to disease (e.g., high-sensitivity C-reactive protein to detect individuals at high risk for cardiovascular events). Secondary prevention is aimed at treating disease after its onset but before serious complications occur (e.g., Pap test to detect and enable treatment of cervical cancer in its earliest stages). Tertiary prevention is used in the later or final states of a disease with the aim of minimizing the degree of disability caused by the disease (e.g., blood tests to estimate the glomerular filtration rate to monitor the severity of chronic kidney disease and manage its complications).

Many serious medical conditions can be prevented or detected early with screening and effective treatment. Evidence-based screening and preventive services have contributed to sizable reductions in morbidity and mortality in heart disease, stroke, cancer, and other major...
sources of national disease burden in recent years.\textsuperscript{21, 22} For example, routine screenings for cervical and colorectal cancer contributed to a 14\% decline in overall mortality from cancer (the second leading cause of death overall) during this same period.\textsuperscript{23}

The growing body of evidence demonstrating the ability to prevent disease by screening for risk factors and adopting early intervention strategies is leading to increased use of testing for primary prevention in asymptomatic individuals. Advances in genetic and molecular testing may increasingly support the role of such tests in primary prevention through identification of individuals with predispositions for disease.\textsuperscript{24} To date, however, clinical guidelines and payer coverage policies have focused largely on laboratory testing in secondary and tertiary prevention.\textsuperscript{25, 26}

\section*{II. Validity and Utility of Screening and Diagnostic Tests}

The development, adoption, and diffusion of laboratory tests are influenced by a widening group of stakeholders that seek well-founded evidence to support decisions about whether or how to develop technology, allow it on the market, acquire it, use it, pay for its uses, and more.\textsuperscript{27} These stakeholders include regulators, clinicians, patients, laboratory directors, hospital managers, payers, and government policymakers, as well as manufacturers and laboratories.

Value may be assessed through clinical trials and other studies, as well as health technology assessments and other systematic appraisals of evidence that examine safety, efficacy, feasibility, effectiveness, appropriateness, and cost of tests, as well as other clinical, social, economic, and ethical implications related to their use.\textsuperscript{28, 29} Such assessments contribute to the knowledge base for improving the quality of health care, including the development and updating of various standards and guidelines.

Assessing the value of tests may involve multiple determinations grouped into three main concepts: analytic validity, clinical validity, and clinical utility, as shown in Figure 1. While these concepts are useful in all laboratory testing, they are gaining increased attention and further definition in the realm of genetic and genomic testing. In addition to these components, cost-related assessments are of growing interest among clinicians, health care institutions, public and private sector payers, and policymakers seeking to improve value per health expenditure.\textsuperscript{30, 31}
A. **Analytic Validity**

The overarching value of laboratory testing is the generation of objective, scientific data about patient health to inform clinical decision-making. High levels of technical performance are required to ensure quality of test results. Regarding analytic validity, laboratory test systems are designed to ensure test-to-test accuracy, precision, and robustness. Measurements aim to determine whether the test performs reliably to specifications and delivers accurate information in a laboratory setting. **Accuracy**, which is the degree of closeness of a measured or calculated quantity to its actual (true) value as compared to a reference “gold standard,” includes analytic sensitivity and analytic specificity. **Analytic sensitivity** is the probability that a test will detect a specific analyte (e.g., a biomarker or genotype) when it is truly present in a specimen. **Analytic specificity** is the probability that a test will be negative when a specific analyte is truly absent in a specimen. **Precision** (or reliability, reproducibility or repeatability) is the degree to which further measurements show the same or similar results. **Robustness** refers to the resistance of test results to small changes in preanalytic or analytic variables associated with testing. 32,93

B. **Clinical Validity**

Clinical validity refers to the ability of a test to detect the condition (disease or disorder) that is associated with an analyte measurement and predict the probability of having the condition based on the test result. It includes clinical sensitivity, clinical specificity (incorporating analytical validity), and positive and negative predictive values. In genetic testing, clinical validity may also be affected by factors that confound the association between a genotype and a phenotype, such as reduced penetrance (i.e., the proportion of individuals with a disease-related genotype or mutation who develop disease), variable expressivity of the disease among individuals with the same genotype, and other genetic or environmental factors. 93
Clinical sensitivity refers to the proportion of individuals with a specified condition whose test results indicate that the condition is present, e.g., how often the genetic mutation (genotype) that is associated with the condition (phenotype) is identified in people who truly have the condition. Tests with high clinical sensitivity are useful for “ruling out” a condition if an individual tests negative. Clinical specificity refers to the proportion of individuals who do not have a specified condition whose test results indicate that the condition is not present. Tests with high specificity are useful for “ruling in” a condition if a person tests positive.

Positive predictive value (PPV) refers to the probability that an individual with a positive test has, or will develop, the specified condition that the test is designed to detect. Negative predictive value (NPV) refers to the probability that an individual a negative test result actually does not have the specified condition. Cutoff values for a test (defining presence or absence of a condition by level of the test result) represent a compromise between clinical sensitivity and clinical specificity. PPV and NPV depend not only on sensitivity and specificity but on the prevalence of a condition within the population being tested and changes in disease patterns over time. As such, PPV and NPV are not constant performance characteristics of a test. For example, if a disease is very rare in the population, even tests with high sensitivity and high specificity can have low PPV, generating more false-positive than false-negative results. In these instances, without educational efforts, some providers may not fully understand the limits of a test’s predictive value and may overestimate the probability of disease in patients with a positive result. The example of rapid testing for the influenza virus (Box 3) demonstrates the fluctuations in PPV and NPV according to prevalence in the population.

Box 3
Example: Rapid Testing for Influenza

An example of the inter-relationships between sensitivity, specificity, and predictive values can be demonstrated with rapid testing for influenza virus as disease patterns fluctuate from low- to mid-prevalence periods to peak flu season, then tapering back down to mid- and low-prevalence periods. Influenza viral pathogens cause significant mortality and morbidity; the recent re-emergence of a novel human influenza A virus (H1N1) posed a serious personal and public health threat. Several types of rapid molecular tests have helped to detect the virus, including point-of-care tests used by physicians in clinical practice and conventional and real-time polymerase chain reaction (PCR) assays that usually have high specificity. One study estimated PPV and NPV using six-year averages of weekly influenza activity reported via CDC surveillance data and a rapid molecular test with sensitivity of 70% and specificity of 90%. During a typical flu season, PPV fluctuated from 17% during low prevalence to 71% at peak while the NPV decreased from 99% during non-flu periods to 89% during flu season.

The following two examples of screening for fragile X syndrome and KRAS gene mutation support the clinical validity of a genetic test (Box 4) and a pharmacogenomic test (Box 5) for clinical decision-making.
Fragile X syndrome is one of the most common inherited causes of mental impairment ranging from learning disabilities to autism and severe mental retardation. Because of the lack of definitive clinical diagnostic criteria, molecular tests are important for detection of individuals with fragile X. The standard screening/diagnostic test is Southern blot analysis although there is increasing interest and evidence to support use of PCR-based methods. Diagnostic tests can ‘rule in’ or ‘rule out’ disease in children with developmental, speech, language, or motor delay. Individuals with a family history of fragile X can be tested to determine if they may be asymptomatic carriers of the disorder. Individuals known to be carriers also can use screening tests to determine a prenatal diagnosis of the fragile X mutation. While there is no cure for fragile X syndrome, early diagnosis facilitates clinical decision-making and therapeutic planning of interventions for speech and language, behavior, cognitive and gross motor development, sensory integration, and daily living activities, in order to improve quality of life for those with the disease.

Colorectal cancer is the third most commonly diagnosed cancer and third-highest cause of cancer death for men and women in the U.S. Up to 20% of patients with colorectal cancer will present with metastases, with a 5-year survival of less than 10%. Cetuximab (Erbitux®) and panitumumab (Vectibix®) are monoclonal antibodies that bind to the epidermal growth factor receptor (EGFR), inhibiting growth of metastatic colorectal cancer. However, these drugs have considerable adverse effects. Moreover, a proportion of patients with colorectal cancer have tumors with a somatic KRAS mutation that affects tumor response to EGFR inhibitors. Diagnostic testing for KRAS gene mutational status is an important predictor of non-response to EGFR-targeted therapy, with great value to clinical decision-making. Retrospective analyses of data from several randomized controlled trials involving patients receiving combination cetuximab and chemotherapy demonstrated that individuals with normal (or wild-type) KRAS had significant improvements in tumor response and that few or none of those with mutated-type KRAS responded to cetuximab. In July 2009, the FDA announced revisions to the prescribing information for EGFR inhibitors and colorectal cancer, requiring inclusion of information on variations in the KRAS gene that may affect patient response to the drugs.

As noted above, clinical validity can vary with the prevalence of a condition within the population being tested, changes in disease patterns over time, and other factors, e.g., age, race/ethnicity, family history, severity of disease, and comorbidities. For many conditions, patients with severe disease are more likely to have positive tests than those with mild or early-stage disease, and healthy patients are more likely to have negative tests than those with significant comorbidities. For example, C-reactive protein (CRP), a sensitive marker of the acute-phase response, has been associated with future cardiovascular endpoints independently of other risk factors. A recent study of myocardial infarction survivors receiving regular high-sensitivity CRP (hsCRP) monitoring tests examined the effect of comorbidities and environmental factors on diagnostic value. There were larger variations in hsCRP values for males, smokers, and patients with increased HbA1c levels >6.5%. As a result, one or two hsCRP measurements may not be sufficient to adequately characterize cardiac risk in different patient groups after myocardial infarction.
A complicating factor in examining the effect of individual patient characteristics on clinical validity is the lack of information available for these stratifications. Several reports in the literature state that, often, researchers do not consistently, or in some cases adequately, report on important patient characteristics and study design features associated with their investigations of test accuracy and predictive value. Improvements in the reporting on patient characteristics could augment clinical research and guideline development for screening and diagnostic tests as well as facilitate tailored recommendations to specific patient subgroups.

C. Clinical Utility

Clinical utility refers to the evidence that use of a test is associated with improved clinical outcomes and its usefulness to patient and clinician decision-making. It encompasses effectiveness (utility in real clinical settings) or efficacy (utility in controlled settings such as clinical trials) and the net balance of risks and benefits associated with using a test in clinical practice. A test with clinical utility yields results, whether positive or negative, that provide information of value to the patient, clinician, or others involved in making decisions about management of a patient with a given condition. In some instances in which there are no interventions or treatments available to prevent or treat disease, a test result can have clinical utility for life-planning purposes.

The specific outcomes for which tests can provide clinical utility vary by condition, but include the categories of mortality, morbidity, adverse events, and quality of life. Some biomarkers are used as intermediate or surrogate outcomes, in that they are known to be predictive of long-term health outcomes. Patient functional status, patient satisfaction, and other patient-reported outcomes, sometimes known as humanistic outcomes, are of increasing importance in patient-centered care.

One of the challenges of demonstrating clinical utility of testing is generating direct evidence of the impact of a test on health outcomes, which can require long follow-up times and is subject to the various intervening decisions and other environmental factors that can influence ultimate patient outcomes. Still, there are notable examples of such direct evidence, including that of FOBT screening for colorectal cancer, described in Box 6.
In many instances, the impact of laboratory testing on health outcomes is inferred through linkages between the tests and intermediate outcomes (or endpoints) or surrogate outcomes that are known to be strongly associated with long-term health outcomes. Surrogate outcomes, such as \( \text{HbA1c} \) for diabetes morbidity and CD4 cell counts and viral RNA levels for progression of AIDS, are generally easier to measure than the long-term health outcomes with which they are associated and, thus, are used more commonly to determine the effects of testing and other health care interventions. These associations must be validated in appropriately long-term natural history studies or other appropriate study designs. Some surrogate outcomes are better predictors of patient outcomes than others. Examples of other validated ones include: cardiac troponin for identifying and predicting outcome of acute coronary syndrome, prostate-specific antigen (PSA) levels for predicting the incidence and course of prostate cancer, INR/prothrombin time for predicting thrombotic events, and \( \text{HER2/neu} \) for predicting response to treatment with trastuzumab (Herceptin) in women with breast cancer.

1. Comparative Effectiveness Research

Recent efforts to augment the national capacity for conducting and using comparative effectiveness research (CER) will increase the interest in such research, particularly for clinical utility, of clinical laboratory testing. As is clear in recent reports by the IOM and the Federal Coordinating Council for Comparative Effectiveness Research on recommended national priorities for CER, this research will address the full range of interventions, including drugs,

---

\( \text{A biological marker (biomarker) is an objectively measured characteristic that is used as an indicator of normal biological processes, natural history of disease, or effect of a therapy. A clinical endpoint is a characteristic or variable that reflects how a patient feels, functions, or survives. An intermediate (nonultimate) endpoint is a true clinical endpoint (e.g., a symptom or measure of function, such as symptoms of angina frequency or exercise tolerance) but not the ultimate outcome of the disease (e.g., survival or the rate of other serious and irreversible morbid events). A surrogate endpoint is intended to substitute for or predict a clinical endpoint on the basis of epidemiological, therapeutic, pathophysiological, or other scientific evidence.} \)
biologics, tests, imaging, and medical and surgical procedures, as well as health care delivery system organization, delivery, and financing.

While there is no standard definition of CER, most address a combination of the following attributes or emphases:

- Direct comparisons of alternative interventions (as opposed to comparison with placebo or indirect comparisons)
- Effectiveness (in realistic health care settings) rather than efficacy (in ideal circumstances)
- Health care outcomes (e.g., morbidity, mortality, QoL, adverse events, and symptoms) rather than intermediate or surrogates endpoints
- Use of primary and secondary data collection, with emphasis on head-to-head comparisons in RCTs and practical/pragmatic clinical trials as well as observational studies (using registries, claims data, electronic health records) and systematic reviews

The definition used by the IOM is:

> The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels. CER’s distinguishing characteristics include informing a specific clinical or policy decision, comparing at least two approaches or interventions, describing results at the subgroup level, measuring benefits in real-world populations, and applying appropriate methods and data sources. — Institute of Medicine

Two main ways in which laboratory testing will be involved in CER are as an indicator of intermediate and long-term health outcomes in comparisons of other interventions and as the index (focal) intervention in a direct comparison with alternatives (Box 7). Pursuant to involvement as index interventions, among the top tier of 25 CER priorities recommended by the IOM in its June 2009 report were:

- Compare the effectiveness of various screening, prophylaxis, and treatment interventions in eradicating methicillin resistant *Staphylococcus aureus* (MRSA) in communities, institutions, and hospitals.
- Compare the effectiveness of genetic and biomarker testing and usual care in preventing and treating breast, colorectal, prostate, lung, and ovarian cancer, and possibly other clinical conditions for which promising biomarkers exist.
Box 7
Main Roles of Clinical Laboratory Tests in CER

- As an indicator of patient intermediate and long-term outcomes in comparisons of other treatments/interventions, particularly those related to priority health conditions, such as diabetes, obesity, heart disease, stroke, kidney disease, HIV/AIDS, mental health/substance abuse, pneumonia, cervical and colon cancer, and pregnancy. For example, HbA1c could be used as one indicator to evaluate the effectiveness of alternative oral medication regimens (e.g., sulfonylureas, biguanides, dipeptidyl peptidase IV inhibitors, or combination therapy) to treat type 2 diabetes. Laboratory values can be useful in quantifying baseline characteristics, assessing intermediate outcomes, conducting subgroup analysis, and more.

- As the subject of analysis for head-to-head comparisons of alternative laboratory tests or comparisons of a laboratory test to another test intervention for a particular health care condition. For example, a CER study may compare a new molecular test assessing the gene expression profile of RNA from the peripheral blood of heart transplant recipients (i.e., AlloMap HTx based on blood draws) to the current standard of care assessing RNA from invasive endomyocardial (heart tissue) biopsy procedures. Another CER study might evaluate the effectiveness of HPV testing for cervical cancer compared to conventional Pap smear testing.

2. Challenges of Demonstrating Causal Effect of Testing on Outcomes

The size of the literature documenting clinical utility of laboratory testing, including patient outcomes, is small relative to the corresponding literature on analytical validity and clinical validity. Typically, there are substantial difficulties in establishing direct causal links between ordering a test and changes in mortality, morbidity, quality of life, and other major patient health outcomes, as a test is likely to be just one of many interventions and environmental and behavioral determinants of patient outcomes. In order to influence outcomes, a laboratory test must be ordered, conducted, returned with results on a timely basis, appropriately interpreted, and affect a decision for further diagnosis or treatment that results in changes in outcomes. This includes instances in which laboratory testing is used to monitor intermediate outcomes of treatment in order to guide the next treatment decision toward desirable outcomes.

As part of the larger clinical process, laboratory testing can affect other aspects of patient care, such as timing, efficiency, and patient and provider satisfaction. Providers may interpret and act on laboratory test results differently and unpredictably, which can affect patient outcomes. Providers and patients may ignore positive results or proceed with a clinical intervention despite a negative test.

Several factors can confound the ability to conduct population-based laboratory outcomes studies. These include the lack of standardization of data collection and reporting methods, lack of agreement regarding appropriate analysis of data (e.g., whether or not to risk-adjust data), and the high cost of collecting outcomes data. Outcomes measurement can be severely constrained by sample size, missing or incomplete test results in patients’ medical records, limited ability to perform risk adjustment with data abstracted from administrative records, and higher cost to abstract data from medical records. Other challenges can include the inability to conceal the identity of tested versus non-tested patients (in blinded RCTs), the number of patients or volunteers required for a study to achieve statistical significance, and what can be long periods of time between a test and a patient outcome of interest, as suggested by the example of screening for colorectal cancer noted above. For these reasons, the assessment of the impact of
laboratory tests on health outcomes has relied more often on intermediate or surrogate outcomes, computer simulations (modeling), and use of observational studies and other non-experimental study designs. However, several trends should improve the capacity for conducting outcomes studies of clinical laboratory testing. These include the current emphasis on evidence-based decision-making, advances in data mining and other techniques for analyzing claims data and other administrative and observational data sets, expansion of electronic health records systems and networks, and the increase in funding for CER and related methods and infrastructure development.

D. Economic Outcomes and Impact

Interest in cost-effectiveness analyses and related studies of the economic impact of health care technologies continues to increase among stakeholders. In screening and diagnostic testing, the body of such evidence is small, though growing. Recent debate about whether savings can be realized from greater federal investment in preventive and wellness services and their cost-effectiveness relative to other types of health care highlights the importance of rigorous, policy-relevant demonstrations of the economic impact of laboratory testing used in screening and diagnosis. The main types of analysis and their application to testing are summarized below. While these analyses are used more widely and various expert groups and journals are working to improve their quality, considerable variation persists in the methods, results, and reporting of these analyses.

Cost-of-illness analysis is a determination of the economic impact of an illness or condition (typically on a given population, region, or country) e.g., of heart disease, diabetes, or cancer, including associated treatment costs. Cost-of-illness studies are used to estimate the burden of disease and provide an economic context for opportunities to lowering such burdens with health care interventions. For example, Box 8 describes how cost-of-illness provides the context for potential savings from MRSA testing.

**Box 8**

**Example: Cost-of-Illness Analysis of Rapid MRSA Testing**

The economic burden of health care-acquired infections is substantial. Adjusted to 2007 dollars, CDC estimates that the direct cost per case for all HAIs ranges from $20,549 to $25,903, for a total of roughly $36 to 45 billion in annual costs. Of these costs, $3.5 to 10 billion are associated with surgical site infections; $670 million to $2.7 billion to central venous line associated bloodstream infection, $1 to $1.5 billion in ventilator-associated pneumonia, and $1.2 to $1.6 billion catheter-associated urinary tract infections. Approximately 50% of all HAIs are MRSA-related. A nine-year study at Brigham and Women’s Hospital (Boston) found that routine surveillance cultures and subsequent contact precautions decreased incidence of bacteremia by 75% in ICUs, 40% in non-ICU areas, and 67% hospital-wide. After issuing a mandate for MRSA testing in all high-risk units, the VA reduced infection rates in ICUs by 79% at their Palo Alto site. This large reduction in rates of MRSA infection rates at a major facility suggests the potential for broader-scale dramatic reductions in the prevalence of MRSA and the morbidity, mortality, and costs associated with that pathogen.

Cost-minimization analysis is a determination of the least costly among alternative interventions that are assumed to produce equivalent outcomes. Several studies have examined the cost of POCT relative to central laboratory testing using metrics such as cost-per-test (including estimates for the
cost of equipment, supplies, labor, and other variables) and laboratory test turnaround time.\textsuperscript{70, 71} Box 9 provides an example of cost-minimization analysis for bedside glucose testing.

**Box 9**

**Example: Cost-Minimization Analysis of Bedside Glucose Testing**

A study reported in 2004 compared the analytical costs of central laboratory glucose testing and semi-automated bedside glucose testing (BGT) among 445 hospitals enrolled in the College of American Pathologists’ Q-Probes quality assurance program. Results showed different distributions of costs across three main types of sites. The median (10th-90th percentile range) analytical costs per glucose test were $1.18 dollars ($0.36-$5.59) for central laboratories, $1.96 ($0.77-$9.51) for high-volume BGT sites, and $4.66 ($1.02-$27.54) for low-volume BGT sites.\textsuperscript{72} In addition to being higher than costs for central laboratories, costs for BGT were highly variable and dependent on volume. In this instance, the cost-per-test for central laboratories was the lowest, although ranges of medians showed considerable overlap. Of further consideration is the potential clinical value to the provider and patient in having an instantaneous (bedside) result, which could lead to cost savings if it avoids helps to avoid unnecessary care.

Cost-effectiveness analysis (CEA) compares costs in monetary units with outcomes in quantitative non-monetary units, e.g., reduced mortality or morbidity. CEA aims to weigh the health and economic tradeoffs of alternative health interventions. It calculates the incremental (marginal) cost per incremental unit of effectiveness achieved through use of an intervention versus the standard of care or other alternative. Results are presented as net cost per health outcome, such as cost per case prevented or cost per life saved; this is also known as an incremental cost-effectiveness ratio (ICER).\textsuperscript{73, 74} One type of cost-effectiveness analysis, sometimes referred to as cost-utility analysis (CUA), compares costs in monetary units with outcomes in terms of their utility, usually to the patient, measured, e.g., in quality-adjusted life years (QALYs), a unit combining quality of life and length of life.\textsuperscript{74} Although U.S. payers use no formal threshold for an acceptable cost per QALY, incremental cost-effectiveness ratios of $50,000-$100,000 or less per QALY are generally regarded as acceptable value.\textsuperscript{75} CEA/CUA can be conducted from different economic perspectives, e.g., of the clinician, payer, patient, or society at large. Various types of cost-effectiveness analyses have been conducted in such areas as prenatal genetic screening, screening for *Chlamydia trachomatis*, FOBT for colon cancer, HER-2/neu testing of breast cancer, HIV screening, and nucleic acid testing for safety of donated blood.\textsuperscript{76-81} An example is highlighted in Box 10.

Cost-benefit analysis (CBA) compares costs and benefits, both of which are quantified in common monetary units. Two basic approaches for cost-benefit analysis (CBA) are ratio approach and the net benefit approach. The ratio approach indicates the amount of benefits (or outcomes) that can be realized per unit expenditure on a technology vs. a comparator. In the ratio approach, a technology is cost beneficial vs. a comparator if the ratio of the change in costs to the change in benefits is less than one. The net benefits approach indicates the absolute amount of money saved or lost due to a use of a technology vs. a comparator. In the net benefits formulation, a technology is cost-beneficial vs. a comparator if the net change in benefits exceeds the net change in costs. Due largely to the difficulty associated with assigning monetary values to years of life or health outcomes, few true CBAs have been conducted for health care technologies, including for laboratory testing.
Budget impact analysis (BIA) provides an estimate of the financial impact on capital and operating budgets (and may include analysis of cost offsets) of the adoption and use of a health care intervention in a given patient population, health care setting, or other context of the introduction of a technology or service. BIA often includes analysis of cost offsets. CEA and BIA can provide complementary analyses for evaluating technology acquisition and use decisions by health care providers, policy makers, and others.

Box 10
Example: Cost-effectiveness Analysis of Factor V Leiden Test for Thrombophilia

Thrombophilia is the propensity to develop potentially venous thromboembolism (VTE) caused by hereditary defects in one or more of the clotting factors. The most common mutation is factor V Leiden associated with activated protein C (APC) resistance (an anticoagulation enzyme). The factor V Leiden mutation has a relatively high prevalence in the general population, including 5% in Caucasians, and accounts for 85-95% of APC resistance cases. Patients at high-risk for VTEs include those using oral estrogen preparations, who are pregnant, or having major surgery. Increased screening for the factor V Leiden mutation has been suggested as a strategy for preventing VTEs. The factor V Leiden test is a genetic test for detecting the presence of the mutation.

A recent systematic review and CEA conducted in the U.K. examined the cost effectiveness of screening regimens for four types of high-risk individuals. The cost-effectiveness ratio was expressed as costs per adverse clinical complication prevented when comparing universal screening and selected screening to no screening for each of the four risk groups. The study was based on a hypothetical model of 10,000 patients in each of group. The results for universal screening versus selective screening for each of the four groups were as follows. Shown are the number of complications prevented by each strategy and the ICER, i.e., the ratio of the incremental cost for each strategy to the number of complications prevented.

<table>
<thead>
<tr>
<th></th>
<th>Universal Screening</th>
<th>Selective Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complications Prevented</td>
<td>ICER*</td>
</tr>
<tr>
<td>Combined oral estrogen:</td>
<td>3 £200,402 ($323,230)</td>
<td>1 £79,085 ($127,560)</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td>42 6,824 (11,010)</td>
<td>15 2,446 (3,950)</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>59 81,554 (131,540)</td>
<td>7 81,250 (131,030)</td>
</tr>
<tr>
<td>Orthopedic surgery</td>
<td>88 14,129 (22,790)</td>
<td>26 9,136 (14,740)</td>
</tr>
</tbody>
</table>

*ICER: Cost in UK£ (and US$) per complication prevented

The analysis offers important tradeoffs for consideration. First, for all four risk groups, universal screening led to prevention of more complications than did selective screening, e.g., 42 vs. 15 in the hormone replacement therapy risk group. Second, for all four risk groups, selective screening yielded lower ICERS than universal screening, demonstrating better cost-effectiveness for that strategy in general. The greatest improvements in cost-effectiveness resulting from selective screening were in the combined oral estrogen risk group (61% decrease) and the hormone replacement therapy risk group (64% decrease). Third, despite their relative magnitudes, the ICERS of both universal and selective screening of the hormone replacement therapy group were the lowest among the full set of eight strategies.

In its August 2009 letter to Congress about the potential economic impact of greater federal spending on preventive and wellness services, the CBO focused on how it would “score” such spending. Scoring by the CBO for determining the net impact of proposals to expand governmental support for a designated program is a form of BIA. CBO stated that, although
particular types of preventive care would have different effects on costs, available evidence suggests that expanded use of most preventive services results in higher, not lower, overall spending.\(^1\) CBO also distinguished its scoring from determinations of the cost-effectiveness of spending on such services, noting that, while most would add to spending, many would be cost-effective, i.e., their benefits to health would be justified by the additional net costs.\(^64\)

Further deliberations on these matters in the context of national health reform and other proposals to improve the U.S. health care system should address how to assess the economic impact of these services. This should include, for example, the extent to which CBO scorekeeping rules and procedure are limited with respect to the broader costs and benefits (i.e., beyond federal budgetary effects) that would accrue to the nation, how some preventive services (or selective use of them in certain high-risk populations) are more likely to be cost saving or cost effective than others, and various methodological aspects such as the time horizon of analysis and inclusion of direct and indirect costs.

As is so for many other types of health care technology, greater investment is needed to build the body of evidence on the economic impact of laboratory testing. Findings about laboratory tests that are truly cost-saving or highly cost-effective, such as colorectal cancer screening in adults aged 50-75 and screening young women for chlamydial infection,\(^22,83\) should be broadly recognized and used as models for further work. Among key considerations in such analyses are that the economic impacts of specific tests can vary significantly depending on the populations targeted (with testing of high-risk populations more likely to be cost-effective) and tradeoffs between the cost of greater testing frequency and yield of cases detected.

III. Analytic Frameworks to Appraise Evidence of Clinical Value

Health care decision-makers increasingly rely on health technology assessments (HTAs), systematic reviews, and other evidence reports to support decisions and policies related to clinical care, practice guidelines, coverage policies, public health services, and other purposes.\(^27\) Usually, these evidence reports are prepared after clinical laboratory tests have received regulatory approval by the FDA or met specifications of CLIA, as appropriate. Stakeholder organizations may undertake or sponsor their own HTAs for various reasons, such as to support decisions regarding health plan benefits, clinical practice guidelines, technology acquisition, public health services, coverage and payment, and investment in technology.\(^27\)

HTAs pertaining to testing are often organized according to analytic frameworks such as the one shown in Figure 2. These analytic frameworks map the linkages between the populations to be tested, test results, interventions (performance of a test, treatment), change in intermediate outcomes, and change in patient outcomes.\(^84\) A set of specific key questions are developed to provide the conceptual context for the study and guide retrieval of the available scientific evidence to assess these linkages.

\(^1\) Among the points made by CBO with regard to cost offsets is that some of the proposed additional federal spending on these services would only substitute for existing spending on these services from other sources, resulting only in a cost shift to the federal government with no additional benefits.
A. Organizations that Conduct Evidence Appraisals

Many different organizations conduct HTAs or other evidence reports of health technologies. Some organizations evaluate technologies for the spectrum of clinical applications, including screening, diagnosis, treatment, and monitoring, whereas others limit their evaluations to certain types of interventions or services.

The **U.S. Preventive Services Task Force (USPSTF)** is an independent, non-federal advisory group sponsored by the Agency for Healthcare Research and Quality (AHRQ) that conducts evidence reviews of screening and preventive services used in primary care. The panel of experts that serve on the USPSTF evaluates the strength of the scientific evidence and makes recommendations for use of preventive services based on its findings. The type of clinical analytic framework used by the USPSTF is recognized as a standard approach for mapping linkages between patient populations, interventions and outcomes, shown in Figure 2.

The **Evaluation of Genomic Application in Practice and Prevention (EGAPP)** initiative, sponsored by CDC, conducts rigorous evaluations of genetic tests and other genomic applications for clinical and public health practice in the U.S. An independent, non-federal advisory group, EGAPP currently focuses on evaluation of tests with wide population application (e.g., high disorder prevalence, higher frequency of test use), those with potential to affect clinical and public health practice (e.g., emerging prognostic and pharmacogenomic tests), and those for which there is significant demand for information. EGAPP has an evidence grading system that is based on hierarchies for each of analytical validity, clinical validity, and clinical utility (described below).

The **Community Services Task Force**, sponsored by the Centers for Disease Control and Prevention (CDC), conducts evidence reviews to develop guidelines for screening and prevention for services related to infectious diseases and other public health concerns, e.g., health care-acquired infections such as *S. aureus* (staph infection), sexually transmitted diseases such as *Chlamydia trachomatis* and HIV/AIDS, and potential epidemic and pandemic diseases such as influenza.

Through its **Effective Health Care Program**, including its 14 Evidence-based Practice Centers (EPCs), located primary at academic health centers, and other programs, AHRQ funds systematic evidence reviews of relevant scientific literature on clinical, behavioral, and organizational topics. These are used by a wide array of stakeholders to inform the development of educational materials and tools, quality measures, guidelines, coverage decisions, and research agendas. In particular, the Oregon EPC, based at Oregon Health & Science University, supports the work of the USPSTF, and the EPCs also support EGAPP. Although these particular EPC evidence reviews are prepared initially for USPSTF and EGAPP, they are placed in the public domain and are used by others. In addition to biomedical research that supports the development and validation of many and diverse tests, the **National Institutes of Health**, including the National Heart, Lung, and Blood Institute and the National Cancer Institute, undertake studies to support development of clinical guidelines on use of screening and diagnostic tests.

**Health professional societies**, particularly those associated with medical specialty areas, have guideline development committees and other expert advisory groups and panels that conduct independent evaluations of scientific evidence and make recommendations. Examples of the many such organizations engaged in these activities are the American Academy of Family Physicians, American Academy of Pediatrics, American College of Cardiology, American
College of Physicians, and American Society of Clinical Oncology. Non-profit associations devoted to certain health problems, such as the American Cancer Society, American Diabetes Association, and American Heart Association, are also active in sponsoring advisory groups and expert panels that evaluate evidence and develop guidelines related to testing. Nearly 300 health professional organizations have clinical guidelines included in the National Guideline Clearinghouse, managed by AHRQ.8

Many private sector payers have internal research arms that undertake assessments to determine coverage and reimbursement. Examples of major commercial health plans that conduct formal reviews of new technologies include Aetna, CIGNA, UnitedHealthcare, and WellPoint.14 HTA vendors such as the Cochrane Collaboration, BlueCross BlueShield Association Technology Evaluation Center (TEC), ECRI Institute, and Hayes, Inc.,h produce HTAs of tests using in screening and diagnosis.87, 88 BlueCross BlueShield TEC and ECRI Institute are also EPCs.

Figure 2. Generic Clinical Analytic Framework for Screening and Diagnostic Tests

Key questions correspond to numbers above.

(1) Is there direct evidence that the test reduces morbidity, mortality, and/or quality of life?
(2) What is the prevalence of disease in the target group? Can a high-risk group be reliably identified?
(3) Can the test accurately detect the target condition? (a) What are the sensitivity and specificity of the test? (b) Is there significant variation between examiners in how the test is performed? (c) In actual testing programs, how much earlier are patients identified and treated?
(4) Does treatment reduce the incidence of the intermediate outcome? (a) Does treatment work under ideal, clinical trial conditions? (b) How do the efficacy and effectiveness of treatments compare in community settings?
(5) Is the intermediate outcome reliably associated with reduced morbidity and/or mortality?
(6) Does treatment improve health outcomes for people diagnosed clinically? (a) How similar are people diagnosed clinically to those diagnosed by screening? (b) Are there reasons to expect people diagnosed by screening to have even better health outcomes than those diagnosed clinically?
(7) Does testing result in adverse effects? (a) Is the test acceptable to patients? (b) What are the potential harms, and how often do they occur?
(8) Does treatment result in adverse effects?


h Hayes, Inc., provides technology assessment reports for health plans, managed care companies, hospitals, and health networks and offers training programs to facilitate participants’ understanding of the HTA process.
B. Direct and Indirect Evidence of Test Impact on Outcomes

The analytic framework of the type shown in Figure 2 and used by such groups as the USPSTF and EGAPP provides routes for evaluating direct and indirect evidence of the impact of testing on patient outcomes. The direct route (#1 in Figure 2) follows a population that has been tested long enough to determine whether the test had any ultimate impact on specified patient outcomes, e.g., of improvements in mortality, morbidity, or quality of life. Prospective experimental study designs, particularly RCTs, are the most rigorous for assessing test impact along this direct route. As discussed below, generating direct evidence in this manner can be challenging in many instances, and such studies often are not available.

When direct evidence of the impact of testing on patient outcomes using RCTs or similarly rigorous methods is unavailable, impact on outcomes may be established through a chain of indirect evidence, also shown in Figure 2. This would include assembling evidence from separate studies establishing the ability of the test to detect a target condition (#3) and a treatment choice that affects intermediate outcomes (#4) or patient outcomes directly (#5), and evidence that these intermediate outcomes are strongly associated with patient outcomes (#6). Separate studies can be used to assess adverse effects of the test (#7) or of the treatments (#8). Often at issue is whether such indirect chains of evidence are sufficient to provide expert panels enough confidence to conclude that a test has an impact on outcomes and can be recommended for clinical use.

C. Evidence Hierarchies Used for Screening and Diagnostic Testing

HTAs and other evidence reports appraise different types of studies in the available bodies of evidence pertaining to the interventions being assessed. The various study designs offer characteristic strengths and limitations for answering evidence questions. In most instances, whether for evidence reports sponsored by government, private non-profit, or commercial programs, determinations of the methodological quality of the available studies are based on evidence hierarchies and related quality criteria. Most of these hierarchies are arranged according to the relative rigor of various study designs to account for the internal validity of the causal effect of an intervention on specified outcomes. Typically, these hierarchies rank RCTs, or systematic reviews of RCTs or meta-analyses of RCTs, as the highest form of evidence. Although systematic reviews (some of which include meta-analyses) and meta-analyses do not themselves constitute primary evidence, they are compilations or syntheses of evidence that may hold greater weight than the individual RCTs (or other primary studies) of which they are comprised.

Two evidence hierarchies of particular importance for laboratory tests used in screening and diagnosis are those of the USPSTF and EGAPP, shown in Table 2 and Table 3, respectively. The USPSTF and EGAPP account not only for study design but for the quality of study execution, e.g., “properly powered and conducted RCTs” and “well-designed longitudinal cohort studies.” In appraising the strength of evidence for clinical preventive services, the USPSTF places RCTs and systematic reviews or meta-analyses of RCTs at the top level. EGAPP has evidence hierarchies for each of analytic validity, clinical validity, and clinical utility. EGAPP places meta-analysis of RCTs at the highest level, followed by single RCTs. In addition to its evidence hierarchies, the USPSTF has a grading system for its recommendations, shown in Table 4 and Table 5. EGAPP has a similar system (not included here).
Despite its traditional emphasis on evidence from RCTs, the USPSTF has become somewhat more flexible in its approach to grading of evidence and recommendations, as described in its methods revisions made during 2007-08. The USPSTF’s approach may benefit further by continuing to consider alternative evolving methods for evidence appraisal, including those described below. In formulating their recommendations, both USPSTF and EGAPP rely not only on strength or quality of evidence but on the anticipated magnitude of net benefit. Weighing both enables providing favorable recommendations even when the evidence is not of the highest level. For example, as is apparent in Table 4, the certainty of net benefit (i.e., benefit minus harms based on available evidence) of a preventive intervention may be only moderate, due, for instance, to having level II-2 or II-3 evidence rather than level I evidence. However, if the expected magnitude of net benefit is high or moderate, USPSTF can assign a “B” grade recommendation, meaning that the USPSTF recommends offering or providing the intervention, as indicated in Table 4.

### Table 2. USPSTF Hierarchy of Research Design

<table>
<thead>
<tr>
<th>Level</th>
<th>Evidence Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Properly powered and conducted randomized controlled trial (RCT); well-conducted systematic review or meta-analysis of homogeneous RCTs</td>
</tr>
<tr>
<td>II-1</td>
<td>Well-designed controlled trial without randomization</td>
</tr>
<tr>
<td>II-2</td>
<td>Well-designed cohort or case-control analytic study</td>
</tr>
<tr>
<td>II-3</td>
<td>Multiple time series with or without the intervention; dramatic results from uncontrolled Experiments</td>
</tr>
<tr>
<td>III</td>
<td>Opinions of respected authorities, based on clinical experience; descriptive studies or case reports; reports of expert committees</td>
</tr>
</tbody>
</table>


### Table 3. EGAPP Hierarchies of Data Sources and Study Designs for Components of Evaluation

<table>
<thead>
<tr>
<th>Level</th>
<th>Analytic Validity</th>
<th>Clinical Validity</th>
<th>Clinical Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (highest)</td>
<td>Collaborative study using a large panel of well-characterized samples</td>
<td>Well-designed longitudinal cohort studies</td>
<td>Meta-analysis of RCTs</td>
</tr>
<tr>
<td></td>
<td>Summary data from well-designed external proficiency testing schemes or interlaboratory comparison programs</td>
<td>Validated clinical decision rule</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Other data from proficiency testing schemes</td>
<td>Well-designed case-control studies</td>
<td>A single RCT</td>
</tr>
<tr>
<td></td>
<td>Well-designed peer-reviewed studies (e.g., method comparisons, validation studies)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert panel reviewed FDA summaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Less well designed peer-reviewed studies</td>
<td>Lower quality case-control and cross sectional studies</td>
<td>Controlled trial without randomization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unvalidated clinical decision rule</td>
<td>Cohort or case-control study</td>
</tr>
<tr>
<td>4</td>
<td>Unpublished and/or non-peer-reviewed research, clinical laboratory, or manufacturer data</td>
<td>Case series</td>
<td>Case series</td>
</tr>
<tr>
<td></td>
<td>Studies on performance of the same basic methodology, but used to test for a different target</td>
<td>Unpublished and/or non-peer-reviewed research, clinical laboratory, or manufacturer data</td>
<td>Unpublished and/or peer-reviewed studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consensus guidelines</td>
<td>Clinical laboratory or manufacturer data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert opinion</td>
<td>Consensus guidelines</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Expert opinion</td>
</tr>
</tbody>
</table>

Table 4. USPSTF Recommendation Grid: Letter Grade of Recommendation

<table>
<thead>
<tr>
<th>Certainty of Net Benefit</th>
<th>Magnitude of Net Benefit</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Substantial</td>
<td>Moderate</td>
</tr>
<tr>
<td>High</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Moderate</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Grade A** indicates that the certainty of evidence is high that the magnitude of net benefits is substantial.

**Grade B** indicates that the certainty of evidence is moderate that the magnitude of net benefits is either moderate or substantial, or that the certainty of evidence is high that the magnitude of net benefits is moderate.

**Grade C** indicates that the certainty of the evidence is either high or moderate that the magnitude of net benefits is small.

**Grade D** indicates that the certainty of the evidence is high or moderate that the magnitude of net benefits is either zero or negative.

**Grade I** indicates that the evidence is insufficient to determine the relationship between benefits and harms (i.e., net benefit).


Table 5. What the USPSTF Grades Mean and Suggestions for Practice

<table>
<thead>
<tr>
<th>Grade</th>
<th>Grade Definitions</th>
<th>Suggestions for Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>The USPSTF recommends the service. There is high certainty that the net benefit is substantial.</td>
<td>Offer/provide this service.</td>
</tr>
<tr>
<td>B</td>
<td>The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.</td>
<td>Offer/provide this service.</td>
</tr>
<tr>
<td>C</td>
<td>The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is moderate or high certainty that the net benefit is small.</td>
<td>Offer/provide this service only if there are other considerations in support of the offering/providing the service in an individual patient.</td>
</tr>
<tr>
<td>D</td>
<td>The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.</td>
<td>Discourage the use of this service.</td>
</tr>
<tr>
<td>I</td>
<td>Statement</td>
<td>Read “Clinical Considerations” section of USPSTF Recommendation Statement. If offered, patients should understand the uncertainty about the balance of benefits and harms.</td>
</tr>
</tbody>
</table>


In addition to periodic refinement of evidence hierarchies and grading systems used by the USPSTF and EGAPP, other systems are being developed and adopted that have components that are devoted in particular to interventions used in screening and diagnostics. Of particular note are the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system and the Strength of Recommendation Taxonomy (SORT) system. These systems provide explicit differentiation, including separate evidence hierarchies, of evidence requirements for screening and diagnosis, and related guidance about using indirect evidence.
to make inferences about impact of these interventions on patient outcomes. GRADE is being used (sometimes in adapted forms) by an increasing number of organizations, including AHRQ EPCs, the Cochrane Collaboration, and certain medical professional societies and journals. Developed by journal editors and other experts, SORT emphasizes a patient-centered approach to appraising evidence.

**D. Challenge of Relying on RCTs**

The strength of RCTs is in their internal validity—how well the study design and execution diminish the opportunity to introduce bias or inferential error that can affect results regarding the true impact of interventions on outcomes under the particular conditions of the study. In RCTs, randomization ensures that intervention groups (e.g., one getting a new intervention and one getting standard care) differ only in their exposure to the intervention, so that differences in observed impacts can be attributed to differences in the intervention.

Notwithstanding their strengths, there are disadvantages to RCTs in the context of evidence appraisals, particularly for laboratory tests. Given the multiple intervening steps between accurate diagnostic results and improved health outcomes, it can be time-consuming, complex, costly, and, in certain instances, not feasible, to conduct RCTs on laboratory tests that yield direct evidence on patient outcomes. The time between the use of a laboratory test until the time that patients experience a change in health outcomes that may have been influenced by that test can be years or even decades. Blinding of patients (as well as clinicians and investigators) in an RCT, e.g., as to whether patients received one test vs. another, or a test vs. no test, may be impractical. It may be difficult to account for all confounders that may affect patient outcomes in an RCT (e.g., “washing out” previous or current other treatments that may affect outcomes). The impact of a test may be difficult to isolate relative to the impacts of other diagnostic and therapeutic interventions, as well as any relevant environmental and behavioral factors experienced by a patient, between the time of the test and assessment of ultimate patient outcomes. Results of the RCTs, typically conducted with carefully selected patients under carefully controlled conditions, may not be generalizable to the broader patient population in real-world clinical practice. RCTs may not be able to consider or account for patient preferences, which may reflect on their relevance for patient-centered care. To the extent that a substantial body of evidence from other study designs has accumulated, it may be strong enough to obviate the need for RCTs. Further, enrolling sufficient numbers of patients in RCTs and identifying clinical investigators may be impractical under those circumstances.

Methods other than RCTs are better suited to answering certain questions about the clinical impact or value of health care interventions, including screening and diagnostic testing. For example, the best method for assessing prognosis of a condition (disease, disorder) may be a patient cohort study of people with the condition with follow-up at uniform time intervals in the clinical course of the condition. Case control studies can be used to identify risk factors for a condition. Test accuracy (sensitivity, specificity, positive and negative predictive value) can be assessed by a large cross-sectional study of patients known to have a condition. Registries and surveillance studies can be used to monitor patient populations for the incidence of serious or rare adverse effects that may not arise in RCTs that are too small, insufficiently representative of target patient populations, or too short in duration to detect such adverse effects. Although clinical trials that are not randomized and observational studies tend to be less rigorous (for
minimizing sources of bias) than RCTs, a well-designed studies of these types can provide evidence that is sufficiently strong for making clinical and policy decisions.23, 85, 27, 95

As noted above, evidence requirements and grading approaches that are applied to laboratory testing are gradually changing and are being affected by trends in generation of primary data and analyses of observational data. While RCTs will remain the preferred study design for establishing causal effects of interventions on patient outcomes, strengthening of other study designs and emergence of others are helping to supplement and, in some instances, substitute for RCTs. Included are evolving variations in traditional clinical trial designs (e.g., practical clinical trials and adaptive and Bayesian trial designs) and “mining” and other analyses of clinical data sources (e.g., electronic medical records, patient registries) to evaluate populations, interventions, and outcomes. New methods also may involve studies by clinical laboratories that can link de-identified patient samples gathered for testing purposes to patient outcomes. New methods and tools for developing and evaluating evidence are necessary to adequately address the effectiveness of interventions on risks, the changing disease patterns of comorbidities, and heterogeneity of treatment effects based on individuals’ genetic variations.96

E. Challenges Associated with Implementation of Screening in Clinical Practice

1. High Value Screening Tests Underused by Clinicians

There is substantial evidence that clinicians often fail to order appropriate clinical laboratory tests to screen for, diagnose, and monitor patient health conditions, including those tests recommended in clinical guidelines or incorporated into performance measures to evaluate quality of care.97, 98 Underuse is a common problem for several laboratory tests that reduce mortality by detecting disease early or that prompt interventions to control risk. For example, the American Cancer Society recommends that, beginning at age 50, men and women at average risk for developing colorectal cancer receive (depending on patient preference and other factors) one of a set of tests that detect polyps and cancer (flexible sigmoidoscopy, colonoscopy, double-contrast barium enema, or CT colonography) or that mainly find cancer (fecal occult blood test [FOBT], fecal immunochemical test [FIT], or stool DNA test [sDNA]) at designated intervals, including follow-up colonoscopy when certain of these tests are positive.99 Yet, the 2006 National Healthcare Quality Report (NHQR) reported that only 52% of adults over age 50 had colorectal cancer screening by any method in the previous 2 years, indicating that just under half of adults in this population are not following the recommended screening schedule.100 Some tests are overused relative to existing guidelines; one example concerns Pap smears. A study published in 2004 reported that half of all women who had undergone hysterectomy (or nearly 10 million women) received unnecessary Pap smears within the three years leading up to 2002, excluding those Pap smears that may have preceded a recent hysterectomy and hysterectomies that spared the cervix or were performed for cervical neoplasia.101

Clinical laboratory tests also are important tools to control, manage, and monitor chronic conditions. They can detect complications of care and prevent the development of additional comorbidities. Underuse also is prevalent in this regard. For example, elevated blood cholesterol (i.e., LDL and total cholesterol) is an especially important risk factor for heart disease and contributes to the management of diabetes. Significant progress has been made in raising patients’ awareness of the importance of cholesterol screening. However, according to an analysis by the National Center for Health Statistics, only 65% of men and 70% of women 20
years or older have had their cholesterol level checked in the last five years.\textsuperscript{102} According to the NHQR, the lack of adequate screening contributes to overall poor control of high cholesterol in the population.\textsuperscript{103} Of course, measurement alone does not assure desired outcomes; rather, patients must receive needed therapeutic adjustments and counseling from clinicians for effective metabolic control.\textsuperscript{104} Nevertheless, without reliable test results, the clinician cannot make appropriate adjustments or give informed medical advice.

Inadequate use of certain types of testing contributes to the great disparities in health care services in the U.S.\textsuperscript{105} For example, living in a low socioeconomic area has been a key determinant in late-stage diagnosis of colorectal cancer, more so than age, race, gender, and source of care. Poor African-Americans are at increased risk for both occurrence of and mortality from colorectal cancer.\textsuperscript{106, 107} One study found that 50\% of excess mortality observed in African-Americans was due to late-stage diagnosis.\textsuperscript{108} A prominent barrier to adequate screening is lack of health insurance. Providers may not make recommendations for screening or diagnostic tests due to patients’ inability to pay for the test or follow-up treatment or providers may focus “triage” to the patients’ most pressing health need at the moment.\textsuperscript{105} Low-income populations also may be challenged by lack of health literacy, language barriers, negative perceptions about certain diseases, and practical considerations (e.g., job demands).

Multiple reasons are offered regarding why clinicians are not adequately promoting the benefits of well-established screening and preventive services consistent with clinical practice guidelines. Box 11 lists some of the main reasons cited in the literature.

<table>
<thead>
<tr>
<th>Box 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provider Barriers to Preventive Care</strong></td>
</tr>
<tr>
<td>- Insufficient time during the clinical encounter</td>
</tr>
<tr>
<td>- Organizational barriers (e.g., offering only those preventive services associated with performance measures, administrative costs)</td>
</tr>
<tr>
<td>- Provider attitude toward patients (e.g., anticipated lack of patient cooperation with recommended test, inattentiveness to language barriers)</td>
</tr>
<tr>
<td>- Provider beliefs and lack of knowledge concerning the evidence supporting the guidelines (e.g., that evidence is inconclusive, insufficient, or conflicting; high number of false negatives)</td>
</tr>
<tr>
<td>- Lack of financial incentives encouraging prevention (e.g., capitation, low payment, lack of short-term savings).\textsuperscript{21, 105, 109-112}</td>
</tr>
</tbody>
</table>

**2. Insufficient Patient Knowledge of High Value Tests**

Patients often expect to receive a laboratory test during physician office visits.\textsuperscript{113} They may even see such tests as evidence of sound medical treatment or acknowledgement that their concerns have been heard. Generally, patient expectations are tied to routine screening tests, including blood counts, urine tests, blood chemistries, Pap smears for women, and prostate specific antigen (PSA) tests for men.\textsuperscript{114, 115} Patient satisfaction does not seem to be linked to whether expectations were met or not for tests or specialist referrals, although less satisfaction has been reported when new medications are expected but not received.\textsuperscript{116}
There are reported marked differences in patient knowledge and requests of certain widely used screening and diagnostic tests, such as FOBT or fecal immunochemical tests (FIT) for colorectal cancer, C-reactive protein for heart disease, and serologic tests for sexually transmitted diseases. For example, a study published in 2005 assessed barriers to screening (via FOBT, FIT, or colorectal endoscopy) in adults 50 years or older who had never been tested or were not current for recommended screening for colon cancer. Patient-related lack of knowledge, awareness, and motivation was cited as the main barrier for about 77% of these individuals and lack of access or provider recommendation were cited as the main barrier for about 22%. Only 10% of adults not current with testing and who had a doctor visit in the past year reported receiving a screening recommendation.\textsuperscript{117} Even when patients are aware of screening recommendations, they may not follow through due to other barriers, such as lack of health insurance. Underserved populations tend to be most vulnerable in this regard.\textsuperscript{118}

Many providers believe that meeting patient expectations, when medically warranted, is important to satisfaction with care and contributes to health outcomes by improving compliance with treatment and follow-up.\textsuperscript{119} However, the actual effect of patient expectations on satisfaction with care is mixed.\textsuperscript{113} In some instances, patients may want information rather than specific actions (e.g., test ordering). Patient expectation for testing may vary by the prospects of unfavorable test results or out-of-pocket costs associated with diagnostic tests.\textsuperscript{120} Pursuant to patient-centered care, providers must manage patient demands and expectations for laboratory tests and counsel them as to those that would best serve their health needs.\textsuperscript{121}

### 3. Lag Times and Varying Evidence Expectations

Two factors that introduce uncertainty and inconsistency to incorporation of evidence into practice are lag times and varying evidence expectations among organizations that generate clinical practice recommendations. Lag time from appearance of relevant new evidence pertaining to a laboratory test through updating recommendations about use of the test remains a significant challenge to clinicians, patients, payers, and other stakeholders. This applies to any organizations engaged in appraising evidence, including the USPSTF, EGAPP, health professional expert panels, HTA vendors, and others. To the extent that such recommendations are used to inform practice guidelines and coverage policies, this time lag can slow access to beneficial laboratory tests and other interventions. Given its important role in generating evidence-based recommendations pertaining to laboratory testing and other services used in prevention, it is helpful to consider the USPSTF with regard to these factors.

The USPSTF tends to have among the more rigorous evidence requirements of the expert advisory groups that generate recommendations or guidelines for clinical preventive services. Also, the USPSTF is among the organizations that places explicit emphasis on direct evidence of the impact of preventive interventions on patient outcomes, not just intermediate or surrogate outcomes. As such, this can result in “I” statements (insufficient evidence to recommend for or against using a service) rather than “A” or “B” recommendations (to offer or provide a service) in some instances where other organizations recommend offering or providing the service.

Like other public and private sector efforts that generate evidence-based recommendations for clinical practice, the USPSTF is subject to delays from the publication of new evidence on a topic to the time at which this evidence can be incorporated into a new or revised recommendation. Given resource constraints and the backlog of topics, there are time lags between updates of any
The Value of Laboratory Screening and Diagnostic Tests for Prevention and Health Care Improvement

topic. As such, relevant evidence may come into the public domain and be incorporated by some expert advisory groups’ processes before others.

The USPSTF 2008 update of its 2003 recommendation about screening for prediabetes and diabetes provides a useful example about varying evidence requirements and time lags. In the 2008 update, and drawing on clinical trial evidence generated since the 2003 recommendation, the USPSTF narrowed its grade B recommendation for screening for type 2 diabetes in asymptomatic adults by retaining screening for those with sustained high blood pressure but dropping the recommendation to screen those with hyperlipidemia. Even so, the USPSTF stated that clinicians should assess overall cardiovascular disease risk in patients, and that if the patient’s risk is near a threshold for treatment with lipid-lowering drugs, they should screen for diabetes to assess the patient’s cardiovascular disease risk. The USPSTF found that there is still insufficient evidence to establish the impact of routine screening for type 2 diabetes on patient outcomes for asymptomatic individuals with normal blood pressure. In 2009, the American Academy of Family Physicians modified its recommendation in a manner that aligned with the USPSTF. In contrast, in its 2009 statement, the American Diabetes Association recommends screening for diabetes and prediabetes beginning at age 45. The Canadian Task Force on Preventive Health Services guideline, last updated in 2005, recommends screening for patients with hypertension or hyperlipidemia. (See accompanying case study on screening for type 2 diabetes and prediabetes.)

In its most recent review of the topic in 2003, the USPSTF concluded that the evidence was insufficient to (1) recommend for or against the routine use of new technologies (i.e., HPV testing, liquid-based cytology, computerized rescreening, and algorithm based screening) as a primary screening tool, or (2) determine whether new technologies are more effective than conventional Pap smear screening, in reducing incidence of or mortality from invasive cervical cancer. Currently, the USPSTF is revisiting the topic of cervical cancer screening and, as such, is expected to consider new evidence from several RCTs published since 2002 confirming the value of adding HPV testing to cervical cytology. (See accompanying case study on HPV screening.)

In screening for hepatitis C virus (HCV), the CDC recommends screening of a broad range of high-risk individuals. In contrast, in its 2004 statement, the USPSTF found insufficient evidence to recommend for or against routine screening for HCV infection in adults at high risk for infection. In particular, USPSTF found no evidence that screening for HCV infection in high-risk adults leads to improved long-term health outcomes, although there is good evidence that antiviral therapy improves intermediate outcomes, such as viral loads of HCV.

IV. Effect of Value Assessments on Reimbursement Policy

A considerable set of challenges to the laboratory medicine sector involves payer reimbursement policies that govern coverage decisions, payment rates, and coding of new tests. Public and private sector insurers use independent processes to conduct HTAs and other evidence appraisals that inform their respective reimbursement decisions. The policies of the Centers for Medicare and Medicaid (CMS), the nation’s largest health care purchaser, have an especially strong influence on the reimbursement policies of other federal, state, and private sector payers, particularly with regard to payment rates.
Specific challenges involving payers include the following.

- Coverage of certain laboratory tests based on medical necessity criteria is variable across payer groups, particularly for certain screening tests and for genetic and other newly developed molecular tests.

- Among public and private sector insurers, the multitude of different payment schedules for inpatient and ambulatory care services used through local payer entities can be burdensome to manage and difficult to assess for periodic fee increases, and can result in payment shortfalls and inconsistencies across carriers.

- The standardized coding systems used to list a test on a fee schedule and process claims, as well as the mechanism for updating codes, are inadequate, which leaves providers to use existing codes and underdeveloped code modifiers that lack specificity for newer and emerging tests.

Many of these challenges have been brought to the attention of policy makers in other reports; however, progress in resolving them remains slow.

A. Coverage Decisions

Medicare Coverage. Medicare’s authorizing legislation in 1965 established broad categories of coverage for hospital, physician, and laboratory services, but limited payment to expenses deemed reasonable and necessary for the “diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member” (emphasis added). This means that the original Medicare statute effectively restricts payment for preventive and screening interventions in patients without signs, symptoms, complaints or personal history of disease or injury, unless otherwise specified by Congress.

In general, policymakers understand the importance of screening and preventive services in improving patient health and reducing expenditures. In 2003 and 2005, Congress expanded Medicare coverage to include certain screening tests, such as those for cardiovascular disease and diabetes. More recently, under the Medicare Improvement for Patients and Providers Act of 2008, CMS gained authority to consider adding preventive services that “that identify medical conditions or risk factors” that the Secretary of Health and Human Services deems “reasonable and necessary for the prevention or early detection of an illness or disability,” and those recommended by the USPSTF. As such, the advantages and disadvantages of the USPSTF’s approach to developing recommendations for clinical preventive services may be reflected in Medicare coverage policies.

As is so for other payers, Medicare coverage decisions are subject to time lags, reflecting processes to review and revise coverage decisions and the time lag of peer-review publications. Such time lags may delay coverage determinations concerning some innovative technologies and testing methods. Coverage decisions are also affected by the quality of relevant available evidence and some requests for further evidence. The CMS Medicare Evidence Development and Coverage Advisory Committee (MEDCAC) examines available evidence, including in the form of HTAs prepared by EPCs at the request of CMS, in advising CMS regarding pending national coverage determinations (NCDs) for Medicare. For 85% of the various technologies evaluated from 1999 to 2007, including some diagnostic tests and procedures, CMS considered the evidence fair or poor.
Positive NCDs were made in 60% of cases, although almost always with conditions placed on the population or setting to which coverage applies. The conditions for coverage vary, although they increasingly involve restrictions tied to disease severity and treatment regimens as well as requirements for clinical trials, registries, or other data collection.130

Consistent with the Medicare statute, most genetic tests are not covered by Medicare unless they are performed on symptomatic individuals or are used to identify treatment-responsive populations.8,131 In 2009, CMS131 is reviewing scientific evidence on genetic tests and assess their value for coverage decisions, including innovative oncology tests and pharmacogenomic tests.

**Private Payer Coverage.** Private payers maintain their own processes for making coverage decisions and may choose to adjust policies following the introduction of new technologies to the market. In addition to matters of clinical benefit for their beneficiaries, these policies may be affected by a variety of sources, such as state mandates, consumer preference, or financial considerations. Payers often negotiate specific coverage policies with the groups or employers purchasing the health plan.133 Private payers increasingly draw on HTAs prepared by HTA vendors and other HTA programs and agencies to inform their coverage decisions. Generally, there is consistency in coverage among private sector payers in routine laboratory screening and diagnostic testing associated with standard of care, although some variations remain.133

Due in part to how recently these tests have become available and their innovative nature, private payers vary widely in their coverage of pharmacogenomic and other genetic and molecular-based tests, making their own decisions on a case-by-case basis. Consistent with higher and more explicit evidence requirements, payers seek evidence on clinical utility as well as analytical validity and clinical validity for these tests. Private plans generally cover a pharmacogenomic test when it is recommended on a drug label, such as the HER-2/neu test to determine whether Herceptin treatment should be recommended for patients with breast cancer.134 Coverage decisions may vary significantly across individual payers in instances where there evidence is limited on clinical utility and costs associated with a test.14,133-135

**B. Payment Systems**

Methods of paying for laboratory tests are complex and vary by provider site (e.g., inpatient, outpatient, ambulatory care) and type of test (e.g., clinical or anatomic pathology).8 In most instances, the Medicare prospective payment systems for inpatient and ambulatory care serve as the basis for federal and private sector payment, although other payers may assign their own payment rates. The ambulatory care payment schemes have proved to be the most challenging for the diagnostics industry.

---

1 Aside from a few CMS Medicaid-mandated newborn genetic screening tests, coverage decisions are the responsibility of the state Medicaid programs and can vary substantially from state to state.132

1 For inpatient care, technical fees for laboratory services are bundled into diagnosis-related groups (DRGs) based on the patient’s diagnosis; ambulatory patient classifications (APC) are used for payment of hospital outpatient surgery.
Laboratory tests for ambulatory care are paid according to predetermined, fixed-fee schedules, negotiated contracts, or competitive bidding contracts. The main CMS fee schedules are the:

- Medicare physician fee schedule (MPFS), which covers clinician services, including pathologist interpretive services for certain anatomic pathology or molecular tests
- Medicare Clinical Laboratory Fee Schedule (CLFS) for payment of clinical laboratory tests

Private payers often use the MPFS and CLFS for setting their own payment rates as a multiple or percentage of Medicare rates, although private payers typically negotiate contracts with providers and suppliers, including clinical laboratories.

Policies related to the CLFS have proven to be the greatest challenge to laboratories for several reasons. The CLFS was established in 1984 and is based on prevailing charges from 1983. The CLFS contains payment ceilings, or national limitation amounts (NLA). While Congress originally intended to adjust the CLFS annually for inflation, this has rarely occurred. Over the past 25 years, the NLAs have been reduced seven times; also, there were only two pricing updates, 1.1% (2003) and 4.5% (2009), during the period 1997-2009. Because the CLFS historically has been managed at the local level by 56 local carriers, there have been essentially 56 different fee schedules until recently. Now that CMS has consolidated these carriers into 15 Medicare Administrative Contractor (MAC) jurisdictions, some consolidation of these fee schedules is occurring. According to a recent report by the Office of the Inspector General (OIG) at HHS, 83% of carrier rates were at the NLAs and 89% of claims are paid at the NLAs. However, 97% of lab tests had at least one carrier rate that varied from the NLA. For some of the most frequently used tests, OIG found wide variations in payment levels, including variations of up to 40% for complete blood count, 30% for a urinalysis, 28% for glycosylated hemoglobin, 27% for lipid panel, 19% for prothrombin time, and 11% for basic metabolic panel. As noted in our earlier report and elsewhere, the CMS processes for developing CLFS payment rates for new test HCPCS codes pose significant uncertainties and other challenges for laboratories and manufacturers that hinder incentives to innovate in this area.

V. Key Findings and Conclusions

Overarching conclusion:

Innovation, demonstrated clinical benefit, and appropriate use of laboratory screening and diagnostic tests are essential to achieving goals of health system reform. Clinical laboratory testing is integral to evidence-based improvements in health care quality, patient outcomes, efficiency, and accountability.

1. Screening and diagnostic tests contribute to health care value across the spectrum of care. Information from clinically appropriate testing contributes to early detection, diagnosis, patient and clinician decision-making, choice of treatment, therapeutic monitoring, reducing adverse events, improved health outcomes and quality of life, and more cost-effective care.

Refer to the CDC-sponsored report, Laboratory Medicine: A National Status Report, for a full review of reimbursement policy and payment calculation methodology at [http://www.futurelabmedicine.org](http://www.futurelabmedicine.org)
2. Laboratory testing, including existing, new, and emerging testing technologies, aligns with and will have an integral role in meeting major goals of national health reform. These aspects include wider access to testing; early detection and treatment; support of personalized medicine and other tailored services for priority populations and others; more appropriate, efficient and cost-effective care for chronic conditions; slower disease progression; and faster recovery and less disability.

3. A growing body of evidence demonstrates the value of laboratory screening tests in primary prevention—that is, in early detection at a time when it is possible to prevent the onset of disease, not just treat disease after it has occurred. Primary prevention offers opportunities for cost-effective care and net savings in some instances. Advances in genetic and molecular testing are enhancing the potential of primary prevention through identification of individuals with pre-dispositions for disease.

4. Key HTA and clinical practice guideline groups are setting high evidence bars, emphasizing explicitly the need for demonstrating the clinical utility of tests used in screening and diagnosis. In addition to analytic validity (test accuracy, precision, robustness), which is typically established by regulatory processes, and clinical validity (detect and predict probability of having a disorder based on a test result), there is greater demand for evidence of clinical utility (impact on clinical outcomes and usefulness to patient and clinician decision-making) by the USPSTF, EGAPP, various HTA agencies, private payers, and others.

   - RCTs that provide direct evidence of impact of testing on patient outcomes remain preferred to other study designs and indirect evidence of such impact. This remains a challenge to, and can be impractical for, demonstrating value of testing, particularly given the confounding effects of intervening decisions, interventions, and environmental factors between testing and ultimate outcomes, and the costs and time (which can be years or decades) needed to assess impact on outcomes using RCTs.

   - Evidence hierarchies and grading systems, such as recently revised ones by the USPSTF and those of EGAPP and GRADE, should continue to be adapted for screening and diagnostic testing, accounting for emerging, practical study designs for validating tests and accompanying evidence appraisal methods.

   - The explicitness of evidence expectations by regulators, guideline groups, and other gatekeepers provides opportunities for test manufacturers and clinical laboratories to build evidence requirements into innovation, development, validation, and marketing of laboratory tests.

5. Varying evidence standards and time lags for assessing evidence and updating recommendations and guidelines introduce inconsistency and uncertainty to incorporating evidence about some laboratory tests into practice. Lag time from appearance of relevant new evidence through updating and disseminating recommendations can affect health care quality and access. Different evidence requirements, including the generally more rigorous ones of USPSTF compared to others, along with varying review cycles of these groups, can result in divergent findings in such areas as screening for diabetes, human papillomavirus, and hepatitis C virus. These factors also pose risk and uncertainty to laboratories and test manufacturers.
6. **New methods and analytical tools are emerging for assessing and demonstrating the clinical and economic impact of laboratory tests.** While RCTs remain the preferred study design for establishing causal effects of interventions on patient outcomes, strengthening and emergence of other study designs are helping to supplement and, in some instances, may substitute for traditional RCTs. Included are variations in traditional clinical trial designs; “data mining” of claims data, patient registries, and electronic medical records; retrospective studies of specimen remnants; and analyses of linked data sets of laboratory data and patient outcomes.

7. **Laboratory testing has prominent roles in the national agenda for comparative effectiveness research that provide opportunities for broad demonstration of value in “real-world” health care.** Two main roles for laboratory testing in CER are: (1) for patient selection and tracking intermediate and long-term health outcomes in CER of other interventions, and (2) as the index interventions for CER, e.g., in head-to-head comparisons of alternative laboratory tests or comparisons of laboratory tests to other tests for particular health care conditions. Included among the Institute of Medicine’s recommended top national CER priorities are screening for methicillin resistant *S. aureus* (MRSA) and genetic and biomarker testing for multiple types of cancer.

8. **Greater investment is needed in the small but growing body of evidence on the economic impact of laboratory testing.** As is so for many other types of health care technology, wider demand by clinicians, provider institutions, payers and policy makers, including in national deliberations on health reform, calls for evidence pertaining to the distinct analyses of cost-savings, cost-effectiveness, and cost-utility of laboratory testing. Findings about laboratory tests that are truly cost-saving or highly cost-effective, such as colorectal cancer screening in adults aged 50-75 and screening young women for chlamydial infection, should be broadly recognized and used as models for further work. Key considerations are that economic impact of specific tests can vary significantly depending on the population targeted (with testing of high-risk populations more likely to be cost-effective) and tradeoffs between the cost of greater testing frequency and yield of cases detected.

9. **Laboratory testing has a central role in personalized medicine, whose extraordinary potential is recently emerging into practice.** Recent scientific and technological advances have led to molecular-level and genetic testing, including pharmacogenomics, that enable tailoring therapies to subgroups and individuals, i.e., to ensure ‘the right treatment for the right patient at right time.’ In parallel, EGAPP and other groups are developing evidence frameworks to guide assessments of such genetic and genomic testing technologies.

10. **Payment policies that govern coverage decisions, payment rates, and coding of new tests remain major challenges to the laboratory medicine sector.** Coverage of certain laboratory tests, particularly certain screening tests and for genetic and other newly developed molecular tests, is variable across payers. The multitude of payment schedules for inpatient and ambulatory care services used through local payer entities can be burdensome to manage and difficult to assess for periodic fee increases, and can result in payment shortfalls. The coding systems used to list tests on fee schedules and process claims, and the mechanisms to update them, are inadequate, often leaving providers to use existing codes and underdeveloped code modifiers that are not specific to newer and
emerging tests. Medicare’s processes for developing CLFS payment rates for new test HCPCS codes pose significant uncertainties and other challenges for laboratories and manufacturers that hinder the incentive to innovate in this area. Progress in resolving these longstanding problems remains slow.
References


The Value of Laboratory Screening and Diagnostic Tests for Prevention and Health Care Improvement


Appendix
Case Studies