Teaching Pack: Using Test Information I


KEY CHAPTERS IN DECISION ANALYSIS TEXTBOOKS


**SCREENING AND DIAGNOSIS IN CLINICAL DECISION MAKING**


Getting the right diagnosis is a key aspect of health care - it provides an explanation of a patient's health problem and informs subsequent health care decisions. The diagnostic process is a complex, collaborative activity that involves clinical reasoning and information gathering to determine a patient's health problem. According to Improving Diagnosis in Health Care, diagnostic errors—incorrect or delayed diagnoses—persist throughout all settings of care and continue to harm an unacceptable number of patients. It is likely that most people will experience at least one diagnostic error in their lifetime, sometimes with devastating consequences. Diagnostic errors may cause harm to patients by preventing or delaying appropriate treatment, providing unnecessary or harmful treatment, or resulting in psychological or financial repercussions. The committee concluded that improving the diagnostic process is not only possible, but also represents a moral, professional, and public health imperative.


The clinical utility of medical tests is measured by whether the information they provide affects patient-relevant outcomes. To a large extent, effects of medical tests are indirect in nature. In principle, a test result affects patient outcomes mainly by influencing treatment choices. This indirectness in the link between testing and its downstream effects poses practical challenges to comparing alternate test-and-treat strategies in clinical trials. Keeping in mind the broader audience of researchers who perform comparative effectiveness reviews and technology assessments, the authors summarize the rationale for and pitfalls of decision modeling in the comparative evaluation of medical tests by virtue of specific examples. Modeling facilitates the interpretation of test performance measures by connecting the link between testing and patient outcomes, accounting for uncertainties and explicating assumptions, and allowing the systematic study of tradeoffs and uncertainty. The authors discuss challenges encountered when modeling test-and-treat strategies, including but not limited to scarcity of data on important parameters, transferring estimates of test performance across studies, choosing modeling outcomes, and obtaining summary estimates for test performance data.


The physician's estimate of the probability that a patient has a particular disease is a principal factor in the determination of whether to withhold treatment, obtain more data by testing, or treat without subjecting the patient to the risks of further diagnostic tests. Using the concepts of decision analysis, we have derived expressions for two threshold probabilities involved in this choice: a “testing” threshold and a “test-treatment” threshold. Values can be assigned to these thresholds from data on the reliability and potential risks of the diagnostic test and the benefits and risks of a specific treatment. Treatment should be withheld if the probability of disease is smaller than the testing threshold, and treatment should be given without further testing if the probability of disease is greater than the test-treatment threshold. The test should be performed (with treatment depending on the test outcome) only if the probability of disease is between the

---

This teaching pack was developed by the Center for Health Decision Science at the Harvard T.H. Chan School of Public Health. All materials produced by the Center for Health Decision Science are free and publicly accessible for educational use.

This resource is licensed Creative Commons Attribution-Non Commercial-NoDerivs3.0 Unported

chds.hsp.harvard.edu
two thresholds. The method exposes important principles of decision making and helps the clinician develop a rational, quantitative approach to the use of diagnostic tests.

**Realdi G et al.** *Selection of Diagnostic Tests for Clinical Decision Making and Translation to a Problem Oriented Medical Record.* Clinica Chimica Acta 2008; 393 (1): 37-43. [https://doi.org/10.1016/j.cca.2008.03.024](https://doi.org/10.1016/j.cca.2008.03.024). The leading function of the physician is the clinical reasoning, which involves appropriate investigation of the problems of the patient, formulation of a diagnostic suspect based on the patient's symptoms and signs, gathering of additional relevant information, to select necessary tests and administration of the most suitable therapy. The problems of the patient are expressed by symptoms or signs or abnormal test results, requested for a variety of reasons. The entire scientific, as well as diagnostic approach, is based on three steps: to stumble in a problem; to try a solution through a hypothesis; to disprove or to prove the hypothesis by a process of criticism. Clinicians use the information obtained from the history and physical examination to estimate initial (or pre-test) probability and then use the results from tests and other diagnostic procedures to modify this probability until the post-test probability is such that the suspected diagnosis is either confirmed or ruled out. When the pre-test probability of disease is high, tests characterized by high specificity will be preferred, in order to confirm the diagnostic suspect. When the pre-test probability of disease is low, a test with high sensitivity is advisable to exclude the hypothetical disease. The above mentioned process of decision making has been transferred to a problem oriented medical record that is currently employed in our Clinic.

**Grimes DA, Schulz KF.** *Refining Clinical Diagnosis with Likelihood Ratios.* The Lancet 2005; 365: 1500-1505. [Open Access:](https://doi.org/10.1016/S0140-6736(05)66422-7) Likelihood ratios can refine clinical diagnosis on the basis of signs and symptoms; however, they are underused for patients' care. A likelihood ratio is the percentage of ill people with a given test result divided by the percentage of well individuals with the same result. Ideally, abnormal test results should be much more typical in ill individuals than in those who are well (high likelihood ratio) and normal test results should be most frequent in well people than in sick people (low likelihood ratio). Likelihood ratios near unity have little effect on decision-making; by contrast, high or low ratios can greatly shift the clinician's estimate of the probability of disease. Likelihood ratios can be calculated not only for dichotomous (positive or negative) tests but also for tests with multiple levels of results. When combined with an accurate clinical diagnosis, likelihood ratios from ancillary tests improve diagnostic accuracy in a synergistic manner.

**Šimundić A.** *Diagnostic Accuracy – Part 1: Basic Concepts: Sensitivity and Specificity, ROC Analysis, STARD Statement.* Point of Care 2012; 11 (1): 6-8. [Open Access:](http://journals.lww.com/poctjournal/Citation/2012/03000/Diagnostic_Accuracy_Part_1__Basic_Concepts___.3.aspx). Diagnostic accuracy relates to the ability of a test to discriminate between the target condition and health. This discriminative potential can be quantified by the measures of diagnostic accuracy such as sensitivity and specificity, predictive values, likelihood ratios, the area under the ROC curve, Youden's index and diagnostic odds ratio. Different measures of diagnostic accuracy relate to the different aspects of diagnostic procedure: while some measures are used to assess the discriminative property of the test, others are used to assess its
predictive ability. Measures of diagnostic accuracy are not fixed indicators of a test performance, some are very sensitive to the disease prevalence, while others to the spectrum and definition of the disease. Furthermore, measures of diagnostic accuracy are extremely sensitive to the design of the study. Studies not meeting strict methodological standards usually over- or under-estimate the indicators of test performance as well as they limit the applicability of the results of the study. STARD initiative was a very important step toward the improvement the quality of reporting of studies of diagnostic accuracy. STARD statement should be included into the Instructions to authors by scientific journals and authors should be encouraged to use the checklist whenever reporting their studies on diagnostic accuracy. Such efforts could make a substantial difference in the quality of reporting of studies of diagnostic accuracy and serve to provide the best possible evidence to the best for the patient care. This brief review outlines some basic definitions and characteristics of the measures of diagnostic accuracy.

Bainbridge T. Confidence in Medical Tests. Practice Uses of Math and Science 2016. Open Access: https://pumas.nasa.gov/examples/index.php?id=162. Medical tests are not always correct. When a doctor and a patient receive the results of a test, what is the probability that the result is correct?

Grimes DA, Schulz KF. Uses and Abuses of Screening Tests. The Lancet 2002; 359: 881–884. http://dx.doi.org/10.1016/S0140-6736(02)07948-5. Screening tests are ubiquitous in contemporary practice, yet the principles of screening are widely misunderstood. Screening is the testing of apparently well people to find those at increased risk of having a disease or disorder. Although an earlier diagnosis generally has intuitive appeal, earlier might not always be better, or worth the cost. Four terms describe the validity of a screening test: sensitivity, specificity, and predictive value of positive and negative results. For tests with continuous variables—eg, blood glucose-sensitivity and specificity are inversely related; where the cutoff for abnormal is placed should indicate the clinical effect of wrong results. The prevalence of disease in a population affects screening test performance: in low-prevalence settings, even very good tests have poor predictive value positives. Hence, knowledge of the approximate prevalence of disease is a prerequisite to interpreting screening test results. Tests are often done in sequence, as is true for syphilis and HIV-1 infection. Lead-time and length biases distort the apparent value of screening programmes; randomised controlled trials are the only way to avoid these biases. Screening can improve health; strong indirect evidence links cervical cytology programmes to declines in cervical cancer mortality. However, inappropriate application or interpretation of screening tests can rob people of their perceived health, initiate harmful diagnostic testing, and squander health-care resources.

TOOLS AND NOMOGRAMS


In 1975, Fagan published a nomogram to help practitioners determine, without the use of a calculator or computer, the probability of a patient truly having a condition of interest given a particular test result. Nomograms are very useful for bedside interpretations of test results, as no test is perfect. However, the
practicality of Fagan's nomogram is limited by its use of the likelihood ratio (LR), a parameter not commonly reported in the evaluation studies of diagnostic tests. The LR reflects the direction and strength of evidence provided by a test result and can be computed from the conventional diagnostic sensitivity and specificity of the test. This initial computation is absent in Fagan's nomogram, making it impractical for routine use. The authors have integrated the initial step to compute the LR and the resulting two-step nomogram allows the user to quickly interpret the outcome of a test. With the addition of the sensitivity and specificity, the nomogram, for the purposes of interpreting a dichotomous test result, is now complete. This tool is more accessible and flexible than the original, which will facilitate its use in routine evidence-based practice.

DocNomo is a convenient graphical tool to enhance the bedside interpretation of a diagnostic test result. From the diagnostic sensitivity and specificity of the test and the probability of the patient having the target disorder before running the test (Pre-Test Probability), DocNomo calculates for you the probability of the patient having the target disorder after running the test (i.e. Post-Test Probability). Using the convenience of tactile sliders, each of these parameters may be changed to visually assess their impact on the Post-Test Probability. DocNomo is the digital adaptation of the Two-Step Fagan Nomogram which is the updated version (2013) of the original Fagan's nomogram developed by Dr. Terrence J. Fagan (1975).

In this letter to the editor, Fagan introduces a solution to Bayes’s formula in the form of a nomogram.

In this commentary, Abushouk reflects on the evolution of Bayes’ related nomograms from Fagan’s in 1975 through a 2011 reimagining in use today.

ASSESSING TEST PERFORMANCE

 Diagnostic test evidence is challenging to grade because standard tools for grading evidence were designed for questions about treatment rather than diagnostic testing; and the clinical usefulness of a diagnostic test depends on multiple links in a chain of evidence connecting the performance of a test to changes in clinical outcomes. Grading the strength of a body of diagnostic test evidence involves challenges over and above those related to grading the evidence from health care intervention studies. This chapter identifies challenges and outlines principles for grading the body of evidence related to diagnostic test performance.

There may be genuine differences between test accuracies in different settings, such as primary care or hospital, in different types of hospital, or between countries. Deciding whether estimates of test accuracy are transferable to other settings depends on an understanding of the possible reasons for variability in test discrimination and calibration across settings. The transferability of measures of test performance from one setting to another depends on which indicator of test performance is to be used. Real variation in the performance of diagnostic tests (such as different test types, or a different spectrum of disease) needs to be distinguished from artifactual variation resulting from study design features. These features include the target condition and reference standard used, the population and the clinical question studied, the evaluated comparison, and the way the index test was performed, calibrated, and interpreted. In preparing studies on diagnostic accuracy, a key question is how to design studies that carry more information about the transferability of results. To ensure that estimates of diagnostic accuracy will travel, before starting to design a study the following questions must be answered: How are the target condition and reference standard defined? - Is the objective to estimate global test performance or to estimate probability of disease in individuals? - What is the population and clinical problem? - Is the test being considered as a replacement or incremental test? - To what extent do you want to study the reasons for variability of the results within your population? - To what extent do you want to study the transferability of the results to other settings? Designing studies with heterogeneous study populations allows exploration of the transferability of diagnostic performance in different settings. This will require larger studies than have generally been carried out in the past for diagnostic tests.


Effectiveness and comparative effectiveness reviews, systematic reviews of existing research on the effectiveness, comparative effectiveness, and comparative harms of different medical tests, are intended to provide relevant evidence to inform real-world health care decisions for patients, providers, and policymakers. In an effort to improve the transparency, consistency, and scientific rigor of the work of the Effective Health Care (EHC) Program, the Evidence-based Practice Center Program (EPC) has developed a Methods Guide for Medical Test Reviews through a collaborative effort. This guide is intended to be a practical resource for investigators in the Evidence-based Practice Center program as well as other investigators interested in preparing, conducting, and using systematic reviews of medical tests. Originally posted as a draft for public comment in November 2010, the papers have been revised in response to peer and public review comments. The revised versions are posted below and co-published with the Journal of General Internal Medicine on June 1, 2012 at: http://link.springer.com/journal/11606/27/1/suppl/page/1.

- Samson D, Schoelles KM. Chapter 2: Medical Tests Guidance (2) Developing the Topic and Structuring Systematic Reviews of Medical Tests: Utility of PICOTS, Analytic Frameworks, Decision Trees, and


RECEIVER OPERATING CHARACTERISTIC CURVES

Sensitivity and specificity are the basic measures of accuracy of a diagnostic test; however, they depend on the cut point used to define "positive" and "negative" test results. As the cut point shifts, sensitivity and specificity shift. The receiver operating characteristic (ROC) curve is a plot of the sensitivity of a test versus its false-positive rate for all possible cut points. The advantages of the ROC curve as a means of defining the accuracy of a test, construction of the ROC, and identification of the optimal cut point on the ROC curve are discussed. Several summary measures of the accuracy of a test, including the commonly used percentage of correct diagnoses and area under the ROC curve, are described and compared. Two examples of ROC curve application in radiologic research are presented.

A representation and interpretation of the area under a receiver operating characteristic (ROC) curve obtained by the "rating" method, or by mathematical predictions based on patient characteristics, is presented. It is shown that in such a setting the area represents the probability that a randomly chosen diseased subject is (correctly) rated or ranked with greater suspicion than a randomly chosen non-diseased subject. Moreover, this probability of a correct ranking is the same quantity that is estimated by the already well-studied nonparametric Wilcoxon statistic. These two relationships are exploited to (a) provide rapid closed-form expressions for the approximate magnitude of the sampling variability, i.e., standard error that one uses to accompany the area under a smoothed ROC curve, (b) guide in determining the size of the sample required to provide a sufficiently reliable estimate of this area, and (c) determine how large sample sizes should be to ensure that one can statistically detect differences in the accuracy of diagnostic techniques.

APPLIED EXAMPLES BY CLINICAL AREA
Cardiovascular Disease and Stroke Diagnostics
The authors evaluated the diagnostic accuracy of exercise-induced ST-segment depression in detecting coronary-artery disease by applying the likelihood-ratio formulation of Bayes's theorem to stress test data, which were partitioned into half-millimeter ranges of depression. The graphic relation between the predictive value of a given test result and the pretest risk of disease in the test subjects was obtained for each of these half-millimeter intervals. This method reveals that the predictive value of testing depends on the degree of ST-segment depression, and that the pretest risk of coronary-artery disease is an important determinant of the predictive value of any test result in the individual patient. These findings suggest that the use of the terms "positive" and "negative" are inappropriate to describe most stress-test results – rather results should be interpreted in terms of a continuum of risk based on the ST-segment depression.

TB Diagnostics

The landscape of diagnostic testing for tuberculosis (TB) is changing rapidly, and stakeholders need urgent guidance on how to develop, deploy and optimize TB diagnostics in a way that maximizes impact and makes best use of available resources. When decisions must be made with only incomplete or preliminary data available, modelling is a useful tool for providing such guidance. Following a meeting of modelers and other key stakeholders organized by the TB Modelling and Analysis Consortium, we propose a conceptual framework for positioning models of TB diagnostics. The authors use that framework to describe modelling priorities in four key areas: Xpert® MTB/RIF scale-up, target product profiles for novel assays, drug susceptibility testing to support new drug regimens, and the improvement of future TB diagnostic models. In order to maximize the impact and cost-effectiveness of TB diagnostics, these modelling priorities should figure prominently as targets for future research.

Also see:

The Xpert MTB/RIF test enables rapid detection of tuberculosis (TB) and rifampicin resistance. The World Health Organization recommends Xpert for initial diagnosis in individuals suspected of having multidrug-resistant TB (MDRTB) or HIV-associated TB, and many countries are moving quickly toward adopting Xpert. The authors evaluated potential health and economic consequences of implementing Xpert in five southern African countries—Botswana, Lesotho, Namibia, South Africa, and Swaziland—where drug resistance and TB-HIV coinfection are prevalent.

Cancer Screening

Experts, professional societies, and consumer groups often recommend different strategies for cancer screening. These strategies vary in the intensity of their search for asymptomatic lesions and in their value. This article outlines a framework for thinking about the value of varying intensities of cancer screening. The authors conclude that increasing intensity beyond an optimal level leads to low-value screening and speculate about pressures that encourage overly intensive, low-value screening.


Several factors are changing the landscape of cervical cancer control, including a better understanding of the natural history of human papillomavirus (HPV), reliable assays for detecting high-risk HPV infections, and
a soon to be available HPV-16/18 vaccine. There are important differences in the relevant policy questions for different settings. By synthesizing and integrating the best available data, the use of modeling in a decision analytic framework can identify those factors most likely to influence outcomes, can guide the design of future clinical studies and operational research, can provide insight into the cost-effectiveness of different strategies, and can assist in early decision-making when considered with criteria such as equity, public preferences, and political and cultural constraints.


Cervical-cancer screening strategies that involve the use of conventional cytology and require multiple visits have been impractical in developing countries. Authors used computer-based models to assess the cost-effectiveness of a variety of cervical-cancer screening strategies in India, Kenya, Peru, South Africa, and Thailand.


New screening technologies and vaccination against human papillomavirus (HPV), the necessary cause of cervical cancer, may impact optimal approaches to prevent cervical cancer. Authors evaluated the cost-effectiveness of alternative screening strategies to inform cervical cancer prevention guidelines in Norway.


Utilizing data from the Cervical Cancer Prevention in El Salvador (CAPE) demonstration project, authors assessed the health and economic impact of HPV-based screening and two different algorithms for the management of women who test HPV-positive, relative to existing Pap-based screening.

**Ebola Diagnostics**


Early identification of patients suspected to have Ebola virus disease is important for patient management and transmission interruption. WHO is currently working with the Foundation for Innovative New Diagnostics (FIND) and other partners to assess additional rapid assays for potential use in the current or future Ebola outbreaks. This interim guidance is intended to provide Ministries of Health and other organizations information on the potential role for rapid diagnostic tests detecting Ebola antigen for outbreak response and control.

**Also see:**

---

This teaching pack was developed the Center for Health Decision Science at the Harvard T.H. Chan School of Public Health. All materials produced by the Center for Health Decision Science are free and publicly accessible for educational use.

Authors review Ebola rapid diagnostic tests approved by the World Health Organization and those currently in development, and use mathematical modelling to explore the potential benefits of diagnostic testing strategies involving rapid diagnostic tests alone and in combination with polymerase chain reaction testing. Such rapid diagnostic tests could allow early triaging of patients, thereby reducing the potential for nosocomial transmission. In addition, despite the lower test accuracy, rapid diagnostic test-based diagnosis may be beneficial in some contexts because of the reduced time spent by uninfected individuals in healthcare settings where they may be at increased risk of infection; this also frees up hospital beds.

Laboratory diagnosis of Ebola virus disease plays a critical role in outbreak response efforts; however, establishing safe and expedient testing strategies for this high-biosafety-level pathogen in resource-poor environments remains extremely challenging. Since the discovery of Ebola virus in 1976 via traditional viral culture techniques and electron microscopy, diagnostic methodologies have trended toward faster, more accurate molecular assays. Importantly, technological advances have been paired with increasing efforts to support decentralized diagnostic testing capacity that can be deployed at or near the point of patient care. The unprecedented scope of the 2014-2015 West Africa Ebola epidemic spurred tremendous innovation in this arena, and a variety of new diagnostic platforms that have the potential both to immediately improve ongoing surveillance efforts in West Africa and to transform future outbreak responses have reached the field. In this review, the authors describe the evolution of Ebola virus disease diagnostic testing and efforts to deploy field diagnostic laboratories in prior outbreaks. They then explore the diagnostic challenges pervading the 2014-2015 epidemic and provide a comprehensive examination of novel diagnostic tests that are likely to address some of these challenges moving forward.

Rapid case finding and diagnosis have been cornerstones of the strategy to stem the 2014–15 Ebola epidemic. However, during most of this epidemic, timely laboratory confirmation of Ebola virus disease has been complicated by the delay between recognition of a patient with suspected Ebola virus disease in the community, transport or self-presentation of this patient to a treatment site for phlebotomy, and the time needed to transport the sample to a laboratory equipped to run molecular diagnostics safely. This delay has had devastating effects not only at the population level, by extending opportunities for transmission, but also for individual patients.

**Diagnostic Tests for Appendicitis**
Abdominal pain is a common presenting symptom for patients seeking care at emergency departments, with approximately 3.4 million expected cases per year in the United States. Appendicitis is a frequent cause of abdominal pain and occurs in approximately 8 to 10 percent of the population over a lifetime. Appendicitis has its highest incidence between the ages of 10 and 30 years. The ratio of incidence in men and women is 3:2 through the mid-20s and then equalizes after age 30. Appendicitis is the most common abdominal surgical emergency, with over 250,000 appendectomies performed annually in the United States. Untreated appendicitis can lead to perforation of the appendix, which typically occurs within 24 to 48 hours of the onset of symptoms. Perforation of the appendix can cause intra-abdominal infection, sepsis, intraperitoneal abscesses, and rarely death. In order to avoid the sequelae of perforated appendicitis, a low percentage of "negative" appendectomies (i.e., removing a normal noninflamed appendix in patients mistakenly diagnosed with appendicitis) is generally accepted from a surgical standpoint.

**Diagnoses for Pulmonary Embolus**


Authors perform a systematic review and meta-analysis to define the diagnostic performance of pulmonary embolism rule-out criteria (PERC) in deferring the need for D-dimer testing to rule out pulmonary embolism in the emergency department (ED). They find the existing literature suggests consistently high sensitivity and low but acceptable specificity of the PERC to rule out pulmonary embolism in patients with low pretest probability.


Symptoms associated with pulmonary embolism can be nonspecific and similar to many competing diagnoses, leading to excessive costly testing and treatment, as well as missed diagnoses. Objective studies are essential for diagnosis. This study evaluates the cost-effectiveness of different diagnostic strategies in an emergency department (ED) for patients presenting with undifferentiated symptoms suggestive of pulmonary embolism.

**Breast Cancer Screening**

About half of the United States has legislation requiring radiology facilities to disclose mammographic breast density information to women, often with language recommending discussion of supplemental screening options for women with dense breasts. The objective of this study was to examine variation in breast density assessment across radiologists in clinical practice.


In this editorial, Elmore introduces the issue of overdiagnosis, specifically of breast cancer, and discusses Welch’s findings in the same issue. She suggests also that we must improve communications around the subject, including dissemination of scientific findings at the population level through educating patients at the individual level.


The goal of screening mammography is to detect small malignant tumors before they grow large enough to cause symptoms. Effective screening should therefore lead to the detection of a greater number of small tumors, followed by fewer large tumors over time. Authors used data from the Surveillance, Epidemiology, and End Results (SEER) program, 1975 through 2012, to calculate the tumor-size distribution and size-specific incidence of breast cancer among women 40 years of age or older. Although the rate of detection of large tumors fell after the introduction of screening mammography, the more favorable size distribution was primarily the result of the additional detection of small tumors. Women were more likely to have breast cancer that was over diagnosed than to have earlier detection of a tumor that was destined to become large. The reduction in breast cancer mortality after the implementation of screening mammography was predominantly the result of improved systemic therapy.

**Visceral Leishmaniasis Diagnostic Tests**


Rapid diagnostic tests for visceral leishmaniasis (VL) are amongst the most important innovations in the control of VL, where early case detection/treatment can improve patient prognosis and reduce transmission. This user guide provides general information to facilitate proper use of rapid diagnostics tests for, and improve the quality of care of VL.


Visceral leishmaniasis (VL) is a deadly disease caused by infection with the Leishmania parasite. The majority of cases are found in South Asia, east Africa, and Brazil. As many as 310 million people are at risk of infection, and it is estimated that between 20,000 and 50,000 deaths result from VL annually. VL is spread through the bite of the sandfly vector, and it can be harbored by human and canine reservoirs. The parasite causes nonspecific symptoms such as fever and splenomegaly; if left untreated, VL typically leads to death. The
WHO set a number of goals for the Neglected Tropical Diseases (NTD) to be achieved by 2020, and the London Declaration on NTDs backed these goals with commitments from public and private institutions. The 3rd progress report of the London Declaration indicated that “priorities for progress” towards reaching VL goals include early detection of cases, improved access to diagnosis, and scale-up of diagnostic services. In support of the London Declaration goals, PATH aimed to catalyze engagement of the diagnostics industry and product development efforts. As part of this work, PATH assessed needs and landscaped potential solutions to improve diagnostic tools used to support VL elimination efforts. PATH identified four use cases for human VL diagnostics: diagnosing acute infection, diagnosing VL-HIV co-infection, diagnosing post-kala-azar dermal leishmaniasis (PKDL), and treatment monitoring.


Visceral Leishmaniasis (VL), a severe parasitic disease, could be fatal if diagnosis and treatment is delayed. Post kala-azar dermal leishmaniasis (PKDL), a skin related outcome, is a potential reservoir for the spread of VL. Diagnostic tests available for VL such as tissue aspiration are invasive and painful although they are capable of evaluating the treatment response. Serological tests although less invasive than tissue aspiration are incompetent to assess cure. Parasitological examination of slit-skin smear along with the clinical symptoms is routinely used for diagnosis of PKDL. Therefore, a noninvasive test with acceptable sensitivity and competency, additionally, to decide cure would be an asset in disease management and control. Authors describe the development of antibody-capture ELISA and field adaptable dipstick test as noninvasive diagnostic tools for VL and PKDL and as a test of cure in VL treatment.

**Teaching Tools**


To familiarize new investigators with the methodological framework and guidance outlined in the Methods Guide for Medical Test Reviews, AHRQ has created these training modules for faculty engaged in educating investigators and clinicians interested in systematic review methods for medical tests.


