CUMITECH 31

CUMULATIVE TECHNIQUES AND PROCEDURES IN CLINICAL MICROBIOLOGY

FEBRUARY 1997

Verification and Validation of Procedures in the Clinical Microbiology Laboratory

B. LAUREL ELDER, SHARON A. HANSEN, JAMES A. KELLOGG, FREDERIC J. MARSIK, and RONALD J. ZABRANSKY

COORDINATING EDITOR

BRENDA W. McCURDY

ASM PRESS

AMERICAN SOCIETY FOR MICROBIOLOGY

Cumitech 1A • Blood Cultures II • June 1982
Cumitech 2A • Laboratory Diagnosis of Urinary Tract Infections • March 1987
Cumitech 3A • Quality Control and Quality Assurance Practices in Clinical Microbiology • May 1990
Cumitech 4A • Laboratory Diagnosis of Gonorrhea • April 1993
Cumitech 5A • Practical Anaerobic Bacteriology • December 1991
Cumitech 6A • New Developments in Antimicrobial Agent Susceptibility Testing: a Practical Guide • February 1991
Cumitech 7A • Laboratory Diagnosis of Lower Respiratory Tract Infections • September 1987
Cumitech 8 • Detection of Microbial Antigens by Counterimmunoelectrophoresis • December 1978
Cumitech 9 • Collection and Processing of Bacteriological Specimens • August 1979
Cumitech 10 • Laboratory Diagnosis of Upper Respiratory Tract Infections • December 1979
Cumitech 11 • Practical Methods for Culture and Identification of Fungi in the Clinical Microbiology Laboratory • August 1980
Cumitech 12A • Laboratory Diagnosis of Bacterial Diarrhea • April 1992
Cumitech 13A • Laboratory Diagnosis of Ocular Infections • September 1994
Cumitech 14A • Laboratory Diagnosis of Central Nervous System Infections • March 1993
Cumitech 15A • Laboratory Diagnosis of Viral Infections • August 1994
Cumitech 16A • Laboratory Diagnosis of the Mycobacterioses • October 1994
Cumitech 17A • Laboratory Diagnosis of Female Genital Tract Infections • June 1993
Cumitech 18 • Laboratory Diagnosis of Hepatitis Viruses • January 1984
Cumitech 19 • Laboratory Diagnosis of Chlamydial and Mycoplasmal Infections • August 1984
Cumitech 20 • Therapeutic Drug Monitoring: Antimicrobial Agents • October 1984
Cumitech 21 • Laboratory Diagnosis of Viral Respiratory Disease • March 1986
Cumitech 22 • Immunoserology of Staphylococcal Disease • August 1987
Cumitech 23 • Infections of the Skin and Subcutaneous Tissues • June 1988
Cumitech 24 • Rapid Detection of Viruses by Immunofluorescence • August 1988
Cumitech 25 • Current Concepts and Approaches to Antimicrobial Agent Susceptibility Testing • December 1988
Cumitech 26 • Laboratory Diagnosis of Viral Infections Producing Enteritis • September 1989
Cumitech 27 • Laboratory Diagnosis of Zoonotic Infections: Bacterial Infections Obtained from Companion and Laboratory Animals • February 1996
Cumitech 28 • Laboratory Diagnosis of Zoonotic Infections: Chlamydial, Fungal, Viral, and Parasitic Infections Obtained from Companion and Laboratory Animals • February 1996
Cumitech 29 • Laboratory Safety in Clinical Microbiology • July 1996
Cumitech 30 • Selection and Use of Laboratory Procedures for Diagnosis of Parasitic Infections of the Gastrointestinal Tract • September 1996
Cumitech 31 • Verification and Validation of Procedures in the Clinical Microbiology Laboratory • February 1997

Cumitechs should be cited as follows, e.g.: Elder, B. L., S. A. Hansen, J. A. Kellogg, F. J. Marsik, and R. J. Zabransky. 1997. Verification and Validation of Procedures in the Clinical Microbiology Laboratory. Coordinating ed., B. W. McCurdy. American Society for Microbiology, Washington, D.C.

Editorial Board for ASM Cumitechs: Frederick S. Nolte, *Chairman*; Lorraine Clarke, Janet Hindler, Brenda W. McCurdy, Alice S. Weissfeld, and Stephen Young.

The purpose of the Cumitech series is to provide consensus recommendations by the authors as to appropriate state-of-the-art operating procedures for clinical microbiology laboratories which may lack the facilities for fully evaluating routine or new methods. The procedures given are not proposed as "standard" methods.

Copyright © 1997 American Society for Microbiology
1325 Massachusetts Ave., N.W.
Washington, DC 20005

VERIFICATION AND VALIDATION OF PROCEDURES IN THE CLINICAL MICROBIOLOGY LABORATORY

B. LAUREL ELDER, *CompuNet Clinical Laboratories, Moraine, Ohio 45439*

SHARON A. HANSEN, *Food and Drug Administration, Rockville, Maryland 20850*

JAMES A. KELLOGG, *Department of Pathology, York Hospital, York, Pennsylvania 17405-7198*

FREDERIC J. MARSIK, *Food and Drug Administration, Rockville, Maryland 20857*

RONALD J. ZABRANSKY, *Pathology and Laboratory Medicine, VA Medical Center, Cleveland, Ohio 44106*

COORDINATING EDITOR

BRENDA W. McCURDY, *VA Medical Center, Detroit, Michigan 48201-1932*

The key desirable attribute of a laboratory test is its ability to produce accurate and precise results consistently over extended periods of time with an appropriately rapid turnaround time so that the test results are of clinical utility. Accuracy and precision are readily measurable and have become the hallmarks for comparing tests and laboratories, and they have even been used as criteria for granting certification, licensure, or accreditation. Furthermore, when a new laboratory test is introduced, these factors may be weighed with other issues, such as clinical relevance, cost, instrumentation, and ease of performance. The laboratory must have in place a detailed method for analysis of new tests to evaluate these parameters before the tests are introduced as new offerings or as replacements for older procedures (verification). In addition, the laboratory must have a systematic approach for the ongoing review of established tests which demonstrates that the testing process and the consistency of results have not changed and that testing personnel remain competent to perform tests and report results (validation).

The processes of verification and validation are part of the quality assurance program for the laboratory. Laboratories are required to establish policies and procedures to maintain or improve the reliability, efficiency, and clinical utility of laboratory tests. The purpose of this Cumitech is to provide guidance on the necessary criteria that may be required as new tests are considered for clinical use and as old tests are reevaluated for their clinical relevance. This document is a guide by which a laboratory may verify the performance of a new method or undertake the revision of an established method to extend its applicability or improve its performance. The guidelines within this Cumitech apply equally to simple single-reagent tests and the most complex of instruments generating a variety of analytic results and interpretations.

The guidelines and suggestions found in this Cumitech offer information that users of tests may consider in their efforts to improve the general operation or quality of their laboratory services. These guidelines should not be considered minimal criteria or policy which regulatory, accrediting, licensing, or standard-setting agencies use in assessing either the compliance of a laboratory or the accuracy of individual tests or instruments that may be considered for acceptance by laboratory personnel.

The information in this Cumitech is not all-inclusive. Much information can be found in the literature regarding the verification and validation of test methods; more often, however, this information applies to quantitative analyses, as are commonly found in the chemistry section of the laboratory. When this process is applied to microbiology, where qualitative results are more common, where subjective interpretations are required, or where the results include the identification of microorganisms with biological variation, more flexibility is required. Thus, the focus of this Cumitech is on the common qualitative and semiquantitative test procedures performed in the clinical microbiology laboratory.

DEFINITIONS

The primary processes addressed in this Cumitech are the verification of a new test or method prior to introduction into the laboratory and the ongoing validation of the performance of existing test methodology. However, before these processes can be discussed in detail, relevant definitions must be established.

Accuracy (20, 23, 24) : *Technical accuracy* is the

nearness of an individual measurement to the true value, as determined by a reference method. *Clinical accuracy* is the overall ability of a test to both rule in and rule out an analyte or a specific disease. Accuracy is synonymous with test efficiency and can be expressed mathematically as a percent:

$$\frac{\text{number of correct results}}{\text{total number of results}} \times 100$$

Analyte (20, 23): The component of a specimen or organism which is to be measured or demonstrated. An analyte may be a particular antigen, antibody, nucleic acid, organism, enzyme, species, or metabolic product.

Gold standard: The best available approximation of the truth (20). It is a commonly used term, generally indicating a test method currently accepted as reasonably, but not necessarily 100%, accurate (8, 23). It is used as the reference method for assessing the performance characteristics of another test method. When the true disease status of a patient is unknown and the disease state is being determined by using a test compared with an imperfect gold standard, the results will be skewed and errors of the reference method will be magnified. In this situation, it may be more appropriate to display the agreement and disagreement between the gold standard and new test in graphic or tabular form. Areas of disagreement may then be further investigated by other tests or by following the patient's condition to determine if disease develops. In those instances in which it cannot be determined whether the new test is better than the gold standard, a decision to use the new test alone or in combination with the gold standard may be made by using a cost-benefit analysis (8). Problems associated with use of an imperfect gold standard are more fully discussed in the National Committee for Clinical Laboratory Standards (NCCLS) publication I/LA 18-A (23) and in references 6 and 11.

Home brew test: A procedure developed in-house that uses reagents that are either commercially available or produced in-house *or* any procedure that incorporates modifications of the manufacturer's package insert instructions. Reagents provided by companies for either investigative or research use must be used in accordance with the guidelines described by the company and/or the U.S. Food and Drug Administration (FDA) and must not be used for in vitro diagnostic testing. Since a home brew test is not reviewed by the FDA and since the reagents used therein cannot be reviewed for their intended use, these tests require a more extensive performance verification.

New test: A new test includes any test not previously offered by a laboratory, a procedure or methodology change, or a test performed in-house that was previously performed at a reference laboratory. Such tests include detection or identification of a totally new analyte, the use of totally new methodology, a new approach to detecting an analyte, a change from a manual method to an automated one, a new application of existing technology, or the test of a new matrix (old analyte in a different specimen).

Old test: An old test is any procedure for detection of a disease, analyte, or characteristic (e.g., antimicrobial susceptibility) that had been in use prior to September 1, 1992, the effective date of the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88).

Precision (20, 23): A measure of the extent to which replicate analyses (using identical procedures) of a homogeneous analyte agree with each other. Precision implies freedom from inconsistency and random error but does not guarantee accuracy. Precision is synonymous with reproducibility; however, the use of the term precision is generally applied to quantitative assays, while reproducibility is used with qualitative analyses. Mathematically, precision can be expressed as a percent:

$$\frac{\text{number of repeated results in agreement}}{\text{total number of results}} \times 100$$

Predictive value (8, 20, 23): The predictive value of a test is the probability that a positive result (positive predictive value, or PPV) accurately indicates the presence of an analyte or a specific disease or that a negative result (negative predictive value, or NPV) accurately indicates the absence of an analyte or a specific disease. PPV is expressed as a percent:

$$\frac{\text{number of true-positive results}}{\text{number of true-positive plus false-positive results}} \times 100$$

NPV is expressed as:

$$\frac{\text{number of true-negative results}}{\text{number of true-negative plus false-negative results}} \times 100$$

Predictive values can vary significantly with the prevalence of the disease or analyte unless the test is 100% sensitive (for NPV) or specific (for PPV). The highest predictive values are desired when inappropriate treatment due to false-positive or -negative results has serious clinical, emotional, epidemiological, public health, or economic consequences. Predictive values are most meaningful in evaluating a test's performance in specific risk population groups (see Appendix A).

Prevalence (24): The pretest probability of a particular clinical state in a specified population; the frequency of a disease in the population of interest at a given point in time.

Quality assurance (28, 29): A system for continuously improving and monitoring the reliability, efficiency, and clinical utilization of laboratory tests. Quality control, quality improvement, and method validation are integral components of quality assurance.

Quality control (29): The process of ongoing performance checks, including personnel performance, using known organisms or analytes to ensure on a regular and frequent basis that a method which has gone through the verification process (see below) and is now part of the laboratory's routine test battery is performing as expected. Quality control systematically detects deficiencies in testing by setting limits of acceptable performance (accuracy and precision). It thus allows detection, and corrective action where appropriate, of major problems or errors with test systems and their performance. Quality control implies that there exist standard analytes that have known reactions or reactivity. Quality control is an integral part of the test validation process.

Quality improvement (29): The prevention of test deficiencies and enhancement of a test's clinical utility by establishing a thorough understanding of the test's capabilities and limitations, as gathered from experience and observation, and the subsequent use of this knowledge to make and verify procedural changes for improved test performance.

Reference method (20): A thoroughly investigated method, in which exact and clear descriptions of the necessary conditions and procedures are given for the accurate determination of one or more values; the documented accuracy and precision of the method are commensurate with the method's use for assessing the accuracy of other methods for measuring the same property values or for assigning reference method values to reference materials. A currently used method is unacceptable as a reference method unless there is on-site or peer-review journal documentation of an acceptable level of accuracy and precision of the method.

Reproducibility: See Precision.

Sensitivity (8, 20, 23, 24): A measure of a test's efficiency in detecting an analyte or a specific disease. *Analytic sensitivity* measures the smallest quantity of an analyte that can be reproducibly distinguished from background levels, or a zero calibrator, in a given assay system; it is usually defined at the 0.95 confidence level (± 2 standard deviations) and is more appropriately called "detection limit." In microbiology, the detection limit can be correlated to the number of colonies in culture or the lowest quantity of antigen a test can detect. *Clinical sensitivity* is the percent test positivity in a population of affected patients. The highest sensitivity is desired when a disease is serious and treatable and when false-positive results will not lead to serious clinical or economic problems. Mathematically, sensitivity is expressed as a percent:

$$\frac{\text{number of true-positive results}}{\text{number of true-positive plus false-negative results}} \times 100$$

Specificity (8, 20, 23, 24): A measure of a test's efficiency in ruling out an analyte or a specific disease. *Analytic specificity* is the ability of an analytical method to detect only the analyte that it was designed to measure. *Clinical specificity* is the percent of negative test results in a population without the specified disease. The highest specificity is desired when the disease is serious but not treatable, when disease absence has either psychological or public health value, or when false-positive results might cause serious clinical or economic problems. Mathematically, specificity is expressed as a percent:

$$\frac{\text{number of true-negative results}}{\text{number of true-negative plus false-positive results}} \times 100$$

Validation: The documentation that a test which has already been verified is repeatedly giving the expected results as the test is performed over a period of time. Validation confirms that the test continues to perform satisfactorily according to the laboratory's requirements or the manufacturer's claims or, for home brew tests, according to its intended use. The requirements for test validation may include personnel competency assessment, quality control, internal and external proficiency testing, and correlation with clinical findings. Validation thus becomes an integral part of the laboratory's quality assurance program.

Verification (7, 29): The documentation of either commercial or home brew test accuracy. For commercially obtained tests, it is the process of examination or evaluation of a test system to determine whether the claims stipulated by the manufacturer in the package insert as they relate to the product, the process, the results, or the interpretation can be achieved. Verification requires determination or confirmation of the test performance characteristics, including sensitivity, specificity, and, where appropriate, the predictive values, precision, and accuracy of the test. Verification is a one-time process, completed before the test or system is used for patient testing.

FEDERAL REGULATION: ROLE OF CLIA '88

On February 28, 1992, the Department of Health and Human Services published the final regulation implementing the Clinical Laboratory Improvement Amendments of 1988 (7). CLIA '88 extended federal regulation to cover all laboratories that examine human specimens for the diagnosis, prevention, or treatment of any disease or impairment or for the assessment of the health of human beings. Sections 493.1201 through 493.1285 of subpart K of the regulation set forth the quality control requirements, which became effective September 1, 1992, for tests of moderate or high complexity (waivered tests are not subject to quality control). For tests of moderate complexity, quality control requirements were to be phased in over a two-year period. For tests of high complexity, the category of the majority of tests in microbiology, the requirements became effective in full as of that date.

Section 493.1213 describes the requirement for the establishment and verification of method performance specifications. The requirement only applies to instruments, kits, or test systems introduced after the effective date of the quality control regulations, September 1, 1992. After that date, a laboratory that introduces a new procedure for patient testing which uses a method developed in-house, a modification of the manufacturer's test procedure, or a method that has not been cleared by the FDA as meeting the CLIA requirements for general quality control must, prior to reporting patient test results, verify or establish for each method the performance specifications for the following performance characteristics, as applicable:

- Accuracy
- Precision
- Analytical sensitivity
- Analytical specificity, to include interfering substances
- Reportable range of patient test results
- Reference range(s)
- Any other characteristic required for test performance and interpretation of results

Also, control procedures for patient testing must be established on the basis of the verified performance specifications. The laboratory is required to have documentation of the verification of manufacturer's specifications or establishment of performance specifications for tests developed in-house. All laboratory accrediting programs which have been granted deemed status by the Health Care Financing Administration (HCFA) (including the College of American Pathologists, the Joint Commission on Accreditation of Healthcare Organizations, and the Commission on Laboratory Accreditation) must include these requirements as part of their inspection criteria.

Since to date the FDA has not implemented the process of clearing tests that meet CLIA requirements for general quality control, all tests of high complexity are subject to verification. Verification is not required for tests of moderate complexity unless such tests were developed in-house, are not cleared by the FDA, or have been modified by the laboratory. Validation requirements for tests of both moderate and high complexity are delineated in the quality assurance and specific quality control sections of the regulation.

It should be noted that the 1992 regulations were published as a final rule with comment and, as of May 1996, are undergoing revision on the basis of the comments and recommendations from the Clinical Laboratory Improvement Advisory Committee (CLIAC). Publication of the final regulations has been projected for 1997, but final deliberations are still in process as of this writing.

ROLE OF THE FDA IN REGULATION OF CLINICAL MICROBIOLOGY IN VITRO DEVICES

Considerable confusion often exists concerning the process followed by the manufacturer of a test kit or system to gain FDA clearance for marketing of the product and the relationship of this process to the requirement for later verification in the user laboratory. An understanding of the processes utilized by the FDA can help the end user of a product understand the reasons for the regulatory requirement that the laboratory verify the performance of a test approved and/or cleared by the FDA in its own setting.

The Medical Device Amendments to the Food, Drug, and Cosmetic Act were enacted in 1976 and directed the FDA to regulate medical devices under the appropriate control levels necessary to ensure safety and effectiveness. The Safe Medical Devices Act of 1990, a major revision of the 1976 amendments, added new provisions to better ensure that devices entering the market were safe and effective and provided means for the FDA to learn quickly about serious device problems and remove defective devices from the market. The Food, Drug, and Cosmetic Act and the Safe Medical Devices Act are administered by the FDA's Center for Devices and Radiological Health through the Office of Device Evaluation, which includes the Division of Clinical Laboratory Devices.

A manufacturer can legally place an in vitro diagnostic device into the market in two main ways. The first is by Premarket Notification 510(k), in which a manufacturer demonstrates that its device is substantially equivalent to a preamendment device (a device in use before 1976) or a predicate device (a similar legally

marketed device) in terms of both safety and effectiveness. Typically in a 510(k), the manufacturer must describe the device methodology/technology, its intended use, indications for its use, and its performance characteristics as shown in labeling (product insert) and in promotional material and advertisements. The manufacturer must also compare and contrast the submitted device with a similarly legally marketed device with supporting data, and, in the case of modified devices for which the manufacturer has determined that the modification could affect the safety and/or effectiveness of the device, the manufacturer must provide documentation/data to address that effect(s). The 510(k) review is entirely a paper review; the FDA does not submit the actual products to direct laboratory evaluation. The agency therefore has no hands-on experience with the vast majority of products under review. If the FDA assessment of the 510(k) submission indicates that it is substantially equivalent to a legally marketed device, the device is cleared and the manufacturer is free to market it in the United States. The FDA has granted exemptions from the requirement for 510(k) notification for a variety of generic-type devices, including such microbiology products as anaerobic chambers, incubators, gas-generating devices, and most media.

A second pathway to market is via the premarket approval (PMA) application. In this case, there is no preamendment device and the manufacturer must provide reasonable assurance of safety and effectiveness under conditions of intended use(s). Before approval, a PMA receives an in-depth scientific review, and the firm must undergo a comprehensive good manufacturing practice inspection. Finally, the PMA is reviewed by an FDA advisory panel of outside experts who provide recommendations to the FDA for approval with or without conditions or for disapproval of the application. Examples of microbiology devices requiring PMA applications are devices directed at detection or typing of human papillomavirus; all hepatitis and human immunodeficiency virus diagnostic, detection, and monitoring devices; and devices using nucleic acid amplification techniques for direct detection of *Mycobacterium tuberculosis* from clinical material.

The proposed package insert is a part of the 510(k) or PMA application. After all analytical and clinical data have been critically reviewed, the final step of the Microbiology Branch review is to "clear or approve" the product labeling (thus, product clearance by the FDA does not guarantee test performance). The Branch pays particular attention to the following components of the package insert, all of which are required by law (21 CFR 809.10): intended use; specimen collection, transport, and storage recommendations; warnings and limitations; expected values; validation of cutoff; results and their interpretation; quality control recommendations; and specific performance characteristics. All microbiology products which undergo a scientific evaluation of data to substantiate product performance claims as stated in the product insert (i.e., moderate and high risk devices) are expected by the FDA to maintain that performance throughout the life of the product. Failure to maintain that expected performance could result in compliance or regulatory action. Promotional and advertising material also falls under the labeling regulation. Such material can only be reflective of the information contained in the package insert for the device.

Devices which are in the laboratory research phase of development may not be represented as effective diagnostic products, and the statement "For Research Use Only. Not for use in diagnostic procedures" must be prominently placed in the labeling. A product being shipped or delivered for product testing prior to full commercial marketing must prominently bear the statement "For Investigational Use Only. The performance characteristics of this product have not been established." Only in vitro devices which have been 510(k) cleared or PMA approved by the FDA may legally have the statement "For In Vitro Diagnostic Use" as part of their package insert. In the clinical laboratory, results from the following must not be reported without verification: (i) tests or procedures that have been developed in-house and use class I reagents that do not have an indicated use for that test or (ii) reagents or tests provided by companies for investigative or research use only that are not used within the guidelines described by the company or by the FDA. Reports of such results should indicate that the test is not FDA cleared or approved but has been developed and verified in-house. Since home brew tests are not reviewed by the FDA and since the reagents used therein cannot be reviewed for their use in those tests, home brew tests require a more extensive performance verification. Furthermore, validation must be performed at least every six months when no commercial quality control reagents or proficiency test samples are available (2, 7).

FDA clearance or approval of a product does not predict how that product will perform in the end user's laboratory under actual testing conditions and with the specimen mix encountered in a particular patient population. Thus, initial clearance by the FDA cannot be used as a substitute for verification of test performance by the performing laboratory. In addition, test verification and validation by clinical laboratories now also play a critical role in ensuring that devices and medical products previously cleared/approved by the FDA are performing as expected. On December 11, 1995, the final rule for mandatory reporting of medical product-related serious incidents

was published (4). The rule requires hospitals and other health care facilities for the first time to report deaths and serious injuries connected with the use of medical products (including in vitro diagnostic devices) to the manufacturer and directly to the FDA. The agency evaluates the seriousness of the health hazard, takes corrective action, and communicates that action to the health professional community. In addition, product problems (e.g., suspected contamination, questionable stability, defective components, poor packaging or labeling, and device malfunction) should be reported to the manufacturer.

SELECTION OF A LABORATORY METHOD

Once a laboratory has reached the decision to offer a new test, the next step in the process is generally selection of the method by which the test will be performed. Few laboratories have the time or resources to perform in-house evaluations of the large number of available test systems, kits, or methods which may be available to the microbiology laboratory for detection of an organism, antigen, or other analyte of interest. Thus, it becomes crucial to approach the selection of a new method in an organized fashion, making use of all available information to narrow down the selection of a laboratory method without performing expensive in-house studies. The following steps are designed to serve as a guide for the initial selection of a laboratory method. Although all steps may not be necessary for every test method under consideration by a laboratory, the basic process can be followed for the majority of tests utilized by the microbiology laboratory.

1. Define the purpose for which the method is to be used. Common purposes for tests include the following (23).

- **Screening.** Screening is used for testing large populations of patients for the presence of a disease state or analyte (such as an infectious agent). In general, screening tests should have high (i.e., greater than 95%) clinical sensitivity and negative predictive values. In most cases, the recommended specificity and positive predictive value can be lower than those of diagnostic and confirmatory tests. Thus, a negative screening test result should indicate that the person has a high probability of being free of the characteristic, whereas a positive test result might reflect only the need for confirmatory testing.
- **Confirmation.** Confirmation is used after obtaining a positive screening or diagnostic test result to ensure the accuracy of that initial result. Specificity and positive predictive value, rather than sensitivity and negative predictive value, are usually the primary considerations for confirmatory tests; specificity should exceed 98%. Confirmatory tests may not be necessary when the screening or diagnostic test has high specificity and positive predictive values.
- **Diagnosis.** Diagnosis is used for the evaluation of persons suspected of having a given disease state or characteristic (e.g., a particular type of infection). If the characteristic is important, either for treatment or for prognostic considerations, sensitivity should be as high as possible. When diagnostic test results are not confirmed by additional laboratory or clinical data, specificity may also need to be very high (see Appendix A). However, if an accurate confirmatory test is readily available, a high degree of specificity might not be necessary. The majority of clinical tests for infectious diseases are for diagnostic use.

2. Decide what analyte (e.g., organism, antigen, nucleic acid, etc.) is to be detected and what the reference method or gold standard will be for comparison. Note: if the new test is likely to be more sensitive than the gold standard, then ways to arbitrate discrepant results should be considered (e.g., clinical data, other assays, etc.) (6, 11, 23).

3. In conjunction with the end user of the test (e.g., the physician) and the information from steps 1 and 2, determine the medical usefulness of the test (e.g., whether the test will lead to improved patient care and/or shortened hospital stay) and preliminary clinical and/or microbiological requirements for test sensitivity, specificity, and predictive values (as appropriate).

4. Survey the technical and medical literature for performance claims of various methods that may indicate that one or more methods will meet the initial requirements for sensitivity, specificity, etc. When reviewing the literature, confirm that the method described is actually the test (unmodified) that is to be evaluated in the laboratory.

5. Determine the characteristics of the method(s) of interest. The choice of method may also be based on the following practical parameters (9, 14, 27).

- Cost of the method. What are the comparative costs for material and labor relative to alternatives to the test? What is the extent of reimbursement?
- Practicality in the laboratory setting
  Can the test be performed on all necessary shifts?
  Does the test require special equipment?
  What is the turnaround time for the test?
  What are the personnel requirements?
  Are quality control and proficiency test materials available?
  What is the extent of quality control that will need to be performed?

Is there adequate space in the laboratory to perform the test?
Can the test be automated to reduce labor?
Does the system have an indication for all the uses/organisms that are of interest?

- Specimen requirements
  Volume and type of specimen needed
  Collection requirements
  Transport requirements
  Storage requirements
  Quality of specimen
- Quantities of reagents and controls needed for test; storage requirements
- Shelf-life of reagents and controls before and after opening
- Availability of supplies, service, and/or technical support
- Possible safety hazards related to performing test
- Whether a reference range is appropriate for the test, and how it will be determined for the institution

6. Make a preliminary selection of a test method and perform the in-house verification. A brief outline of this process is outlined in Appendix B.

VERIFICATION OF COMMON MICROBIOLOGY TESTS

Verification of a test's performance parameters is accomplished by performing the new or revised test method in parallel with a reference method that has an established and satisfactory level of accuracy. The results of test verification should indicate one of three possibilities:

- The test is acceptable for routine use
- Further verification studies are required
- Immediate corrective action is required by the manufacturer (if commercially obtained), the user, or both. The test is unsuitable for routine use until its performance parameters can be verified.

Certain commonly used microbiology items are not considered instruments, kits, or test systems under CLIA '88 and may not require a complete verification process prior to use. Items such as media or individual reagents used as components of the identification process (e.g., oxidase, catalase) may instead be monitored through the quality control protocols of the laboratory. However, each laboratory must assess the nature and purpose of each of these components and may choose to perform more elaborate in-house verification. For example, a decision may reasonably be made not to perform verification on common media with a history of limited failures, such as most of the media listed in Table 3 of NCCLS document M22-A (17), but it should be performed on some of the highly complex media which are multifunctional (e.g., a medium which is selective for an organism and contains biochemicals that are then used to presumptively identify the organism). This would include *Campylobacter* agar, media for the selective isolation of the pathogenic *Neisseria* spp., and any other media not listed in Table 3 of reference 17.

In the following sections, suggested methods are included for verification of many of the commonly used tests and test systems found in the microbiology laboratory. These suggestions are not meant to be all-inclusive, and alternative approaches may be utilized in individual laboratories. In addition, it is recognized that the verification process can be timely and expensive, often complicated by a paucity of specimens or samples containing or lacking the desired analyte. In some cases, laboratories will need to make difficult choices about the extent of verification that is possible, taking into consideration how widely the test has been used and accepted by the microbiology community, the extent and results of published evaluations, and the impact of an incorrect test result on the patient. In some cases, repeat testing of selected control material near the test cutoff value(s) may give a level of satisfaction. In other cases, laboratories may decide that they are unable to perform a reasonable verification and may choose to refer the test to another laboratory. Whenever possible, purchase of a new system or test methodology should be made contingent upon the results of the verification studies. Records of the actual test verification results must be maintained at least two years. However, since the laboratory must be able to provide the ordering physician with the performance parameters of each test it performs, good laboratory practice would dictate that the records of test verification be kept for as long as the test is in use.

Antimicrobial Susceptibility Systems

The generation of antimicrobial susceptibility test results is one of the most important functions of the microbiology laboratory, as the results may directly affect the therapy chosen for treatment of an infection. Thus, it is critical that the microbiologist be confident that the system chosen is able to provide accurate and reliable results in the user's own laboratory. Within the past 10 years, a number of recommendations for verification of susceptibility test systems have been discussed in the literature (5, 10, 16). Although it may be difficult for laboratories to perform a rigorous study of a new system, use of selected control and clinical isolates can aid in the effort to verify the claims made in the literature and by the manufacturer regarding the accuracy and reproducibility of a system (10).

The evaluation must be designed to allow detection of the following types of errors (10, 16):

- **Very major errors.** The system under evaluation indicates a "susceptible" response while the reference method indicates a "resistant" response. Clinically, very major errors are the most serious type of error and can only be detected by testing organisms resistant to each antimicrobial agent.
- **Major errors.** The system under evaluation indicates a "resistant" response while the reference method indicates a "susceptible" response. Major errors can only be detected by testing organisms susceptible to each antimicrobial agent.
- **Minor errors.** Either the new system or the reference method indicates an "intermediate" result. The other method indicates either a "susceptible" or "resistant" result.

Evaluation of susceptibility test methods should be done using a distribution of organisms similar to those commonly isolated but should include susceptible and resistant isolates for each antibiotic whenever possible. The distribution should contain examples of clinically relevant isolates appropriate for the class of compounds routinely tested and should include isolates showing important resistance mechanisms. For example, methicillin-resistant *Staphylococcus aureus* and coagulase-negative staphylococci should be included in the evaluation of antistaphylococcal agents (22). Although it may be desirable to test at least 35 resistant isolates for each antibiotic in the test system (10), this is not always possible or practical for laboratories. In these cases, laboratories should carefully scrutinize the organisms to be tested to ensure that they make the best choices possible when selecting isolates for testing.

Since very major errors can only occur with isolates that are resistant, very major error rates should be calculated using the number of resistant strains as the denominator rather than the total number of isolates tested (5, 10, 16). For the same reason, major error rates should be calculated by using the total number of susceptible isolates as the denominator. Very major errors determined for a large sample ($n = \geq 35$) of known resistant isolates should be $\leq 3\%$ (10). The combination of major and minor errors attributable to the new test should be $\leq 7\%$ when determined for a large known-susceptible population or a large unselected sample of clinical isolates (a minimum of 100 strains) (10). More stringent criteria are suggested by NCCLS (22) and required by the FDA for manufacturers (an acceptable error rate of less than 1.5% for very major errors and less than 3% for major errors; however, for the laboratory to demonstrate that it has met the more stringent criteria for very major errors, a minimum of 65 resistant isolates would need to be tested against each antibiotic). For antibiotics without an intermediate interpretive category (where a one-dilution error in the MIC might spuriously appear as a very major or major error), there should be $>90\%$ agreement within one dilution between the two procedures being evaluated (10). Growth failures for susceptibility systems should not exceed 10% of the total number of isolates tested (5). If the chosen limits are exceeded for any of the error types for any drug-isolate combination, the test must be considered "unverified," and a corrective action investigation must be undertaken in conjunction with the manufacturer to attempt to resolve the discrepancies. Following corrective action, the new or revised test should be run again in parallel with the reference method on a minimum of 20 appropriate isolates (isolates which will demonstrate that the problem[s] has been corrected).

Diagnostic Microbiology Tests

Diagnostic microbiology tests (nonculture tests for detection of microorganisms, microbial antigens, nucleic acids, or antibodies from clinical specimens) may be used for screening, confirmatory, or diagnostic purposes. The process for verifying a diagnostic microbiology test may be relatively straightforward for analytes which are common in the population. When the new test will be detecting a relatively uncommon analyte (for example, a direct test for the detection of *Cryptosporidium*), assessment of the sensitivity of the test in the user's own patient population becomes difficult. In these circumstances, the manufacturer, other laboratories, or commercial sources may be able to provide specimens of known content to be used in the evaluation.

It is important to remember that a test may be suitable for one population of patients and not for another. Laboratories must evaluate this individually and be prepared to provide physicians with the specifications of the test for individual populations with different disease prevalences.

Verification of commercially-obtained tests performed exactly as described in the manufacturer's product insert

The new or revised test should be performed in parallel with the existing test or a reference method on a minimum of 20 specimens that contain each target analyte and a minimum of 50 specimens that lack the target analyte. An effort should be made to include weakly positive specimens that will challenge the detection limit of the assay. If these are unavailable, dilution of strongly positive specimens with an appropriate material (e.g., normal serum) can achieve the same effect. In some circumstances, it may be necessary to test more

specimens to document that the test meets the required level of sensitivity, specificity, or positive and negative predictive values. An example of this decision-making process can be found in Appendix A.

After comparing the results of the new or revised test with those of the reference method, the number of true-positive, true-negative, false-positive, and false-negative results obtained with the test to be verified should be documented. When the reference method is known to be an imperfect standard, an attempt to resolve discrepancies should be made. By using these results, the sensitivity, specificity, predictive values, accuracy, and precision of the new or revised test may be calculated. The calculated predictive values may not be indicative of the performance of the test in an actual patient population if the prevalence of the target analyte is different from that used in the verification study. Remember that a test with a specificity of 95% when the prevalence of a target disease is only 1% will have a predictive value of a positive result of only 16.7%. Individual predictive values may be calculated from the sensitivity and specificity data for known different prevalences, as seen in the example in Appendix A. The test may be considered verified if it meets the requirements initially established for performance by the users of the test *and* if the sensitivity and specificity are no lower than 5% below those of the reference method, those appearing in peer-reviewed journals, or those claimed by the manufacturer in the product insert. Whenever possible, data published in a peer-reviewed journal, rather than the manufacturer's marketing data, should be used in the evaluation of test kits and reagents (23). Even the results of published studies should be carefully analyzed for the procedure used (13) .

If the sensitivity or specificity of the new or revised test does not satisfy the verification requirements, the test must be considered unverified and corrective action must be taken by the manufacturer, the user, or both. Following corrective action, the new or revised test should be run again in parallel with the reference method and interpreted as described above.

Verification of home brew diagnostic tests

The new or revised test should be verified as described above, except that a minimum of 50 specimens that contain the target analyte and a minimum of 100 specimens that lack the target analyte should be studied. If either the sensitivity or specificity is lower than 5% below the values of the reference method, verification tests should be repeated, after appropriate corrective action has been taken.

In some instances, test verification will be required because of very small modifications (e.g., a minimal change in incubation time) of an existing protocol which has been previously verified. In this situation, verification must be performed to the extent necessary to demonstrate that the change has not affected the performance of the test, but it may not require the extensive testing performed initially. It may be useful to maintain a panel of a limited number of well-characterized specimens and to only do a complete verification if these specimens do not give satisfactory results.

Microbial or Microbial Product Identification Tests

Microbial or microbial product identification tests include antisera, antigens, chemicals, stains, instruments, reagents, or kits used to identify microorganisms or their products from culture.

Verification of commercially obtained tests performed exactly as described in the manufacturer's product insert

Evaluation of test systems for identification of microorganisms to the species level may be structured around one of the following three suggestions (13).

- Perform at least 1 week of consecutive parallel testing (minimum of 50 strains) with the existing method. Discrepancies must be arbitrated.
- Test known representative strains (stock cultures) of a minimum of 12 to 15 commonly isolated species of organisms, for a total of 50 or more tests.
- Confirm that 20 to 50 organism identifications (12 to 15 different species) agree in concurrent testing with the current method (discrepancies must be arbitrated) or with the results of reference laboratory testing of split samples.

In each case, the appropriate quality control organisms should also be tested and included in the verification process.

For identification methods that detect only one analyte (e.g., coagulase, indole, oxidase, immunofluorescent reagents, etc.), the new or revised test should be run in parallel with the existing test or a reference method on a minimum of 20 microbial isolates that contain the target analyte and 20 isolates that lack the target analyte.

When the results of the verification of tests which identify to the species level are evaluated, both the level of agreement between the new and reference method and the types of errors or disagreements should be evaluated. Overall, there should be at least 90% agreement with the existing system or reference method before the new method is considered verified. Certain groups of organisms will commonly be more challenging for new systems to identify (e.g., nonfermenters, corynebacteria, coagulase-negative staphylococci), and greater flexibility may be necessary in assessing the accuracy of the new method. Other

groups of organisms (e.g., members of the family *Enterobacteriaceae*) should be identified with high levels of accuracy, and these test systems should be held to a higher (e.g., 95%) requirement for agreement (14). In addition, the types of disagreements encountered with the new system should be scrutinized. The new system may misidentify an organism, may require further tests before identifying an organism, or may give no identification at all. Misidentification is the most serious error for an identification system, while a laboratory may choose to accept a certain number of isolates with no or partial identification if other factors (e.g., cost or speed) outweigh the inconvenience of further testing.

When the results of the verification of tests which detect a particular analyte are evaluated, the sensitivity and specificity of the new test should be calculated. The test may be considered verified if the sensitivity and specificity are no lower than 5% below those values of the reference method.

Verification of home brew identification tests

If the new or revised test identifies isolates to the species level, it should be tested in parallel with the existing test or another reference method on a minimum of 200 isolates. Whenever possible, these isolates should include all species identifiable by the new or revised test. The same requirements for method agreement described for commercial methods (at least 90% agreement overall) must be met by home brew tests to consider the method verified.

If the new or revised test identifies a particular analyte, it should be tested in parallel with the existing test or a reference method on a minimum of 50 microbial isolates that contain the target analyte and a minimum of 100 isolates that lack the target analyte. After the results of the new or revised test are compared with those of the reference method, the number of true-positive, true-negative, false-positive, and false-negative results obtained with the test to be verified should be documented. When the reference method is known to be an imperfect standard, an attempt to resolve discrepancies should be made. By using these results, the sensitivity, specificity, predictive values (if appropriate), accuracy, and precision of the new or revised test may be calculated. The test may be considered verified if the sensitivity and specificity are no lower than 5% below those of the reference method or those appearing in the peer-reviewed literature, whichever is higher.

If the new or revised test does not satisfy the verification requirements, the test must be considered unverified and corrective action must be taken. Following corrective action, the new or revised test should be compared again with the reference method as described above.

Blood Culture Systems

Meaningful verification of a new blood culture system is one of the most difficult tasks facing the clinical microbiologist. Parallel testing requires collection of additional blood from each patient and may not be possible in some patients or institutions. The level of positivity for clinical pathogens (usually in the range of 8 to 11%) means that most of the specimens collected will not be of value in the comparison (30). In addition, the incidence of contamination (usually 1 to 3%) and predominance of a limited number of pathogens may result in an evaluation skewed toward only a few of the potentially clinically significant organisms. Thus, in-house verification should be designed to answer the following questions:

- Will the media used by the system support the growth of organisms (including yeasts, anaerobes, or fastidious organisms, where appropriate) commonly seen in the user's patient population?
- Will the instrument (for automated systems) detect, in a timely fashion, the majority of pathogenic organisms from blood cultures which contain these microorganisms?

Two approaches for verification of blood culture systems are discussed below. Laboratories may also choose to combine these approaches to take advantage of the strong points of each (e.g., perform parallel studies to assess the ability of the system to detect commonly isolated organisms and perform seeded blood cultures to assess less common pathogens).

Parallel blood culture studies

Performance of parallel blood cultures allows the laboratory to evaluate all aspects of the new system under actual patient and laboratory conditions. When a laboratory chooses to perform parallel studies of commercially available systems, duplicate sets of blood cultures inoculated with equivalent blood volumes should be obtained until one or more isolates of at least 20 different species of clinically significant organisms have been recovered. The new system may be considered verified if its sensitivity in detection of clinical pathogens is no lower than 5% below that of the reference method.

The method must be considered unverified if the sensitivity of the new system is more than 5% below that of the reference method. Corrective action should be taken by the user (and the manufacturer where appropriate), and the verification study should be repeated. The new system may then be considered verified if its sensitivity for detection of pathogens is no lower than 5% below that of the reference method.

Seeded blood culture studies

Assessment of the ability of the medium to support the growth of pathogens

1. Review past records of positive blood cultures to identify those organisms (a minimum of 20 different species) most prevalent in the patient population served by the institution. In addition, significant pathogens which are relatively uncommon or fastidious (e.g., *Neisseria gonorrhoeae* or *N. meningitidis*, *Cryptococcus neoformans*, *Campylobacter fetus*) should also be included. Whenever possible, actual patient isolates should be used rather than stock strains.

2. Make seeded blood cultures of isolates of each of the above species. To challenge the system, the minimum amount of sterile, antibiotic-free human blood recommended by the manufacturer should be placed in each bottle. In addition, the numbers of organisms placed in each bottle should approximate those found in cases of septicemia (which are often <1 CFU per ml of blood). This is best accomplished by making serial dilutions of the organisms prior to inoculation.

- Dilute each organism in sterile saline or broth to match a 0.5 McFarland standard.
- Make serial dilutions by transferring 0.1 ml of the suspension into 10 ml of saline or broth, mixing, and repeating this transfer two additional times (total of three transfers). This will result in a concentration of approximately 5 to 30 organisms per 0.1 ml in the final dilution tube.
- Inject 0.1 ml of the diluted organism suspension into each culture bottle. Also inoculate two plates of appropriate media with 0.1 ml from the same dilution tube and spread the inoculum for confirmation of the quantitation of the organisms injected into the bottle.

3. Examine the seeded bottles, following the manufacturer's instructions for the system.

4. The method is considered verified if all isolates were detected. Any problems with detection should be investigated by repeating the tests with the same patient strains. If detection is still not obtained, corrective action must be taken by the user and/or the manufacturer prior to instituting use of the system in the laboratory.

Assessment of the ability of an automated system to detect clinical isolates of organisms which have grown in the media (optional)

Once the initial verification of the system indicates that the system will support and allow detection of common isolates in a seeded trial, the system may be used for testing of patient samples. At this time, a concurrent verification may be done for automated systems to ensure that the instrument is detecting a minimum of 98% of organisms in an actual patient setting. This can be done by performing blind subcultures of all negative cultures prior to discard. For institutions with a 10% positivity rate, it will be necessary to subculture at least 500 bottles of each medium type to confirm that ≥98% of the positives have been detected.

VALIDATION OF DIAGNOSTIC TESTS USED IN CLINICAL MICROBIOLOGY

While verification of a new or revised test (commercially obtained or home brew) serves to establish that the performance parameters of the test are satisfactory, it does not provide ongoing assurance that the test is continually performing as expected under routine use over extended periods of time. Test validation is the ongoing process used by the laboratory to provide this assurance. Although validation as a specific process is not addressed in CLIA '88, the components of the process (quality control, proficiency testing, verification of employee competency, and instrument calibration) are all covered. The end result of validation will indicate one of three possibilities: (i) the test continues to be acceptable for routine use, (ii) further investigation is warranted, or (iii) immediate corrective action must be undertaken by the manufacturer (if commercially obtained), the user, or both, and the test must be considered unsuitable for continued routine use until it can be validated. Lot numbers and expiration dates should be documented for all reagents and materials used in the validation process. Records of validation procedures should be retained for at least two years.

Components of the Validation Process

The standard components of a validation process include the following.

1. **Quality control organisms.** Quality control organisms should be as stipulated by the manufacturer of a commercially available test or instrument or chosen by the user. If possible, these should be reference strains (American Type Culture Collection) or have some other recognized source; they should be maintained in a standard manner, so as not to be genetically affected by storage, passage, etc. If nonreference strains are used, the laboratory should have a complete record of the history of the organism, including characterization, storage, and recovery from storage. The frequency of testing and actions to be taken after control failures should follow the manufacturers' recommendations or those of one or more of the various regulatory or advisory (2, 17, 19, 21) agencies. For home brew tests, positive, negative, and other relevant controls should be selected from recommendations for the test found in references such as appropriate NCCLS guidelines, the current *Manual of Clinical Microbiology*

(15), or the *Clinical Microbiology Procedures Handbook* (28).

2. **Quality control analyte.** Quality control analytes are a metabolic product, nucleic acid, enzyme, or antigen usually provided by the manufacturer of a test or system for the routine quality control of a specific procedure or instrument. The analyte should be identified with a lot number or should have other traceable identification; further description should indicate concentration, titer (where appropriate), use, date of preparation, and storage conditions. For an analyte introduced by the user laboratory, a defined record of its development and assay should be available. The frequency of testing should follow the manufacturer's recommendations or those stipulated by one or more of the various regulatory or advisory (2, 17, 19, 21) agencies. For home brew tests, positive, negative, and other relevant controls should be selected from recommendations for the test found in appropriate NCCLS guidelines, the current *Manual of Clinical Microbiology* (15), or the *Clinical Microbiology Procedures Handbook* (28).

3. **Proficiency test (survey) samples.** Proficiency test samples are provided by CLIA '88-approved proficiency testing programs (College of American Pathologists, American Association of Bioanalysts, states, etc.) when a specific laboratory enrolls in a program provided by that agency. HCFA reviews approved proficiency testing programs on an annual basis. The frequency and rotation of the testing of various analytes are usually determined by the provider but have been approved by HCFA.

4. **Instrument calibration.** Certain instruments require that specific components or internal systems be checked on a regular basis. It is imperative that the manufacturer's directions for the calibration be carried out at the specified time intervals. This criterion may not apply to all tests or test systems.

5. **Use of historical data (for blood culture systems).** Laboratories are encouraged to utilize historical data concerning recovery of pathogens in their own populations as an aid in confirming that the system is operating as expected. If significant changes are seen in the distribution and/or frequency of recovery of isolates obtained from patients over a period of time (for example, a substantial reduction in the number of isolates of *Streptococcus pneumoniae* from that obtained with the old system during an equal time period), a more intensive investigation into the ability of the new system to support and detect these species is warranted.

6. **Other.** Other components of the validation process must include a determination and documentation of competency of testing personnel in performing the procedures and obtaining the correct results on an ongoing basis and documentation of all quality control, proficiency testing and calibration results, as well as any corrective action taken. Personnel competency is discussed in a separate section of this document.

When all these components are in place, the user has assurance that the test or test system meets the validation requirements.

Quality Control Reagents and/or Proficiency Test Samples Not Available

There are frequently occurring circumstances in which quality control reagents or proficiency test samples are not available or in which tests are of such a nature that standard analytes cannot be developed for proficiency test purposes. These situations include tests for unusual or rare analytes, tests for labile analytes, relatively new tests, tests that are hard to standardize, or tests that have low sensitivity or specificity. Some examples of these situations include isolation of *Haemophilus ducreyi* from specimens, serum bactericidal tests, or direct fluorescent-antibody tests for *Treponema pallidum.* In the absence of adequate quality control materials or proficiency test samples, such tests would fail the College of American Pathologists accreditation requirement for validation.

Several alternative approaches to satisfy the validation requirement include the following.

1. Split the patient sample and send a portion to a reference laboratory and compare the results. This would not have to be done with every sample, but it should be done at least twice a year. If the test volume is so low that even this approach is not possible, then the laboratory should reexamine whether the test should be offered at all, because of economics and personnel proficiency. It is advisable to inform the reference laboratory that the samples are to be used as a means of validating a test procedure. If this approach can be worked out mutually with another facility, the validation data would serve two laboratories.

2. Split samples may be tested as unknowns in the user's laboratory by ≥ 2 testing personnel. Again, this would only have to be done periodically, and the economics and personnel proficiency of the offering should be reevaluated if the test volume is low.

3. Save known positive and negative samples and prepare in-house quality control and/or unknown (proficiency testing) samples. This approach may work well for higher volume tests and those that have a relatively high degree of positivity, so that samples with various degrees of positivity may be utilized.

4. Obtain the analyte from an outside source and use it as a reference standard. This approach would work for tests or procedures for isolating unusual organisms that might be available from the American Type Culture Collection.

Frequency of Test Validation

Individual laboratories are responsible for ensuring that test validation occurs frequently enough to assure the continued performance of a laboratory test. In most cases, following the manufacturer's guidelines and/or the requirements of the regulatory or accrediting agencies will provide this assurance. For home brew tests, it may be necessary to perform test validation more frequently until it is determined how long the test will continue to perform satisfactorily.

PERSONNEL TRAINING AND DEMONSTRATION OF COMPETENCY

A new test may not be introduced into the laboratory until adequately trained personnel are available to perform the test and accurately report the results. In the microbiology section of the clinical laboratory, the ability to recognize and respond to unusual or critical organisms or results is crucial to good patient service. CLIA '88 states that the laboratory director must ensure that all personnel have the appropriate education and experience, receive the appropriate training for the type and complexity of the services offered, and demonstrate that they can perform all testing operations reliably to provide and report accurate results (7). The director must also ensure that policies and procedures for monitoring individuals who conduct preanalytical, analytical, and postanalytical phases of testing are established to ensure that they are competent and that they maintain their competency to process specimens, perform test procedures, report test results promptly and proficiently, and, whenever necessary, identify needs for remedial training or continuing education to improve skills. Finally, the responsibilities and duties of each person engaged in the performance of preanalytic, analytic, and postanalytic phases of testing must be specified in writing, identifying which procedures the individual is authorized to perform and whether supervision is required for specimen processing, test performance, or result reporting and whether director review is required prior to reporting patient test results.

Personnel competency for persons performing highly complex tests must be verified and documented at least semiannually during the first year the individual tests patient specimens and at least annually thereafter, unless there is a change in methodology or instrumentation. In the latter circumstance, the employee's performance using the changed procedure must be reverified prior to reporting patient test results. If corrective action is required, it should be completed and documented in a timely fashion (i.e., 30 days) (12).

Several resources are available to assist the laboratory in establishing and documenting a complete program of employee training, verification, and competency assessment (12, 25, 26). The basic components of this process as it pertains to incorporating a new procedure into the microbiology laboratory will be discussed in this section.

Writing the procedure(s)

The key to successful training and consistency of test performance is a well-written procedure or procedures which follow the guidelines of the NCCLS document GP2-A2 (18). One suggested method for preparing a procedure which lends itself to training and demonstration of competency is outlined in the NCCLS document GP21-A (25). In this document, the process of procedure writing hinges on identification of the critical steps of the procedure. These can be graphically outlined in a flow chart, and the flow chart can be used as a basis for writing the formal procedure and (if desired) a separate training procedure. Identifying the critical steps also aids in the development of learning objectives which focus on the key elements of the procedure.

Preparing a training document

In order to assess the ability of testing personnel to successfully perform a new procedure, it is necessary for the individual to clearly understand what is expected of him or her when performing each of the steps of the procedure. Thus, a training document differs from a procedure in several significant ways. First, it needs to clearly outline what outcomes are expected once the individual has successfully completed training. These outcomes need to be stated as specific, measurable learning objectives. Second, the training methods and required materials to be used during the training process need to be detailed. Training methods might include reading the procedure, reading background material, viewing a training videotape or working through a computer-based training program, observing the procedure being performed, and performing the procedure. Third, the training document needs to state what measurement tool or tools will be used to document as objectively as possible that the learning objectives have been met. Suggested tools in the NCCLS GP21-A document (25) include:

- testing of blind specimens, proficiency test specimens, or previously analyzed specimens
- administration of a written test
- observation of procedure, process, and outcome
- assessment of response to case studies, problems, or situations related to the procedure
- documentation of response to actual incidents which may have occurred during the performance of the procedure ("critical incidents")

- assessment of response to oral queries related to the procedure

Finally, if desired, the training document can include an additional section which describes what the trained employee will need to do to demonstrate ongoing competency in the performance of the procedure.

One approach to development of a training document would be to include the training information listed above, as well as a specific training procedure (the complete technical procedure modified to include explanatory training notes and information which will be useful as aids to training). This training document would not require the individual to refer to the standard procedure manual during the training process. If this approach is used, it requires extra diligence to ensure that the training procedure is updated and modified whenever changes are made to the standard laboratory procedure.

A second approach to the training document would be to use the standard (unmodified) laboratory procedure when training and to limit the training document to the specific training information alone. This approach has the advantage of eliminating the effort required to ensure that the training procedure and the standard procedure are both modified consistently, and it ensures that the trainee will be using the actual laboratory procedure when training. It has the disadvantage of not allowing inclusion of explanatory notes or instructions which may not be part of the standard procedure. An example of this type of document is presented in Appendix C.

Documentation of training

Documentation of training can be accomplished with a variety of written or computer-based documentation systems. Specific examples of possible checklists or documentation forms are included in the appendices of the NCCLS GP21-A guideline (25), and these could be modified to suit the needs of the laboratory. Procedures may be combined on one document to reduce the amount of paperwork, or individual documents for each employee and procedure may be maintained. The American Society of Clinical Pathologists has developed a computer-based documentation program (ASCP Comptec) (1) which may be purchased and used to document training and ongoing competency. In addition, many commonly used commercial database programs may be used to design individual laboratory-specific computer-based documentation programs. All documentation should be retained for a minimum of two years.

SUMMATION

This document provides guidance in performing test verification and validation in the microbiology laboratory. As stated at the outset, these are guidelines and should not be considered regulatory standards. For those responsible for establishing and maintaining standards in clinical laboratories, there are many excellent documents available, many of which are cited throughout this text. Our goal here is to make this information more "microbiology friendly." Ensuring good laboratory practice, which includes complying with regulations from various agencies, can be a challenge. The availability of clear and useful guidelines for performing verification and validation which specifically address clinical microbiology should make the accomplishment of this aspect of good laboratory practice easier to achieve.

APPENDIX A. METHOD SELECTION EXAMPLE

I am considering a methodology change in my laboratory. I hope to substitute a DNA probe for *Neisseria gonorrhoeae* (GC) in place of my standard GC culture. I will consider the culture the reference test in this situation. I know that about 10% of my GC cultures are positive. How do I go about verifying the method and seeing if it will satisfy the needs of my clinicians? I begin by asking my Obstetrics and Gynecology (OB-GYN) department to send in duplicate samples. I obtain 100 samples, perform both tests, and get the following results:

	Culture results	
New test results	Positive	Negative
Positive	9	2
Negative	1	88

From these data, I can calculate the sensitivity, specificity, and positive and negative predictive values of the new test:

$$\text{Sensitivity} = [9/(9 + 1)] \times 100 = 90\%$$
$$\text{Specificity} = [88/(88 + 2)] \times 100 = 97.7\%$$
$$\text{PPV} = [9/(9 + 2)] \times 100 = 81.8\%$$
$$\text{NPV} = [88/(88 + 1)] \times 100 = 98.9\%$$

I now evaluate the validity of these results. Have I tested enough samples to feel confident in the generated results?

With a prevalence of 10%, I will only expect around 10 positives out of my sample. Thus, the closest I am likely to be when calculating sensitivity is within 10%. In other words, if I tested 100 samples and did not get any false negatives, my sensitivity would be 100%. If the test failed to detect 1 positive, the sensitivity would calculate to 90%. If it failed to detect 2 positives, the sensitivity would be 80%. This is not a very discerning analysis.

How many samples should I test? If I stack the deck and save samples and then perform the new test on 100 patients with positive cultures and 100

patients with negative cultures, I will be able to calculate sensitivity and specificity to the nearest 1%; if I test 50 of each, I can calculate to the nearest 2%. How exact I want or need to be when I evaluate the test will depend in large part on how critical the result is: whether it is to be used for diagnosis or screening, whether it stands alone or can be validated by other laboratory data, and how well the test has been previously evaluated in laboratories and patient populations similar to my own.

In this case, assume that I save samples and test 100 patients with positive GC cultures and 100 patients with negative cultures and get the following results:

	Culture results	
New test results	Positive	Negative
Positive	97	1
Negative	3	99

Now I calculate the sensitivity and specificity.

Sensitivity is 97% and specificity is 99%. PPV and NPV are 99% and 97%, respectively, but they are based on a prevalence of 50% because of the way the study was designed.

Next, I find out that my OB-GYN physicians are adamant that they must have a high level of confidence in my positive results, because they have recently been sued by a patient for pain and suffering who was told that her test was positive when it was not. These physicians do mostly prenatal screening of a low-prevalence population (only 1% of their patients are estimated to have GC). They would like a positive predictive value of 95% for their population. Can I offer this test to them? To make the math easy, assume that I can test 10,000 samples in my evaluation; then, on the basis of the known sensitivity (97%) and specificity (99%) of this test, I fill out the expected results in the table.

	Culture results	
New test results	Positive	Negative
Positive	97	99
Negative	3	9801

I then calculate the PPV and NPV in this case.

On the basis of the new prevalence, the PPV would be 49% and the NPV would be 99.96%. I could feel confident about a negative result, but the positives are much too uncertain (in fact, I could flip a coin and get as accurate a result).

So, what kind of sensitivity and specificity would the test have to have at this low prevalence to satisfy these clinicians?

Sensitivity will not affect PPV; specificity will. So, assuming I am testing the same 10,000 samples, 97 will still be true positives (TP) and 3 will still be false negatives. Now I can calculate how many false positives (FP) I can have to achieve my desired PPV:

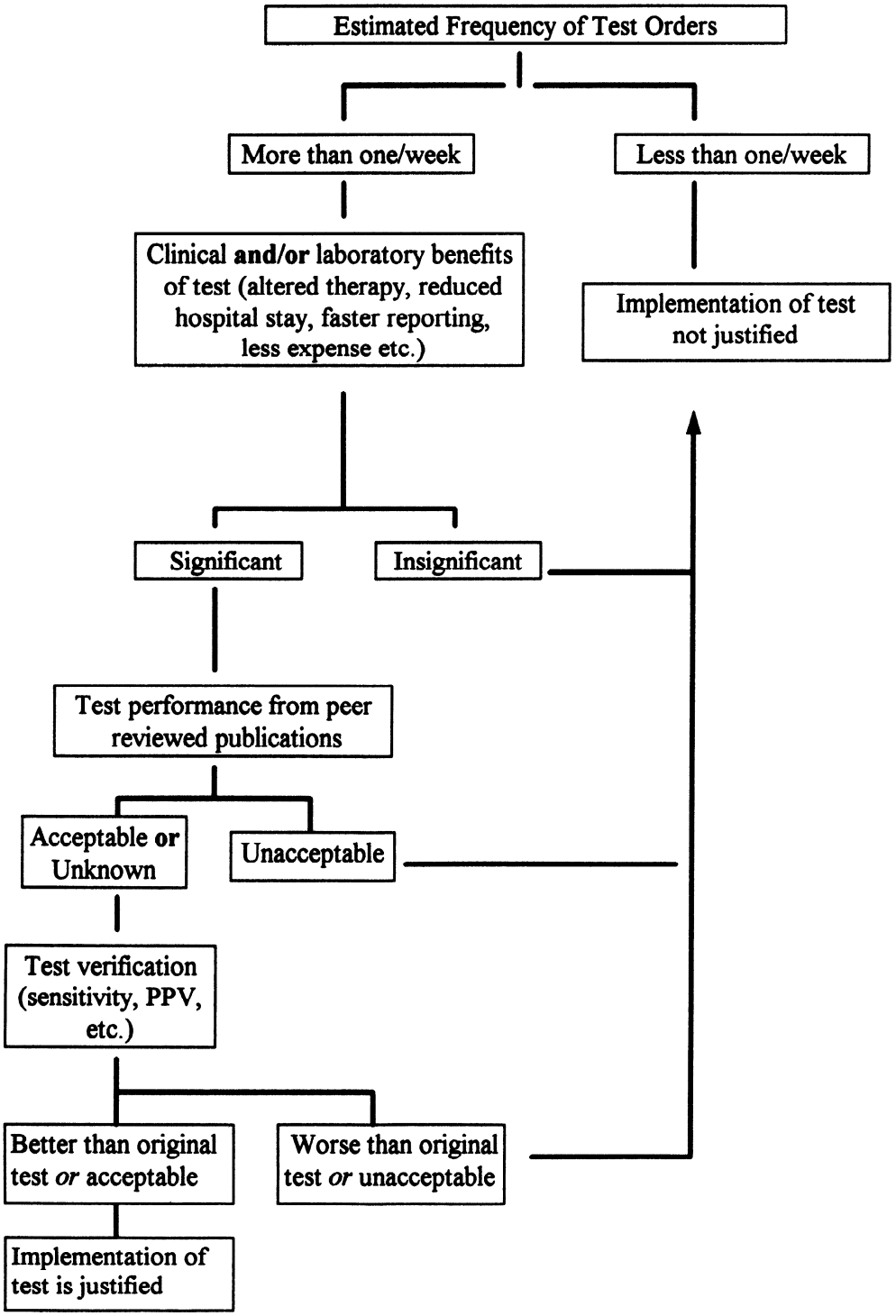
$$95 = [TP/(TP + FP)] \times 100$$
$$95 = [97/(97 + x)] \times 100$$
$$x = 5$$

Thus, of the 10,000 samples, I can have only 5 false positives to achieve this goal, which means of my 9,906 negatives tested, 9895 will have to give a negative result with our new test. Thus, my table will look like this:

	Culture results	
New test results	Positive	Negative
Positive	97	5
Negative	3	9895

Now I can calculate the new specificity required by the clinician by using the data from the table I just filled out. Is the specificity for this test high enough? Statistically, can I even answer this question?

When I calculate the specificity, I see that at this prevalence, to achieve the physician's desired PPV, I need a test with a specificity of 99.95%. Since I only know the new test's specificity to the nearest 1% (based on the 100 positives and negatives I originally tested) I cannot be sure that it will satisfy the clinician's needs (although by common sense, it probably will not!). To know for sure, I would have to be able to detect 5 of 10,000 (or 1 out of 2,000) false-positive specimens. Statistically, this means I would have to test approximately 6,000 negative specimens to be able to assure the clinician with 95% confidence that this test would perform to the desired level of specificity. It is probably better to do culture!


Estimated Frequency of Test Orders
More than one/week
Less than one/week
Clinical and/or laboratory benefits of test (altered therapy, reduced hospital stay, faster reporting, less expense etc.)
Implementation of test not justified
Significant
Insignificant
Test performance from peer reviewed publications
Acceptable or Unknown
Unacceptable
Test verification (sensitivity, PPV, etc.)
Better than original test or acceptable
Worse than original test or unacceptable
Implementation of test is justified

APPENDIX C. SAMPLE OF LABORATORY TRAINING DOCUMENT

Procedure: Urine Culture

Expected Outcome of Training: The trainee will be able to:

1. Describe the various methods used to collect specimens for urine culture and the advantages and disadvantages of each
2. List the requirements for specimen storage and transport to prevent bacterial growth in the specimen, evaluate individual specimens for acceptability on the basis of the criteria, and reject or accept them correctly following laboratory procedure
3. Describe the procedures used to process specimens (clean catch, catheterized, suprapubic aspirate, etc.) for quantitative urine culture
4. After incubation, calculate the original specimen concentration of each of the organisms present in a culture
5. List potential urinary tract pathogens and recognize which organisms most commonly represent urethral or vaginal contamination of the specimen
6. Recognize the morphology of potential pathogens and nonpathogens on blood agar and MacConkey agar
7. Correctly identify which organisms require full identification and susceptibility testing on the basis of their quantities and the presence of other organisms in the culture

Documents and References To Be Used during Training

1. Laboratory Specimen Collection Manual
2. Bacteriology Procedure Manual
3. Reference texts (e.g., *Manual of Clinical Microbiology*, Bailey and Scott, etc.)

Training Methods To Be Used

1. Reading of procedure in manual
2. Supervised set-up, interpretation, workup, and reporting of urine cultures

Evaluation Criteria for Successful Training*

1. When presented with case histories describing collection and storage of urine specimens, the trainee will correctly identify 100% of the time (i) improper procedures, (ii) the likely effect these will have had on the culture results, and (iii) the correct laboratory procedure to follow in acceptance or rejection of the specimen.
2. The trainee will correctly read, interpret, work up, and report at least 100 urine cultures (minimum of 50 positive specimens) without error or deviation from laboratory protocol.

Evaluation Criteria for Ongoing Verification of Competency*

1. Ongoing supervisor review of technologist worksheets of positive cultures indicates that the individual correctly selects appropriate identification and susceptibility tests (less than two errors detected per year).
2. Technologist successfully (95%) identifies pathogenic and nonpathogenic flora on colonial morphology practical exam.**
3. Technologist successfully (95%) identifies potential pathogens and correct workups on a variety of cultures (contaminated, noncontaminated, multiple pathogens, etc.) in a written practical exam.**
4. Technologist correctly identifies potential problems in collection and transport of specimens when presented with case histories as part of a written practical exam.**

*Note: all evaluations are "open book"; any written resources normally available to the employee for performance of the job may be used when demonstrating competency.

**Note: the colonial morphology exam and written exam may be used to evaluate multiple procedures at one time.

REFERENCES

1. **American Society of Clinical Pathologists.** 1994. *ASCP Comptec: the Computerized Ongoing Monitoring Program to Evaluate Competency.* American Society of Clinical Pathologists, Chicago, Ill.
2. **College of American Pathologists.** 1994. Section 1, p. 2, and section 4, p. 16–21. *In Commission on Laboratory Accreditation Inspection Checklist.* College of American Pathologists, Northfield, Ill.
3. **Doern, G.V., R. Vautour, M. Gauder, and B. Levy.** 1994. Clinical impact of rapid in vitro susceptibility testing and bacterial identification. *J. Clin. Microbiol.* **32:**1757–1762.
4. **Federal Register.** 1995. Medical device user facilities and manufacturers; adverse events, reporting requirements, certification, and registration. *Fed. Regist.* **60:**63578–63606.
5. **Ferraro, M. J., and J. H. Jorgensen.** 1995. Instrument-based antibacterial susceptibility testing, p. 1379–1384. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), *Manual of Clinical Microbiology,* 6th ed. American Society for Microbiology, Washington, D.C.
6. **Hadgu, A.** 1996. The discrepancy in discrepant analysis. *Lancet* **348:**592–593.
7. **Health Care Financing Administration.** 1992. Medicare, Medicaid, and CLIA programs. Regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). *Fed. Regist.* **57:**7002–7186.
8. **Ilstrup, D.** 1990. Statistical methods in microbiology. *Clin. Microbiol. Rev.* **3:**219–226.
9. **Jenkins, S. G.** 1995. Evaluation of new technology in the clinical microbiology laboratory. *Diagn. Microbiol. Infect. Dis.* **23:**53–60.
10. **Jorgensen, J. H.** 1993. Selection criteria for an antimicrobial susceptibility testing system. *J. Clin. Microbiol.* **31:**2841–2844.
11. **LeBar, W. D.** 1996. Keeping up with new technology: new approaches to diagnosis of Chlamydia infection. *Clin. Chem.* **42:**809–812.
12. **McCaskey, L., and M. LaRocco.** 1995. Competency testing in clinical microbiology. *Lab. Med.* **26:**343–349.
13. **Miller, J. M.** 1991. Evaluating biochemical identification systems. *J. Clin. Microbiol.* **29:**1559–1561.
14. **Miller, J. M., and C. M. O'Hara.** 1995. Substrate utilization systems for the identification of bacteria and yeasts, p. 103–109. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C.

Tenover, and R. H. Yolken (ed.), *Manual of Clinical Microbiology,* 6th ed. American Society for Microbiology, Washington, D.C.

15. **Murray, P. R., E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.).** 1995. *Manual of Clinical Microbiology,* 6th ed. American Society for Microbiology, Washington, D.C.
16. **Murray, P. R., A. C. Niles, and R. L. Heeren.** 1987. Comparison of a highly automated 5-h susceptibility testing system, the Cobas-Bact, with two reference methods: Kirby-Bauer disk diffusion and broth microdilution. *J. Clin. Microbiol.* **25**:2372–2377.
17. **National Committee for Clinical Laboratory Standards.** 1996. *Quality Assurance for Commercially Prepared Microbiological Culture Media.* Approved standard M22-A2. National Committee for Clinical Laboratory Standards, Wayne, Pa.
18. **National Committee for Clinical Laboratory Standards.** 1992. *Clinical Laboratory Technical Procedure Manuals,* 2nd ed. Approved guideline GP2-A2. National Committee for Clinical Laboratory Standards, Wayne, Pa.
19. **National Committee for Clinical Laboratory Standards.** 1993. *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically,* 3rd ed. Approved standard M7-A3. National Committee for Clinical Laboratory Standards, Wayne, Pa.
20. **National Committee for Clinical Laboratory Standards.** 1993. *Nomenclature and Definitions for Use in NRSCL and Other NCCLS Documents,* 2nd ed. Proposed guideline NRSCL 8-P2. National Committee for Clinical Laboratory Standards, Wayne, Pa.
21. **National Committee for Clinical Laboratory Standards.** 1993. *Performance Standards for Antimicrobial Disk Susceptibility Tests,* 5th ed. Approved standard M2-A5. National Committee for Clinical Laboratory Standards, Wayne, Pa.
22. **National Committee for Clinical Laboratory Standards.** 1994. *Development of In Vitro Susceptibility Testing Criteria and Quality Control Parameters.* Approved guideline M23-A. National Committee for Clinical Laboratory Standards, Wayne, Pa.
23. **National Committee for Clinical Laboratory Standards.** 1994. *Specifications for Immunological Testing for Infectious Diseases.* Approved guideline I/LA 18-A. National Committee for Clinical Laboratory Standards, Wayne, Pa.
24. **National Committee for Clinical Laboratory Standards.** 1995. *Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots.* Approved guideline GP10-A. National Committee for Clinical Laboratory Standards, Wayne, Pa.
25. **National Committee for Clinical Laboratory Standards.** 1995. *Training Verification of Laboratory Personnel.* Approved guideline GP21-A. National Committee for Clinical Laboratory Standards, Wayne, Pa.
26. **Nevalainen, D. E., and L. M. Berte.** 1993. *Training, Verification and Assessment: Keys to Quality Management.* Clinical Laboratory Management Association, Paoli, Pa.
27. **Radetsky, M., and J. K. Todd.** 1984. Criteria for the evaluation of new diagnostic tests. *Pediatr. Infect. Dis.* **3**:461–466.
28. **Sewell, D. L. (ed.).** 1992. Quality assurance, quality control, laboratory records, and water quality, p. 13.0.1-13.4.11. *In* H. D. Isenberg (ed.), *Clinical Microbiology Procedures Handbook,* vol. 2. American Society for Microbiology, Washington, D.C.
29. **Sewell, D. L., and R. B. Schifman.** 1995. Quality assurance: quality improvement, quality control, and test validation, p. 55–66. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), *Manual of Clinical Microbiology,* 6th ed. American Society for Microbiology, Washington, D.C.
30. **Wilson, M. L.** 1994. Blood Cultures—Introduction. *Clin. Lab. Med.* **14**:1–7.