Criteria for Validation of Methods in Microbial Forensics

Bruce Budowle, Steven E. Schutzer, Stephen A. Morse, Kenneth F. Martinez, Ranajit Chakraborty, Babetta L. Marrone, Sharon L. Messenger, Randall S. Murch, Paul J. Jackson, Phillip Williamson, Rockne Harmon and Stephan P. Velsko

Appl. Environ. Microbiol. **2008, 74(18):5599. DOI: 10.1128/AEM.00966-08.**
Published Ahead of Print 25 July 2008.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Sept. 2008, p. 5599–5607 Vol. 74, No. 18
0099-2240/08/$08.00+0 doi:10.1128/AEM.00966-08

MINIREVIEW

Criteria for Validation of Methods in Microbial Forensics▽

Bruce Budowle,[1] Steven E. Schutzer,[2]* Stephen A. Morse,[3] Kenneth F. Martinez,[4] Ranajit Chakraborty,[5] Babetta L. Marrone,[6] Sharon L. Messenger,[7] Randall S. Murch,[8] Paul J. Jackson,[9] Phillip Williamson,[10] Rockne Harmon,[11] and Stephan P. Velsko[9]

Federal Bureau of Investigation Laboratory, Quantico, Virginia[1]; UMDNJ-New Jersey Medical School, Newark, New Jersey 07103[2]; Centers for Disease Control and Prevention, Atlanta, Georgia[3]; Centers for Disease Control and Prevention, Cincinnati, Ohio[4]; University of Cincinnati, Cincinnati, Ohio[5]; Los Alamos National Laboratory, Los Alamos, New Mexico[6]; California Department of Public Health, Richmond, California[7]; Virginia Polytechnic Institute and State University, Alexandria, Virginia[8]; Lawrence Livermore National Laboratory, Livermore, California[9]; University of North Texas Health Science Center, Fort Worth, Texas[10]; and Office of the District Attorney, Alameda County, Oakland, California[11]

A process for validation is essential in the development of methods that microbial forensics uses to generate reliable and defensible results. Law enforcement investigators need to respond quickly to the best leads to counter ever-increasing threats and will rely upon results generated from the analyses of any microbial forensic evidence to attempt to attribute any attack to a person(s) or group. Readily available technology and knowledge are making it easier for an individual or group to carry out biocrimes or bioterrorism using microorganisms and toxins as weapons. The potential that a biological weapon will be used is of serious concern for the safety and security of people and critical infrastructure. If a biocrime is committed, microbial forensic evidence will be sought, collected, and characterized to help investigators identify the perpetrator(s) and exclude innocent suspects. Analyses of collected material are often challenging because the identification of the signatures most useful for attribution often requires substantial effort (3). In addition, some microbial forensic specimens can be limited in quantity and/or quality. Despite these demands, accurate and credible results are needed because the interpretation of such results might seriously impact the course or focus of an investigation, thus affecting the liberties of individuals, or even be used as a justification for a government's military response to an attack. Therefore, the methods for the collection, extraction, and analysis of microbial evidence that could generate key results need to be as scientifically robust as possible so that they are defensible to the legal community (12, 21) and, perhaps, to the international government, law enforcement, and scientific communities. Proper interpretation of the results of microbial forensic analysis relies substantially on understanding the performance and limitations of the methods of collection and the analytical processes, assays, and interpretation involved. Failing to properly validate a method or misinterpreting the results from a microbial forensic analysis or process may have severe consequences.

DEFINING VALIDATION

Validation is frequently used to connote confidence in a test or process. However, frequently, the process of validation is not well defined or properly described in context. Not being explicit about what is meant by validation can result in misinterpretation and misapplication of a properly performed test. It also can lead to a false sense of confidence in a poor method. In the nascent field of microbial forensics (5), there is a need to better describe what constitutes validation. A strict delineation of the steps needed to validate a method or process may be too restrictive; there are a myriad of methods, processes, targets, platforms, and applications. Yet some basic requirements transcend individual differences in methods, and these can be reinforced by contextual description and illustrated with examples. Failing to validate a method or misinterpreting the reliability of a method in a microbial forensic analysis can have dire consequences. This paper provides a framework for developing a validation plan that can be useful for microbial forensics and may have application to other scientific fields where "validation" may be used colloquially.

OBJECTIVE PERFORMANCE CRITERIA AND CATEGORIES OF VALIDATION

Objective performance data are essential for establishing confidence in assays and processes, by defining the inherent quality of a method or process and by demonstrating its applicability for a designated purpose (14, 20). In this context, validation is described as the process that (i) assesses the ability of procedures to obtain reliable results under defined conditions, (ii) rigorously defines the conditions that are required to obtain the results, (iii) determines the limitations of the procedures, (iv) identifies aspects of the analysis that must be monitored and controlled, and (v) forms the basis for the development of interpretation guidelines to convey the significance of the findings. The Quality Assurance Guidelines for microbial forensics addressed the basic need for validation (5). The basic categories of validation have been described in section 8 of those guidelines (Table 1) (5). (See also references 5, 7, and 8 and, for a glossary of terms, the appendix.)

* Corresponding author. Mailing address: UMDNJ-New Jersey Medical School, Newark, NJ 07103. E-mail: schutzer@umdnj.edu.
▽ Published ahead of print on 25 July 2008.

TABLE 1. Elements of validation guidelines for methods and procedures for analyses

Element(s) of validation guidelines or procedure
Developmental validation should be appropriately documented and should address specificity, sensitivity, reproducibility, bias, precision, false positives, and false negatives. Appropriate controls should be determined. Any reference database used should be documented.
Preliminary validation is the acquisition of limited test data to enable an evaluation of a method used to provide investigative support to investigate a biocrime or bioterrorism event. If the results are to be used for other than investigative support, then a panel of peer experts, external to the laboratory, should be convened to assess the utility of the method and to define the limits of interpretation and conclusions drawn.
Internal validation should be performed and documented by the laboratory.
The procedure should be tested using known samples. The laboratory should monitor and document its reproducibility and precision and define reportable ranges of the procedure using a control(s).
Before the introduction of a new procedure into sample analysis, the analyst or examination team should successfully complete a qualifying test for that procedure.
Material modifications made to analytical procedures should be documented and subjected to validation testing commensurate with the modification and have documented approval.

More-specific criteria are presented in this paper for developing a validation plan for microbial forensic methods and procedures. The basic categories of validation are (i) developmental validation, (ii) internal validation, and (iii) preliminary validation. Developmental validation is the acquisition of test data and the determination of conditions and limitations of a newly developed method for analyzing samples. While it is sometimes conceptually convenient to treat the development of an assay and validation as separate issues, in practice, the development and validation processes are intimately intertwined. They should be considered together early in the development process. Once a method or process has been developed and initially validated, it may be transferred to an operational laboratory (or into field operations) for implementation. Internal validation is required in this new operational setting. Internal validation is an accumulation of test data within an operational laboratory to demonstrate that established methods and procedures are carried out within predetermined limits in the laboratory.

A developed and implemented standard operating protocol or procedure (SOP) readily available for all scenarios is not likely. There is a large variety of possible biocrime and bioterrorism scenarios for which a previously validated method may not always be available. Exigency may demand a tool or method (for collection, extraction, or analysis) that has not been validated previously. Clearly, it is irresponsible to wait for months or years for the validation of a procedure when an attack is under way and safety or security is imminently threatened. Therefore, it can be anticipated that some methods of generating investigative leads may not have been through as extensive a validation process as might be carried out when attention is paid to developmental and internal-validation requirements. Consider the scenario where an innocuous bacterial species has been used in an attack and it has been engineered such that it is highly virulent in humans. Because of the lack of interest in the particular species prior to the attack, there might be no fully validated strain identification method for this species in public health or forensic laboratories. A research laboratory protocol might be sought because of the imminent need to respond. However, even in this case, quality assurance and control considerations and understanding the limitations of a method should not be overlooked. In this context, preliminary validation is an early evaluation of a method that will be used to investigate a biocrime or bioterrorism event. A preliminary validation should be carried out to acquire limited test data to enable the evaluation of a method for its investigative-lead value, with the intent of identifying key parameters and operating conditions and of establishing a degree of confidence in the methods of collection, extraction, and analysis. The evaluation may be based on peer review of existing data by a panel of experts that makes recommendations for additional evaluations, studies that may be needed prior to the processing of evidentiary material, or studies that may be carried out after certain results are obtained. The goal is to be able to respond to a biocrime expeditiously, effectively, and efficiently while maintaining scientifically valid and legally defensible approaches. Preliminary validation is expected to be a normal occurrence in those situations where methods have not been through external and internal validation but are deemed necessary to support an investigation (criteria for preliminary validation are under development).

As stressed above, often the basic concept of validation is appreciated; yet the criteria for validation are seldom consistently defined or described and applied. It is common for the term validation to be used vaguely or to remain undefined when applied to a process performance evaluation. The degree of validation varies from nominal to rigorous. The consequences of such varied requirements can be catastrophic if methods used in microbial forensic investigations are poorly constructed, are underdeveloped, and/or generate results that are difficult to interpret. To avoid these shortcomings, it is necessary to develop a set of validation criteria and a validation plan so that procedures are subjected to rigorous evaluation at an acceptable level of quality based on the context in which the procedures may be applied.

CONSTRUCTING A VALIDATION PLAN

Preparation of a "validation plan" begins by defining the criteria that will be used to evaluate the performance of a method. These delineated criteria or parameters for validating a process will guide those who may develop and/or implement a new method and provide a record of what was, and what was not, addressed during validation. However, the generation of a universal list of criteria for all possible methods from evidence collection, preservation, transport, extraction, analysis, and interpretation is not likely to be achieved. There are a multitude of diverse processes that may be employed and myriad targets to be assessed. One set of criteria will never apply a priori to all methods, procedures, and processes (4). Two primary and overarching criteria are reliability and reproducibility. Some criteria, such as specificity, sensitivity, accuracy, and precision, apply to most analytical methods (defined below). However, more-relevant criteria are required for collection tools and

TABLE 2. Elements of a validation plan

Validation plan element(s)
Description of the scope, purpose, and intended application of the method
Description of the performance parameters of the method that will be tested and the data that will be accumulated
Plan for testing those parameters, including: Critical reagents and equipment needed Standard reference materials needed, including positive and negative controls and the ranges of those conditions to be evaluated Validation of test samples in a no. of replicates sufficient to demonstrate reproducibility and reliability Conditions and ranges of conditions to be tested Any aspects unique to the system that require additional specific testing
End results to be assessed, any statistical analysis that will be performed, and criteria for interpretation
Any special personal protective equipment, safety equipment, or safety practices required to perform the method safely
Requirements for personnel performing the method, such as training and competency
Interpretation of results of the analysis according to the interpretation criteria developed above (see the fourth item above)

methods concerning recovery, stability, and yield. Furthermore, collection may be designed intentionally to be generic or specific for a particular target. A different set of criteria is required for the interpretation of the results of the analysis.

Even though it is not possible to provide a specific validation template for the innumerable methods that could assist in a microbial forensic investigation, a checklist of parameters may establish minimal acceptable validation criteria. A researcher and developer can review the checklist and determine which criteria apply and also provide legitimate justification for why some criteria are not applicable. Those who rely on the use of the tool or results obtained will be able to assess the performance in proper context based on the validation criteria used. Those who validate a system should document the tested parameters so that they may be reviewed and assessed by interested parties. The following recommendations should provide a basis within the microbial forensics community for validating methods and processes and for fostering discussion and input for developing the best validation practices as they evolve or for validating new methods as they are developed. Since the process of validation attempts to define the limitations and test the fundamental assumptions of the method being evaluated, the experimental validation design should accumulate performance data on each of the method parameters to enable proper inferences based on the results of the analysis. Thus, a validation plan should define the range of conditions under which the process may be applied so that (i) the interpretation of the analytical results is effective and useful and, equally important, (ii) the conditions under which the results or the standard interpretation is not effective or reliable are understood. The minimal criteria are shown in Table 2 (these criteria are similar to those used to validate methods for use in clinical laboratories [6, 9, 10]).

As seen in Table 2, the first step in the validation process is to define the scope, purpose, and application of the method. This important point sets the direction for validating the method, process, or technology that will follow. Not all criteria will apply to all methods, nor will each criterion apply in the same manner or to the same level of stringency among applications. For example, in clinical medicine, it is acceptable to have a screening test with high false-positive rates (e.g., nontreponemal screening test for syphilis) in order to identify people who would benefit from a more specific evaluation (e.g., a confirmatory treponemal test for syphilis).

A validation plan also needs to address a range of possible relevant conditions that will be evaluated by assessing both reference and, at minimum, mock forensic samples. The samples should cover the relevant domain represented by the population for which the assay is to be validated. The test conditions should span a range of performance criteria. This range of criteria should include the conditions used by the anticipated SOP and indicate where the process begins to and will fail (for example, when the amount of material tested falls below the level of sensitivity of detection).

Once the criteria of a validation plan are satisfied, the procedure itself should be documented by preparing an SOP. Just as documenting the collection and the handling of evidence memorializes the history of an evidentiary sample from initial identification through analysis, a proper SOP should be written with the same detail in mind. This SOP may encompass all processes from collection to data interpretation or portions of the entire process. The SOP is a precise set of instructions for carrying out a process or procedure. It should contain sufficient detail about the procedure so that anyone trained in that field could carry out the assay, and it should include, if appropriate, the following elements: a delineation of each step of the procedure, a list of proper controls (positive, negative, and/or internal), a list of all reagents and descriptions of how to prepare them, calibration requirements for equipment and tests, criteria for the analysis of results, criteria for the interpretation of results, a list of personnel requirements, reporting criteria, and a list of appropriate literature references that support the fundamental theory or established scientific basis of a method. Such data are important for establishing foundations that support the validity of the basic science or the specific method.

PROCESSES OF MICROBIAL FORENSICS THAT REQUIRE VALIDATION

As one moves from the general validation criteria described above to those that are more specific to the situation, it is useful to divide the processes for microbial forensics that may require validation into four categories: (i) sample collection and preservation, (ii) extraction, (iii) analysis, and (iv) interpretation of results. Different validation criteria necessarily apply to the various processes. Each category and their potential checklist criteria are described below.

(i) Collection and preservation. The proper collection and preservation of microbial forensic evidence are crucial to a successful investigation leading toward attribution (4). If exist-

TABLE 3. Checklist of elements of collection and preservation criteria

Element(s) of collection and preservation criteria
Target, namely, the microorganism, toxin, or analyte (e.g., such as DNA, RNA, protein, and elements)
Sample integrity and the maintenance of its true state or condition, such as viability
Sample stability and the maintenance of its true state or condition over time, particularly during transport or storage
Recovery, specifically, the efficiency of collection/recovery of the target from the substrate or from surfaces (especially when the entire specimen is not collected and sent to the laboratory)
An understanding of the influence of the sample matrix and conditions, for example, wet samples, dry samples, or frozen samples, and unwanted materials that may be collected along with the target
Compatibility of the collection device or material, whether the material is inert or interacts with the target, with subsequent signature extraction and analysis methods; a description of any special procedures needed for the efficient recovery of the target from the collection device
Packaging and storage strategies that maintain the integrity and stability of the target and comply with applicable safety and transportation regulations
Personnel requirements, namely, the roles of personnel (e.g., in a 2-person method) and their training, proficiency, and safety

ing evidence is collected inefficiently, degrades, or is contaminated during collection, handling, transport, or storage, the subsequent characterization and attribution analyses may be compromised. Retrieving sufficient quantities and maintaining the integrity of the evidence increase the chances of characterizing the material to obtain the highest possible level of attribution. Thus, methods involved in this aspect of microbial forensics should be validated, as well as is possible, given that each crime scene is unique and all scenarios are not predictable or completely definable. In many undefined situations, it may be more appropriate to develop and validate more-generic approaches. Table 3 illustrates minimum elements to consider for collection and preservation criteria.

The context and how a test may be used are important. Detecting the presence of a microorganism or toxin has a purpose different from trying to prove that it is absent. Recently, the American Society of Tests and Measurements (ASTM) International published a standard protocol for the collection of powders suspected of being biological agents. This protocol was developed in collaboration with participants from multiple federal agencies (17). It is the first validated, standard microbial forensics method developed within the context of the environmental response and recovery efforts following the anthrax mail attacks in 2001. In the aftermath of the anthrax mail attacks (1), and in a context different from an epidemiologic or forensic investigation, national attention focused on the effectiveness of remediation in buildings and specific sources of contamination that may affect individual exposure. A Government Accountability Office audit and subsequent report criticized the lack of validated sampling methods available for assessing remediation following the anthrax attacks (23). The ASTM collection method was developed to address one aspect of the response to and collection of bulk and swab samples from nonporous surfaces visibly coated with powders suspected of being biological agents, illustrating that the same analysis may be applied to different contexts. One application of the analysis attempts to determine the microorganism present at a site that is the cause of a disease or constitutes a weapon. The other application of the analysis is to determine that the microorganism or toxin is not present at the site (or at least no longer active) to evaluate whether it is safe to reenter the site. Each contextual use requires different sampling strategies. The degrees of confidence in a result will differ between these applications.

(ii) Extraction of target for analysis. The extraction of the target from a collected sample may not be necessary before direct analysis. Examples of this are the collection of bacteria from surfaces by using replicate organism direct agar contact plates for culture and the direct collection of visible powder for microscopy (19). In contrast, an extraction process is often necessary in microbial forensics analysis as a sample-processing step between collection and analysis. It may be necessary to extract the target from the collection device (swab, wipe, or filter, etc.), purify the target from the environmental matrix or substrate before analysis, and/or remove inhibitors that may affect subsequent analyses.

Furthermore, genetic analysis of organisms or trace analysis of nucleic acid remnants from some toxin preparations may require the adaptation of a sample-processing procedure. A wide variety of options exist for extracting nucleic acids for analysis. To validate new nucleic acid extraction procedures or modifications of validated procedures for application to new matrices or substrates, researchers may need to consider the nature of the sample being extracted and the potential adverse impact of contaminants in the sample on later analysis. For example, soil samples often contain humic and fulvic acids that can inhibit the PCR (16, 25). Failure to address this effect may lead to false-negative results that might lead to an incorrect interpretation that the target was not present at some level of detection. Signature extraction is a destructive process. Therefore, undue consumption of precious evidence may occur if the extraction process is not well validated for the application. This is particularly important for trace evidence analysis. An SOP should describe the target extraction and, where appropriate, concentration of the target to be analyzed.

For the extraction of targets, a minimum checklist of validation criteria should consider several factors (Table 4).

(iii) Analytical component of the process. For some types of analyses, there are substantial descriptions of validation procedures in the literature. Notable examples of these are analytic procedures for the identification and quantification of chemicals and analytical procedures for the detection and quantification of nucleic acid targets (13, 24). However, sometimes validation of an analytical process of an assay lacks the rigor necessary for high confidence or effective interpretation. One classic example, which demonstrates the need for rigorous validation and the problems that may arise by premature release for general use, is in the human forensic DNA arena. Typing of the HLA-DQA1 locus was based on the amplification by PCR of a specified polymorphic region on chromosome 6 (11). This was one of the first PCR-based identification loci developed for use in a commercial kit assay. Since the design of the molecular assay appeared robust, the manufacturer advocated its use, and some forensic scientists began using the assay for analyses of crime scene evidence before a rigorous validation was completed. Typing inconsistencies began to emerge. Some truly heterozygous individuals typed as apparent ho-

TABLE 4. Checklist of elements of the extraction validation criteria

Element(s) of extraction validation criteria
Specific target, namely, the virus, bacterium, fungus, or toxin, whether spores or vegetative cells are being recovered (spores generally need a more vigorous extraction method)
Assessment of whether the material or analyte needs to maintain viability or activity or is an inactive agent or material acceptable
Analyte(s) that will be assayed, e.g., DNA, RNA, protein, lipid, stabilizers, media, fatty acid, and other trace evidence and possibly combinations of these
Quantity of analyte needed for the subsequent analysis method(s)
Purity of the analyte required by the analytical method
Matrix effect, specifically, the matrix or substrate or material from which the target will be extracted (e.g., food, blood, soil, or carpet, etc.) and any known effects of the matrix on the extraction method; whether the application of the method to this matrix is still within the method scope; i.e., if the matrix is new and the effects are unknown, then the method may need to be revalidated with the new matrix
Recovery and efficiency of yield
Stability of the analyte prior to analysis (optimal storage conditions of the extracted analyte should be described)
Critical reagents
Critical equipment, such as bead beaters for spores and a fume hood for extractions involving volatile chemicals
Controls needed to assess the performance of the extraction process
Personnel training, proficiency, and safety

mozygotes. Typing inconsistencies might lead to false-positive results and false exclusion interpretations. One of the alleles of the HLA-DQA1 locus (allele 1) has four more GC residues than the other alleles (alleles 2, 3, and 4). If the denaturing temperature during the PCR amplification is not sufficient, allele 1 may not denature and thus not amplify, while the other alleles will amplify. This occurred because the first thermal cyclers were not as well designed as the ones in use today. The temperature of the outer wells in the early-model thermal cycler was lower than that in the inside wells. Thus, samples placed in the outside wells could yield incorrect results. This phenomenon was discovered during validation studies (11), well after the product was commercialized. Remedies for this potential allele dropout problem were to (i) use only the internal wells, (ii) add formamide to the PCR mixture to reduce the required denaturing temperature, and (iii) design a better-performing thermal cycler. Thus, validation is required for all aspects of the assay, not just one component.

SOPs for chemical or molecular analysis should identify (i) all reagents critical to the procedure that should be tested before analyzing unknown samples; (ii) critical equipment, calibration, and certification requirements; and (iii) known positive-, negative-, and/or internal-control samples used with the analysis. A validation plan for the analytical portion should consider, when appropriate, a number of parameters (Table 5).

An analysis may have several objectives. One objective that quantitative methods have is to establish the amount of a target analyte that is present in a sample, and many of the parameters listed in Table 5 would apply. Qualitative analysis often seeks to determine if a particular analyte is present or not and therefore does not require validation for some of the criteria listed above. However, personnel considerations and a method's sensitivity, specificity, reproducibility, robustness, limits of detection, precision, accuracy, input values, controls, selectivity, critical reagents, and equipment are all important criteria for validating qualitative methods.

(iv) Interpretation significance or weight of evidence. Assays used in a forensic context usually have a purpose that goes beyond simply identifying or quantifying the presence of an analyte. More often, the result of the analysis is used as evidence to support some inference of forensic relevance. Examples of such inferences are that a particular agent was found in a particular place, that a particular pathogen was transmitted from a suspect to a victim, that a particular method was employed to generate the agent, and that the agent was produced within a particular time frame.

Simply stated, interpretation is the stage of evidence analysis where questions about such issues are addressed. Thus, in the context of microbial forensic evidence, the results of interpretation relate to the identification of the microbial components in the sample and possible source attribution of the samples in the evidence. The interpretation of evidence should be based on well-validated criteria that reflect the existing knowledge surrounding the analysis of the evidence, its collection and storage, and the handling procedures used.

When interpreting microbial forensic evidence, a qualitative and/or quantitative statement about the outcome of the analysis should be provided. The general approaches to these statements should be contained in the interpretation/reporting section of the SOP. The SOP should contain a description of the criteria to be used to assess the significance of results, such as analyte identification. Such criteria include thresholds and the significance of artifacts.

Validation determines the limits of a test. It does not mean that a test must be 100% accurate or have no false-positive or false-negative results to be considered useful. However, solely testing for an analyte may not be sufficient to consider a procedure validated. Often an interpretation of the result is necessary so that proper actions can be taken or decisions can be made, and these can be impacted by specificity and cross-reactivity. A clinical example with the inclusion or exclusion of a diagnosis of Lyme disease illustrates this point and the importance of the context and intended purpose of a test. Most current serologic assays for Lyme disease detect an antigen (flagellin) that also is found in the pathogen that causes syphilis (*Treponema pallidum*) (15) and thus can impact interpretations of what constitutes a positive test. For example, if a symptomatic patient in an area of Lyme disease endemicity has a positive enzyme-linked immunosorbent assay result, a physician's first inclination may be to treat the patient for Lyme disease. However, if the patient has a history of sexually transmitted disease, further testing is warranted for both syphilis and Lyme disease (22).

The use of the "rapid strep test" to diagnose a sore throat as

TABLE 5. Parameters to consider in the analytic component of a validation plan

Parameter(s) of a validation plan	Description
Sensitivity	Minimum amount or concentration of an analyte required to generate a reliable result ("analyte" is used here generically and may refer to an entire microorganism or to an ion)
Specificity	Ability to measure the intended target, analyte, or signature
Reproducibility	Closeness of agreement between the results of successive measurements of the same analyte under similar but not necessarily identical conditions
Precision	Degree to which individual measurements of the same sample are similar
Accuracy	Degree to which the measured material or analyte is similar to its true value
Robustness	Stability of analytical performance under variable conditions
Reference samples	Samples for testing the performance of the assay (e.g., reference panels of the target and mock or nonprobative materials) corresponding to the intended application of the assay. Assays should be evaluated for performance in the presence of potential interfering substances that may mimic some conditions encountered with forensic samples.
Input values	Range of quantities of an analyte that can be analyzed reliably
Quantitation	Amount or concentration of an analyte
Dynamic range	Range of values or limits within which precision is held
Limit of detection	Minimum level at which all replicates are consistently positive
Controls	Test materials of known value for the measured analyte (includes blind samples, negative controls, and positive controls)
Selectivity	Ability to measure and to differentiate the analyte in the presence of other materials that may be present
Window of performance for operational steps of assay	Parameters that define analytical condition variations that will not substantially affect performance or reliability
Critical equipment calibration	Equipment requiring calibration prior to its initial use and on a regular basis thereafter
Critical reagents	Reagents determined by empirical studies or routine practice to require testing on known samples prior to use with evidentiary materials in order to prevent an unnecessary consumption of forensic samples
Databases	Collection of data to be used to support an interpretation of results
Personnel training, proficiency, and safety	Qualifications and education of the personnel to conduct the analysis safely
Interpretation criteria for results	Analysis (statistical or qualitative) used and confidence level of a match or result

an infection with group A beta-hemolytic streptococci is an example of a test with a substantial number of false-negative results (18, 26); yet the assay is very useful when a decision-making algorithm is followed (18, 26). The purpose of this test is to diagnose strep throat rapidly so that antibiotics can be administered immediately. The intent is to treat the patient effectively and to reduce contagion to others. A positive result for a patient with a sore throat is assumed sufficiently definitive to treat. However, a negative "rapid strep test" result, which is known to have a 10 to 20% false-negative rate, is followed up by traditional cultures before doctors conclude that the patient is not infected with group A beta-hemolytic streptococci. This also minimizes unnecessary antibiotic administration. Thus, having a significant false-negative rate does not mean that the procedure is not valid. The validation process determined the limitations of the test, and the physician must appreciate the limits in order to interpret the results properly and to decide what proper follow-up analyses may be needed (18, 26).

Validation should support or provide a basis for the elements relevant to the situation in which a test is being employed (Table 6).

TABLE 6. Checklist of relevant elements of interpretation

Element(s) of interpretation
Qualitative statement
Quantitative statement
Semiquantitative statement
Database (with an understanding of its relevance and representativeness and the quality of data in the database. The issue of rarity is an essential part of any possible source attribution statement of the interpretation process. Inferences of rarity are based on the sample population analyzed and assumptions of relevance and representativeness are basic to the degree of certainty. Therefore, the reference population data used should be defined. Documentation of the construct of an appropriate reference database must be maintained.)
Endemicity or background data (environmental and/or laboratory normal values and reference ranges, etc.)
Criteria for deciding whether a result requires follow-up or further analysis (includes temporal/spatial analysis and effect of passage, etc.)
Determination of alternate (reasonable) explanations
Limits of interpretation based on extant science and context
Statistical approaches used to reach conclusions, such as match, presence or absence, similarity, most recent common ancestor, and identity, etc.
Interpretation thresholds
Software reliability (should include the inherent assumptions underlying the computations that the software performs and the justifications of the assumptions; appropriate citations of prior studies regarding these elements should be made)

VALIDATION AS A DYNAMIC PROCESS

The validation process is not a one-time event for a method. It must be considered dynamic in order to assess periodically the impact of new knowledge and findings and to assess ma-

terial modifications made to existing methods and procedures. Indeed, monitoring and reassessment are tools to ensure that even previously validated processes remain valid if the parameters under which the process is carried out are altered (this can include simultaneous increases in humidity and temperature or other factors that may effect the manufacture or stability of reagents). Similarly significant revisions of existing protocols should undergo validation commensurate with the modification. The revised protocol should be documented, dated, and identified as a more recent version. Already-validated protocols obtained from other laboratories should still undergo an internal validation prior to their use on casework. Before beginning routine casework, scientists and/or other practitioners should successfully complete a qualifying test using the procedure.

One example of dynamic validation arose through the BioWatch program. This program is a nationwide surveillance system for sampling air for the presence of selected pathogens. It should provide an early warning of a potential bioterrorism attack that is more timely than waiting for infected individuals to present to the health care system with symptoms. Material extracted from filter-collected air samples is extracted and subjected to real-time PCR analysis of DNA with primers specific for certain select agents. In October 2003 in the Houston, Texas, area, several BioWatch filters tested positive for *Francisella tularensis* (2). Because the pattern of positive results was not consistent with that of a deliberate release of *F. tularensis*, the interpretation was that these observations were due to a natural event. Soil and water samples collected proximal to the positive BioWatch samples were extensively characterized to determine the source of the positive results. DNA extracts of the environmental samples were screened first by typing the 16S rRNA gene to detect *Francisella* species and related organisms. Subsequent cloning and sequencing of the PCR products indicated the presence of a wide variety of *Francisella*-related species, some of which were quite distinct from known *Francisella* species and appeared to be new species or genera (2). These findings impact the specificity of the PCR-based assay and the ability to differentiate an agent introduced in a bioterrorism attack from a naturally occurring strain. In the concept of method validation, the specificity of the BioWatch tests was validated with existing data, as is any system. Now that a previously unknown type of *Francisella* that is positive by the BioWatch system has been found in the environment, the interpretation criteria for the *F. tularensis* assay should be modified and additional recommendations for follow-up analyses may be required. Given the unknown diversity of the microbial world, additional experience will be gained as more samples are analyzed and further optimization of the method may be indicated.

CONCLUSIONS

In conclusion, validation is an essential process for any scientific discipline, including the evolving field of microbial forensics. It is important to define, or at the very least describe, what one means by the term validation and what parameters were tested in the development of a method. The description of validation should encompass the context and purpose of that which is being validated. As a field still developing, microbial forensics has a motivation and an unencumbered opportunity to define validation more explicitly. Therefore, a minimum set of criteria that should be considered in developing a validation plan for microbial forensic methods have been presented herein. These criteria apply to global goals, the collection process, transport and storage under a chain of custody, laboratory procedures, and interpretation. It is not possible to be all-inclusive because of the wide diversity of samples and sample types for which microbial forensic methods may be applied. Therefore, this outline serves as a guiding document, rather than an exhaustive prescriptive one. Careful consideration of the criteria presented here will be useful for transitioning research efforts into operational microbial forensic settings. Application of these criteria also will be useful for generally improving the quality of microbial forensic research efforts and laying a solid foundation for developing the field of microbial forensic science.

APPENDIX

A glossary of microbial forensics validation definitions (derived and/or modified from references 5, 7, and 8) follows.

Accuracy: the degree of conformity between the result of a measurement and a true value of the analyte, target, or signature.

Analytical procedure: an orderly step-by-step procedure designed to ensure operational uniformity and to minimize analytical drift.

Attribution: the information obtained regarding the identification or source of a material to the degree that it can be ascertained.

Bias: systematic error in measurement of an analyte, target, or signature.

Biocrime or bioterrorism: the threat or use of microorganisms, toxins, pests, or prions or their associated ancillary products to commit acts of crime or terror.

Calibration: a set of operations that establish, under specified conditions, the relationship between values provided by a measuring instrument, a measuring system, and a known material or known values.

Control samples: test materials whose identity, type, and/or values have been established.

Internal control: a control sample placed in the same tube or well as the analyte, target, or signature that is being analyzed so that it will be subjected to the exactly the same conditions as the analyte, target, or signature.

Negative control: a sample similar in nature to what is being analyzed but which does not contain the analyte, target, or signature. This control is run in parallel with the analyte, target, or signature. A positive result from a negative control may invalidate an analysis.

Positive control or known sample: a test material intended

for use in the quality control process whose identity, type, and/or values have been established. These controls are used for day-to-day monitoring of typing activities. Controls are well known and well characterized and are to be analyzed in parallel with the test samples in an assay. A negative or incorrect result from a positive control may invalidate an analysis.

Critical equipment or instruments: those requiring calibration prior to their initial use and on a regular basis thereafter.

Critical reagents: determined by empirical studies or routine practice to require testing on known samples prior to use with evidentiary materials in order to prevent unnecessary consumption of forensic samples.

Error: obtaining an incorrect value, or the difference between a measurement and a true value. Measurement error is composed of random variability and bias (systematic or nonrandom effects on the measurement).

Imprecision: the distribution of independent results of measurements or values derived under specified conditions.

Limit of detection: the lowest concentration of analyte, target, or signature that can be consistently detected in a specified sample and can be distinguished from a sample that does not contain the analyte, target, or signature.

Measuring range: a defined range or limits of values for an analyte, target, or signature within which error of measurement of a system or process is expected to lie.

Microbial forensics: a scientific discipline that examines microorganisms, toxins, pests, or prions or their associated ancillary products for source attribution.

Precision: a measure of the extent (or nearness) of variation in values obtained from replicate determinations. It conveys the degree of repeatability expected for additional replicate measurements or values. It does not necessarily convey trueness or accuracy.

Quality assurance: the system of management activities designed to ensure that a process, item, or service is of the type and quality needed. This includes monitoring activities that are intended to verify whether practices and test results are providing reliable and relevant information.

Quality control: a mechanism or laboratory activity intended to verify whether test conditions are functioning appropriately to yield reproducible results.

Reference material (certified or standard): a material for which identities, types, or values are certified by technically valid procedures and is accompanied by, or traceable to, a certificate or other documentation.

Repeatability: the degree of consistency between or among results from successive measurements of the same analyte, target, or signature obtained under the same conditions.

Reproducibility: the degree of consistency between or among results from successive measurements of the same analyte, target, or signature obtained under changed conditions.

Resolution: the smallest difference between measurements or values that can be reliably distinguished.

Sensitivity: the concentration of analyte, target, or signature that is necessary to produce a reliable result.

Specificity: the ability of an assay to measure or type the analyte, target, or signature that it is intended to analyze, even when other components in the sample may inhibit or interfere, cross-react, or compete with the assay.

Standards: defined analytes, targets, or signatures used to characterize the performance of an assay within defined limits.

Standard operating protocol: a set of explicit instructions (and necessary supporting documentation) for the operation of a specified procedure.

Traceability: the property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards or other documented mechanisms, through an unbroken chain of comparisons.

Trueness: the difference of a measurement or result from that of a true value, accepted standard, or expected value.

Validation: a process by which a procedure is evaluated to determine its efficacy and reliability for analysis. There are three categories of validation defined for microbial forensics: developmental, internal, and preliminary.

Developmental validation: the acquisition of test data and the determination of conditions and limitations of a newly developed methodology for use on samples. Developmental validation should be appropriately documented and should address specificity, sensitivity, reproducibility, bias, precision, false positives, false negatives, and determination of appropriate controls. Any reference database used should be documented.

Internal validation: an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform within determined limits in the laboratory.

Preliminary validation: the acquisition of limited test data to enable an evaluation of a method used to assess materials derived from a biocrime or bioterrorism event. Also termed "validation on the fly," which is utilizing a process during exigent circumstances where the situation has not been encountered and standard operating protocols have yet to be developed.

[*Note:* "Analyte, target, or signature" is used generally and can be anything that may be analyzed, ranging from an intact viable microorganism to an ion.]

ACKNOWLEDGMENTS

This is publication number 08-05 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers

are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

REFERENCES

1. **Barakat, L. A., H. L. Quentzel, J. A. Jernigan, D. L. Kirschke, K. Griffith, S. M. Spear, K. Kelley, D. Barden, D. Mayo, D. S. Stephens, T. Popovic, C. Marston, S. R. Zaki, J. Guarner, W. J. Shieh, H. W. Carver, R. F. Meyer, D. L. Swerdlow, E. E. Mast, and J. L. Hadler.** 2002. Fatal inhalational anthrax in a 94-year-old Connecticut woman. JAMA **287:**863–868.
2. **Barns, S. M., C. C. Grow, R. T. Okinaka, P. Keim, and C. R. Kuske.** 2005. Detection of diverse new *Francisella*-like bacteria in environmental samples. Appl. Environ. Microbiol. **71:**5494–5500.
3. **Budowle, B., S. E. Schutzer, M. S. Ascher, R. M. Atlas, J. P. Burans, R. Chakraborty, J. J. Dunn, C. M. Fraser, D. R. Franz, T. J. Leighton, S. A. Morse, R. S. Murch, J. Ravel, D. L. Rock, T. R. Slezak, S. P. Velsko, A. C. Walsh, and R. A. Walters.** 2005. Toward a system of microbial forensics: from sample collection to interpretation of evidence. Appl. Environ. Microbiol. **71:**2209–2213.
4. **Budowle, B., S. E. Schutzer, J. P. Burans, D. J. Beecher, T. A. Cebula, R. Chakraborty, W. T. Cobb, J. Fletcher, M. L. Hale, R. B. Harris, M. A. Heitkamp, F. P. Keller, C. Kuske, J. E. Leclerc, B. L. Marrone, T. S. McKenna, S. A. Morse, L. L. Rodriguez, N. B. Valentine, and J. Yadev.** 2006. Quality sample collection, handling, and preservation for an effective microbial forensics program. Appl. Environ. Microbiol. **72:**6431–6438.
5. **Budowle, B., S. E. Schutzer, A. Einseln, L. C. Kelley, A. C. Walsh, J. A. Smith, B. L. Marrone, J. Robertson, and J. Campos.** 2003. Public health. Building microbial forensics as a response to bioterrorism. Science **301:**1852–1853.
6. **Clinical and Laboratory Standards Institute.** 2005. Collection, transport, preparation, and storage of specimens for molecular methods. Approved guidelines. CLSI document MM13-A. Clinical and Laboratory Standards Institute, Wayne, PA.
7. **Clinical and Laboratory Standards Institute.** 2005. User verification of performance for precision and trueness. Approved guidelines, 2nd ed. Clinical and Laboratory Standards Institute, Wayne, PA.
8. **Clinical and Laboratory Standards Institute.** 2006. Genotyping for infectious diseases: identification and characterization. Approved guidelines. CLSI document MM10-A. Clinical and Laboratory Standards Institute, Wayne, PA.
9. **Clinical and Laboratory Standards Institute.** 2006. Molecular diagnostic methods for infectious diseases. Approved guidelines. CLSI document MM3-A2. Clinical and Laboratory Standards Institute, Wayne, PA.
10. **Clinical and Laboratory Standards Institute.** 2008. Verification and validation of multiplex nucleic acid assays. Approved guidelines. Clinical and Laboratory Standards Institute, Wayne, PA.
11. **Comey, C. T., and B. Budowle.** 1991. Validation studies on the analysis of the HLA DQ alpha locus using the polymerase chain reaction. J. Forensic Sci. **36:**1633–1648.
12. **Harmon, R.** 2005. Admissibility standards for scientific evidence, p. 381–392. *In* R. G. Breeze, B. Budowle, and S. E. Schutzer (ed.), Microbial forensics. Academic Press, San Diego, CA.
13. **Herrmann-Hoesing, L. M., S. N. White, G. S. Lewis, M. R. Mousel, and D. P. Knowles.** 2007. Development and validation of an ovine progressive pneumonia virus quantitative PCR. Clin. Vaccine Immunol. **14:**1274–1278.
14. **International Organization for Standardization (ISO), International Electrotechnical Commission (IEC).** 1999. General requirements for the competence of testing and calibration laboratories. ISO/IEC 17025. American National Standards Institute, New York, NY.
15. **Jiang, W., B. J. Luft, W. Schubach, R. J. Dattwyler, and P. D. Gorevic.** 1992. Mapping the major antigenic domains of the native flagellar antigen of *Borrelia burgdorferi*. J. Clin. Microbiol. **30:**1535–1540.
16. **Kreader, C. A.** 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl. Environ. Microbiol. **62:**1102–1106.
17. **Locascio, L. E., B. Harper, and M. Robinson.** 2007. Standard practice for bulk sample collection and swab sample collection of visible powders suspected of being biological agents from nonporous surfaces: collaborative study. J. AOAC Int. **90:**299–333.
18. **Mirza, A., P. Wludyka, T. T. Chiu, and M. H. Rathore.** 2007. Throat culture is necessary after negative rapid antigen detection tests. Clin. Pediatr. (Philadelphia) **46:**241–246.
19. **Murray, P. R., E. J. Baron, J. H. Jorgensen, M. A. Pfaller, and R. H. Yolken (ed.).** 2003. Manual of clinical microbiology, 8th ed. ASM Press, Washington, DC.
20. **Peters, F. T., O. H. Drummer, and F. Musshoff.** 2007. Validation of new methods. Forensic Sci. Int. **165:**216–224.
21. **Schutzer, S. E., B. Budowle, and R. M. Atlas.** 2005. Biocrimes, microbial forensics, and the physician. PLoS Med. **2:**e337.
22. **Schutzer, S. E., P. K. Coyle, P. Reid, and B. Holland.** 1999. *Borrelia burgdorferi*-specific immune complexes in acute Lyme disease. JAMA **282:**1942–1946.
23. **U.S. General Accounting Office.** 2005. Anthrax detection: agencies need to validate sampling activities in order to increase confidence in negative results. Document GAO-05-251. U.S. General Accounting Office, Washington, DC. http://www.gao.gov/new.items/d05251.pdf.
24. **Ward, C. L., M. H. Dempsey, C. J. Ring, R. E. Kempson, L. Zhang, D. Gor, B. W. Snowden, and M. Tisdale.** 2004. Design and performance testing of quantitative real time PCR assays for influenza A and B viral load measurement. J. Clin. Virol. **29:**179–188.
25. **Watson, R. J., and B. Blackwell.** 2000. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. Can. J. Microbiol. **46:**633–642.
26. **Wright, M., G. Williams, and L. Ludeman.** 2007. Comparison of two rapid tests for detecting group A streptococcal pharyngitis in the pediatric population at Wright-Patterson Air Force Base. Mil. Med. **172:**644–646.