Battelle
The Business of Innovation

Identifying Best Practices in Laboratory Medicine

Review of Proficiency Testing Services for Clinical Laboratories in the United States – Final Report of a Technical Working Group

By
James C Peterson
Robert H Hill
Robert S Black
James Winkelman
Daniel Tholen

Battelle Memorial Institute
2971 Flowers Road SW
Atlanta, GA 30341

Prepared for
Division of Laboratory Systems
National Center for Preparedness, Detection, and Control of Infectious Diseases
Coordinating Center for Infectious Diseases
Centers for Disease Control and Prevention
Atlanta, Georgia

Contract No. 200-2001-00121, Task No. 0019

April, 2008

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Centers for Disease Control
and Prevention

April 2008

Dear Colleague:

I am pleased to announce a new report based on work sponsored by the Division of Laboratory Systems Laboratory Medicine Best Practices initiative, *Review of Proficiency Testing Services for Clinical Laboratories in the United States – Final Report of a Technical Working Group*, prepared by Battelle Memorial Institute.

The Laboratory Medicine Best Practices initiative aims to:

- Improve patient safety and health care outcomes by improving the use of laboratory testing in screening, diagnosis, monitoring, and management of disease
- Reduce redundancy and waste in laboratory services
- Provide tools that laboratories can use to improve quality of service to clinicians and patients

This report assesses proficiency testing (PT) in relation to the regulatory, educational, and quality improvement objectives of the Clinical Laboratory Improvement Amendments of 1988.

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Work on the Laboratory Medicine Best Practices initiative continues in 2008. For more information about the initiative, please visit our web site at http://wwwn.cdc.gov/dls/bestpractices/. I hope you find the report of interest.

Sincerely yours,

D. Joe Boone, Ph.D.

D. Joe Boone, Ph.D.
Acting Director
Division of Laboratory Systems (DLS)
National Center for Preparedness, Detection,
and Control of Infectious Diseases
Centers for Disease Control and Prevention

The following Battelle staff members were primarily responsible for the content of this report:

James C. Peterson, PhD

Research Leader, Analytical Chemistry Services
Battelle Eastern Science and Technology Center
Aberdeen, Maryland

Robert H. Hill, Jr., PhD

Public Health Program Manager
Atlanta Analytical Chemistry Group
Atlanta, Georgia

Robert S. Black, MPH

Senior Health Research Scientist
Centers for Public Health Research and Evaluation
Atlanta, Georgia

with substantial contributions from subcontractors

James Winkelman, MD, FACP

Professor, Harvard Medical School (Emeritus)
Chestnut Hill, Massachusetts

and

Daniel Tholen, MS

Dan Tholen Statistical Consulting
Traverse City, Michigan

CONTENTS

EXECUTIVE SUMMARY

In 2006, the Division of Laboratory Systems (DLS) at the Centers for Disease Control and Prevention (CDC) assembled a thirteen-member Working Group to assess the effectiveness of clinical laboratory proficiency testing (PT) for regulatory, educational, and quality improvement purposes. Members were chosen for their expertise in PT and laboratory medicine and comprised PT users, PT providers, and representatives of accrediting organizations.

Objectives

DLS/CDC representatives charged the Proficiency Testing Working Group (PTWG) to accomplish four primary objectives:

1. Report on the status of current clinical PT programs in the US, assessing the success of PT in improving the quality of clinical testing and identifying areas in which improvements are needed.
2. Make recommendations for improving the effectiveness of PT in meeting regulatory, educational, and quality improvement objectives.
3. Solicit and consider input from stakeholders including PT providers, PT users in several types of clinical laboratories, and accrediting organizations; assess the feasibility of recommendations.
4. Identify needed improvements to PT that could be addressed during the next five years.

Process

The PTWG met in January and May 2007 to address a series of questions developed by DLS staff and Robert Rej, PhD, the PTWG Chair. Battelle solicited input from stakeholders in the laboratory medicine community in a two-step process. Before the initial Working Group meeting, Battelle invited 79 stakeholder organizations to submit comments and relevant data. After the PTWG's second meeting, Battelle sent a summary of salient points to be considered in the final report to all stakeholders, along with an invitation to provide additional comments, suggestions, or relevant data. A total of 20 organizations submitted comments. Staff from Battelle and DLS/CDC and the PTWG Chair reviewed all comments and integrated them into the report narrative.

Background

With the passage of the 1988 Clinical Laboratory Improvement Amendments (CLIA '88) and its implementing regulations, laboratories performing non-waived testing for selected analytes and tests were required to participate in PT. CLIA established a framework for certification of clinical laboratories operating in the United States and for use of PT in assessing laboratory performance. The 1992 regulations identified analytes in 16 clinical laboratory specialties and subspecialties and specific PT criteria for each analyte. For most analytes, PT programs must include at least three events annually, each of which must include five samples. For most

specialties, satisfactory performance requires that 80% (4/5) of a participating laboratory's test results fall within a specified range of analytical precision. PT providers issue PT samples to each subscriber laboratory, receive results of PT analyses from each laboratory, report PT results back to laboratories, and report laboratory performance evaluated against specific scoring criteria for CLIA-required PT tests for each laboratory to the Centers for Medicare and Medicaid Services (CMS), which is primarily responsible for administering CLIA. The CMS approval process requires PT providers to demonstrate annually how each meets the requirements of the CLIA regulations.

Many tests for analytes that were not included in CLIA regulations are now in common or routine clinical use. For example, tests for cardiac markers such as troponins and many types of molecular genetic testing that are now in general use did not exist in 1992. Although PT is not required for "non-regulated analytes," laboratories are required to verify the accuracy and reliability of test results for these analytes at least twice a year.

Assessment tasks

The PTWG's work was organized into five interrelated tasks, which are listed here. The Working Group discussed each task in light of the current state of the field and attempted to identify gaps; the group discussed potential approaches to fill gaps or improve processes and programs.

Task 1. To assess how well PT programs used by U.S. laboratories meet quality improvement, educational, and regulatory goals

The PTWG considered existing evidence that PT is effective, possible approaches to gather additional data, and criteria for assessing the performance of PT providers.

Task 2: To evaluate the need to improve the use of PT programs

The Working Group considered gaps in current PT program content; potential improvements in the design, integrity, and distribution of PT samples; and possible approaches to improving the evaluation (grading or scoring) of PT results.

Task 3. To evaluate ways to enhance the educational value of PT programs

PT has an important educational role, both in remediation after a PT failure and in raising awareness of the value of monitoring trends in PT results for warning signs of impending failures. The Working Group discussed several ways in which PT providers can reinforce the educational element of PT.

Task 4. To determine how well PT programs are providing challenges to keep up with advances in laboratory testing technology

Programs offered by domestic PT providers cover analytes that comprise the bulk of laboratory testing volume. Gaps in PT coverage exist for newly developed tests, esoteric tests, tests in which analyte stability is problematic, and tests of analytes in matrices such as hair, saliva, and sweat. Reference and specialized laboratories that conduct these less common tests are most likely to encounter problems because of these gaps. PT for molecular diagnostic testing presents

special challenges because of the large number of clinically relevant gene targets. Methods-based PT may be the most practical way to address this gap.

Task 5. To determine whether accreditation of PT programs to an international standard would increase quality or uniformity of programs.

The PTWG considered operational changes and costs, as well as potential benefits of and barriers to accreditation to the consensus international standard, ISO/IEC Guide 43 and corresponding requirements detailed in ILAC document G13, *Guidelines for the Requirements for the Competence of Providers of Proficiency Testing.* These standards are general and intended to apply to a broad range of PT programs. CLIA requirements, in contrast, specify much of the technical content of PT programs and include much less general guidance.

Recommendations

The PTWG developed 21 recommendations that address data collection and analysis, process improvement, process evaluation, and education. If adopted, these recommendations may help improve the state of PT. Although implementing them presents many challenges to the laboratory medicine community, most are feasible. Some recommendations require changes in CLIA regulations or the CLIA statute. Most require new resources (i.e., funding and/or personnel). CDC, CMS, and the larger laboratory community must determine whether the potential benefit from implementing each recommendation warrants the additional costs to PT providers, PT users, and perhaps ultimately to patients.

Data collection and analysis

1. Conduct a study of the existing information in the scientific literature and current databases regarding reasons for unsatisfactory PT results in order to identify areas most in need of improvement or additional research/analysis.
2. Develop and make available a database to collect PT data for characterizing the performance of all laboratories, for identifying reasons for unsatisfactory PT results, for reviewing acceptance criteria used by PT providers, and for identifying a list of analytes that should be regulated.
3. Develop a process to collect, consolidate, analyze, and summarize all complaints received by CMS, state health programs, accrediting organizations, and PT providers about PT. This process includes developing appropriate statistical analyses of data to identify correctable trends and the publication and dissemination of the complaint process for widespread use by all parties.
4. PT providers should publish scientifically credible reports of PT results on a regular basis in peer-reviewed journals.
5. CDC should continue to maintain and update the listing of national and international PT programs on its website.

Process improvement

6. Develop a process to assure that all clinical laboratories, including those that perform waived tests, participate in PT. This recommendation requires a change in the CLIA statute.

7. Develop a process to periodically review, update, and publish the requirements of the CLIA PT program, including the list of regulated analytes and allowable limits.
8. PT providers should seek ways to provide for faster turnaround time for PT results, including developing a system(s) for electronic submission.
9. Before releasing official results, PT providers should consider providing immediate feedback to laboratories when results indicate that PT failure is likely. PT providers should also institute a system that gives warning to laboratories that trends of cumulative results are moving toward PT failure.
10. PT providers should allow for the reporting of analyte results in various units of measure, be able to convert those measures to common units, and evaluate them in accordance with current regulations.
11. PT providers should summarize PT results graphically for end users in a manner that is easy to read and understand.
12. PT Providers should provide samples that mimic patient samples as much as possible with a minimum of artificial matrix effects.
13. Small adjunct studies using fresh frozen samples from a single patient should be conducted in conjunction with routine PT to identify and characterize unrecognized testing problems.
14. An independent advisory board should be established for the purpose of identifying new and evolving technologies and analytes in laboratory medicine, to develop innovative approaches in PT programs, and to alert PT providers of new opportunities for PT offerings.
15. Rather than developing a unique test for each of the rapidly increasing number clinically relevant molecular genetic tests, develop a methodology-based approach for PT that can be used to assess proficiency in process elements common to many tests (e.g., nucleic acid sequencing, PCR amplification and purification, electrophoresis and interpretation).
16. Encourage U.S. PT providers to assess the use of internationally recognized PT standards (ILAC-G13:2006 or ISO Guide 43-1: 1997) for EQA.
17. Assess the benefits and costs of adopting an international standard that requires PT providers to be audited by a qualified third party.

Process evaluation

18. Evaluate alternatives to current CLIA requirements for the frequency of PT events and the number of samples in each event.
19. Evaluate alternatives to the PT scoring approaches currently in use under CLIA.

Education

20. To increase the educational value of PT participation, develop an educational program that teaches laboratory personnel how to evaluate PT results.
21. Using an approach such as that described by publication GP27 from the Clinical and Laboratory Standards Institute (CLSI), PT providers should offer training materials on interpretation and use of PT results in quality improvement processes.

CHAPTER 1. INTRODUCTION

In 2006, the Division of Laboratory Systems (DLS) at the Centers for Disease Control and Prevention (CDC) assembled a thirteen-member Working Group to assess the effectiveness of clinical laboratory proficiency testing (PT) for regulatory, educational, and quality improvement purposes. Membership of the Working Group was selected to provide a balanced representation of PT users, PT providers, and representatives of accrediting organizations. Members were chosen for their expertise in PT and laboratory medicine and not as representatives of their employers, professional associations, or trade associations (Table 1-1).

Objectives

DLS/CDC representatives charged the Proficiency Testing Working Group (PTWG) to accomplish four primary objectives:

5. Report on the status of current clinical PT programs in the US, assessing the success of PT in improving the quality of clinical testing and identifying areas in which improvements are needed.
6. Make recommendations for improving the effectiveness of PT in meeting regulatory, educational, and quality improvement objectives.
7. Solicit and consider input from stakeholders including PT providers, PT users in several types of clinical laboratories, and accrediting organizations. Assess the feasibility of recommended changes.
8. Identify needed improvements to PT that could be addressed during the next five years.

Process

Stakeholders in the laboratory medicine community provided input to the Working Group in response to an invitation sent to a list of PT stakeholders. Initially, stakeholders were asked to provide comments or data they believed to be relevant in time for the Working Group's initial –in person meeting on January 18-19, 2007. Following the initial meeting, a summary of the Working Group's discussions and salient points to be considered in the final report was sent to all stakeholders with an invitation to provide additional comments, suggestions, or relevant data. Seventy nine organizations and individual stakeholders were invited to provide comments; a total of 20 organizations did so. Comments were reviewed by Battelle and DLS/CDC staff and integrated as appropriate into the report narrative.

The Working Group met three times, once in a conference call in December 2006 and twice in person (January and April 2007). Discussion at the two face-to-face meetings was recorded and transcribed to produce a detailed report (not a verbatim transcript), which served as the basis for drafts of the Working Group report.

Report Organization

This report summarizes the PTWG's deliberations and recommendations in response to questions posed by CDC.

Chapter 2 provides background concerning the regulatory framework and current practice of PT in the United States.

Chapter 3 addresses in turn the following five PTWG tasks:

1. Evaluate the PT programs used by U.S. laboratories to determine the extent to which these programs meet quality improvement, educational, and regulatory goals.
2. Evaluate the need to improve the use of PT programs.
3. Evaluate ways to enhance the educational value of PT programs.
4. Determine how well PT programs provide samples that keep up with advances in laboratory testing technology.
5. Determine whether accreditation of PT programs to an international standard would increase the quality or uniformity of programs.

To facilitate discussion and solicitation of input, CDC formulated a series of questions to be addressed by the PTWG in accomplishing each task. These questions constitute the subheadings of Chapter 3.

Chapter 4 summarizes the Working Group's recommendations concerning data collection and analysis, process improvement, process evaluation, and education.

Table 1-1. Proficiency Testing Working Group Members

***Barbara Burmeister*,** MT (ASCP); Supervisor, Proficiency Testing; Wisconsin State Laboratory of Hygiene

***George S. Cembrowski*,** MD, PhD; Director, Medical Biochemistry; Associate Professor, Laboratory Medicine and Pathology, University of Alberta

Kandace Cendejas;* Senior Product Manager, External Quality Assessment, BioRad Laboratories

***Greg Cooper*,** CLS, MHA; Manager, Clinical Standards and Practices, Bio-Rad Laboratories

Daniel Edson, PhD; President, American Proficiency Institute

George K. Fiedler; Vice President and General Manager, PT Division, College of American Pathologists

Judith Gabriel;** Senior Surveyor, The Joint Commission

Verlin K. Janzen, MD, FAAFP; Clinical Assistant Professor, Department of Family and Community Medicine, University of Kansas School of Medicine- Wichita; Family Physician & Laboratory Director, Hutchinson (KS) Clinic, P.A.

Margaret Peck, MS, MT (ASCP); Director, Laboratory Accreditation Program, The Joint Commission

Joseph B Perrone, ScD; Vice President, Standards and Certification, American Type Culture Collection

Robert Rej, PhD [Chair]; Director, Clinical Chemistry, Acting Director, Hematology, New York State Department of Health

Karen A. Rupke, MT (ASCP), MPA; Director, California Quality Assurance, Quest Diagnostics

Nicholas T. Serafy*, Jr.*; Director, Proficiency Testing Service, American Association of Bioanalysts

Max Williams*;* Chief, External Affairs and Global Programs, COLA

James Winkelman, MD; Professor Emeritus, Harvard Medical School; Former Clinical Laboratory Director, Brigham and Women's Hospital, Boston

Ex-Officio Members – Non CDC

Sousan S. Altaie, PhD; Science Policy Advisor, Food and Drug Administration (FDA), Department of Health and Human Services (HHS)

Raelene M. Perfetto; Centers for Medicare and Medicaid Services (CMS), HHS

Kathleen J. Todd; CMS, HHS

Ex-Officio Members – Division of Laboratory Systems, CDC

D. Joe Boone, PhD

Devery Howerton, PhD

Adam Manasterski, PhD

Shahram Shahangian, PhD

Julie R. Taylor, PhD

Dan Tholen, MS

* Ms. Cendejas served as an alternate for Mr. Greg Cooper.

** Ms. Gabriel was an alternate for Ms. Margaret Peck.

CHAPTER 2. BACKGROUND

Brief history of PT implementation in the United States

In 1946, volunteer laboratories in New Jersey, Pennsylvania, and Delaware participated in an event that helped launch national PT when they conducted blood chemistry and hemoglobin tests on 12 different samples.[1] Substantial inter-laboratory differences and a preponderance of "unsatisfactory" results reported in this survey motivated the College of American Pathologists (CAP) to develop the first voluntary PT program. PT participation became mandatory for certain laboratories during the 1960s under the Medicare regulatory provisions of the 1965 Social Security Act and the 1967 Clinical Laboratory Improvement Act (CLIA '67).[2]

Over the ensuing two decades, anecdotal reports appeared which indicated that some laboratories had adopted practices that were designed to ensure satisfactory performance on PT samples but that did not reflect routine daily laboratory practices. These practices included analyzing PT samples in replicate, reporting the mean (but not the variance) of replicates for PT results, and assigning all PT samples to the best-trained and most experienced analyst in the laboratory. There were even comparisons made of results obtained in different laboratories between cooperative participants striving to confirm the correctness of their results before submission to the PT provider. With the passage of the 1988 Clinical Laboratory Improvement Amendments (CLIA) and its implementing regulations,[3] these practices were prohibited. Laboratories must test PT samples in the same manner as patient samples.

PT is one of the essential elements of a quality management system

PT became part of everyday life in clinical laboratories as the CLIA regulations extended mandatory participation in PT for certain tests and subspecialties to all laboratories performing non-waived testing. PT is one of the three main pillars of a quality assurance (QA) program. The other two pillars are 1) qualification of laboratory workers at many different levels based on education and experience and 2) conformity with quality control (QC) and other good laboratory practices, as monitored by surveys and site visits conducted by Centers for Medicare and Medicaid Services (CMS) and accrediting organizations. Although PT is used in a regulatory manner to monitor laboratory performance, PT providers have, over time, increased their emphasis on the educational component of the PT exercise. When optimally employed, PT provides regular, periodic learning experiences for the laboratory testing personnel. Enlightened laboratory directors use PT samples and review of PT surveys as an opportunity to involve, educate, and motivate the laboratory staff at many levels. PT also reassures laboratories that they are getting the "right" results – to the extent that such reassurance can be provided by comparisons with other laboratories employing the same instruments and methods.

Ramifications of performance on PT can extend beyond conformity with inspection and accreditation requirements. For example, laboratory directors or supervisors may use an

individual laboratorian's PT performance as an element of his or her annual performance review. Competitive, for-profit laboratories may trumpet their PT performance. Even hospitals seeking outreach testing business employ successful PT performance as part of a marketing program.

Attitudes toward PT vary widely among laboratorians and responsible executives in hospitals and independent laboratories. Some regard PT as an integral component of total quality management, with substantial educational value, while others regard it as a requirement for certification with no substantial utility. Nonetheless, evidence gathered by the CAP[4] indicates that laboratory decision makers use PT data to assess the relative performance of analytical methods (the extent to which different laboratories obtain the same result on the same samples) and the relative biases of different methods. Methods that are not reproducible across laboratories or that exhibit large relative bias tend to lose market share over time unless manufacturers take steps to improve reproducibility or reduce the relative biases of their instruments.

Roles and responsibilities of regulatory bodies, PT providers, and PT participants

CLIA sets out roles and responsibilities for the Secretary of the Department of Health and Human Services (HHS), providers of PT, and laboratories as PT users. Congress assigned regulatory responsibility for laboratory quality improvement to the HHS. Within HHS, the Secretary assigned primary responsibility for administration of CLIA, including PT, to the CMS – formerly the Health Care Financing Administration. Under an interagency agreement, CDC provides technical and scientific support to assist CMS in carrying out its regulatory responsibilities. Although it has no direct responsibility for PT, the FDA is responsible for approving new diagnostic tests and for determining the complexity of these tests under CLIA.

CMS is responsible for approving PT programs as defined by the CLIA regulations, monitoring PT performance of certified laboratories, enforcing sanctions related to unsuccessful participation in PT by laboratories, developing and revising the CLIA regulations as necessary, and considering recommendations made by the Clinical Laboratory Improvement Advisory Committee (CLIAC). CDC provides technical and scientific assistance with these responsibilities and manages CLIAC.

Fifteen PT providers currently are approved by CMS (Table 2-1). A detailed listing of the tests for which each provider is authorized to provide PT programs is found at: http://www.cms.hhs.gov/CLIA/14_Proficiency_Testing_Providers.asp#

Table 2-1. CMS-Approved Providers of PT Programs for Clinical Laboratories, United States, 2007

Accutest, Inc.
American Academy of Family Physicians (AAFP)
American Association of Bioanalysts (AAB)
American Proficiency Institute (API)
American Society of Clinical Pathology (ASCP)
California Thoracic Society (CTS)
College of American Pathologists (CAP) – SURVEYS
College of American Pathologists (CAP) – External Comparative Evaluation for Laboratories
Maryland Department of Health and Mental Hygiene
Medical Laboratory Evaluation Program (MLE)
New Jersey Department of Health and Senior Services
New York State Department of Health
Department of Health, Commonwealth of Pennsylvania
Puerto Rico Proficiency Testing Service
Wisconsin State Laboratory of Hygiene (WSLH)

Both public and private sector elements of the health care system rely on PT providers to monitor the performance of clinical laboratories. PT providers issue PT samples to each subscriber laboratory, receive results of PT analyses from each laboratory, report all PT results back to laboratories, and report laboratory performance evaluated against specific scoring criteria for all CLIA-required PT tests for each laboratory to CMS. Accreditation organizations (Table 2-2) use PT as one means of assessing laboratories' competence in performing tests. PT results complement on-site laboratory review by surveyors, who examine PT performance over a series of samples for patterns of results that may indicate systematic performance problems.

Table 2-2. List of deemed accreditation organizations under CLIA

AABB (formerly American Association of Blood Banks)*,* 8101 Glenbrook Road, Bethesda, Maryland 20814-2749
American Osteopathic Association (AOA)*,* 142 East Ontario Street, Chicago, Illinois 60611
American Society for Histocompatibility and Immunogenetics*,* 15000 Commerce Parkway, Suite C, Mt. Laurel, New Jersey 08054
College of American Pathologists (CAP)*,* 325 Waukegan Road, Northfield, Illinois 60093-2750
COLA, 9881 Broken Land Parkway, Suite 200, Columbia, Maryland 21046-1195
The Joint Commission (formerly Joint Commission on Accreditation of Health Care Organizations—JCAHCO)*,* One Renaissance Boulevard, Oakbrook Terrace, Illinois 60181

In practice, accreditation assessors from state laboratory licensing agencies or organizations with "deemed status" to operate as agents for CMS, such as CAP, often begin the assessment process with a review of the PT results for the previous one or two years. A laboratory that has a record of unsatisfactory PT performance (i.e., fewer than 4 out of 5 values within acceptable levels of accuracy for a given analyte) is likely to be required to demonstrate that it has thorough records that document a vigorous pursuit of satisfactory explanations for those results.

CMS approval of PT programs and oversight of PT providers are intended to ensure that both PT users (clinical laboratories) and accreditation bodies can have confidence that the PT providers conduct their programs in conformity with regulations and the requirements of accrediting organizations. The CMS approval process requires PT providers to demonstrate annually how each meets the requirements listed in Subpart I of the CLIA regulations. The information for PT provider applications or re-applications is due to CMS by July 1st of the current year. Usually before May 1st of each year, previously approved PT providers are sent a solicitation letter that contains the required elements for CMS approval. The review process for previously approved PT providers includes the following steps:

1. A CMS team reviews information in submissions received by July 1 for completeness and responsiveness to requests in the solicitation letter.
2. CMS PT database reports are generated and evaluated to determine if all required PT scores were submitted to CMS within 60 days after the date by which the laboratory must report results, as required in 42 C.F.R. §493.903 (a)(1).
3. CMS also checks the Missing Score Report. This process is ongoing and occurs each time scores are submitted.
4. CMS sends unresolved questions or concerns electronically to PT providers including issues from CMS regional offices and State agencies.
5. This process is iterative and continues until all concerns are resolved.

CMS review and approval is usually completed, and approval letters are transmitted to PT programs prior to December 31st of that year. A list of approved PT programs for the calendar year is posted on the CMS website.

The PT process as currently practiced in the United States

CLIA and the HHS regulations implementing the Amendments set forth requirements for PT programs including the number and frequency of samples, a list of specific analytes, and specific criteria for evaluating performance in PT.[3] For most analytes, regulations stipulate that PT programs must include at least three events annually, each of which must include five samples. With the exception of transfusion medicine testing where the only passing score is 100%, a "passing" PT score (satisfactory performance) requires that 80% (4/5) of a participating laboratory's test results fall within a specified range of analytical precision, defined either in relation to a "known" value established with a reference method or in relation to the values measured by a peer group of laboratories or a group of reference laboratories.

CLIA directed the Secretary of HHS to "establish criteria for acceptable performance under a proficiency testing program…for all [laboratory] examinations and procedures." and "…criteria shall be established for all examinations and procedures and shall be uniform for each examination and procedure." (42 U.S.C. §263(f)(3)(B)). In developing regulations to implement this directive, HHS identified analytes in each of 16 clinical laboratory specialties and subspecialties and developed specific PT criteria for each of these analytes. The list was promulgated in 1992. In the intervening 15 years, many additional diagnostic and monitoring tests for analytes that did not appear on the 1992 list have come into common or routine clinical use. Many additional tests for what might be described as uncommon or esoteric analytes have become available, and entire new fields of clinical laboratory testing (e.g., molecular genetic testing) have come into general use. Some of the analytes on the 1992 list (e.g., lactate dehydrogenase [LDH]) are no longer widely used. PT providers routinely offer programs that cover many of these "unregulated analytes." Since the list of regulated analytes was promulgated, the number of unregulated analytes has grown steadily, so that the unregulated analytes now outnumber those that are regulated. To meet the CLIA requirement for QA of non-regulated analytes, these programs offer a minimum of two events per year, but more frequent intervals are also offered. Some nongovernmental accrediting agencies (e.g., CAP) require participation in PT for unregulated analytes when it is offered, in effect extending CLIA requirements for the laboratories that choose these accreditors. Typically, programs for unregulated analytes involve fewer samples (one to three) at more frequent intervals than do those for regulated analytes. Evaluation criteria for unregulated analytes vary substantially among providers.

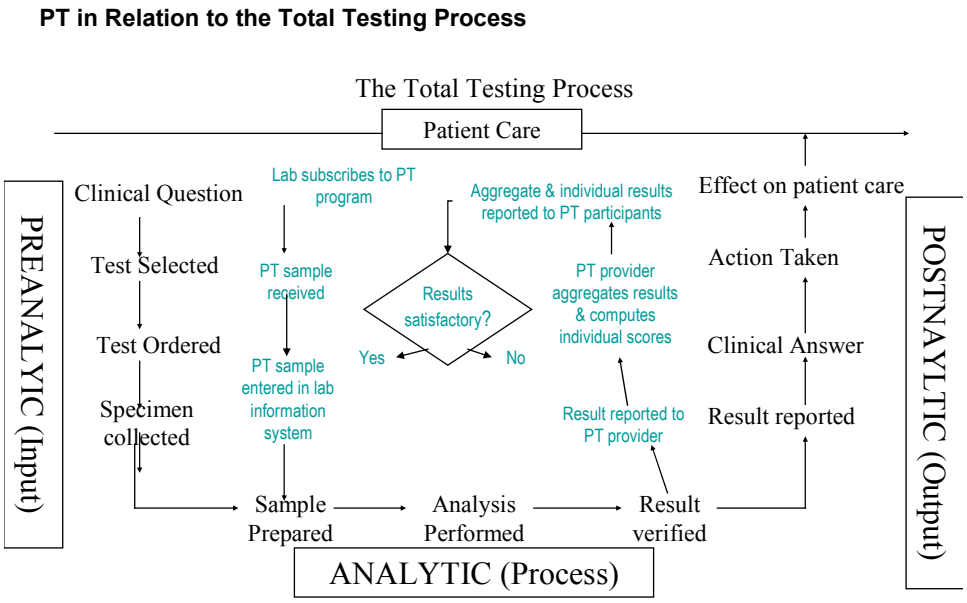
Limitations of PT

PT does not always measure or reflect accuracy, per se (i.e., the proximity of the subject laboratory's test result to the "true" value of the analyte being tested, as measured by a

"gold standard" reference method). In general, PT does not employ reference materials or methods but assesses an individual laboratory's performance in relation to results obtained by a group of referee laboratories or "peers," defined by instrument and method, into which each laboratory interpolates its results. In cases where one or several instruments or methods exhibit(s) a systematic bias with respect to a "true" value or all-method mean, the use of peer group means to evaluate PT performance may have substantial implications for quality of patient care, especially if monitoring of patient response to treatment is critical to disease management. For example, monitoring of low density lipoprotein (LDL) cholesterol in management of hyperlipidemia and of hemoglobin A_{1c} (HbA_{1c}) in management of diabetes are situations in which several different test methods exist that exhibit a bias with respect to values determined by a reference method.

PT focuses specifically on the analytical process in the laboratory and does not address the many other sources of error or inaccuracies that may adversely affect the quality of laboratory services. PT typically does not address certain pre- and post-analytical steps in the testing process – even those that are wholly under the control of the laboratory (Figure 2-1).

Figure 2-1. Proficiency Testing in the Total Testing Process for Clinical Laboratories



Another limitation of PT is that it may not allow laboratories to detect and correct problems affecting quality in a timely way. An underlying assumption concerning the validity of PT as a measure of laboratory performance is that samples are treated in the same way as patient samples. In practice, several factors can contribute to differences in

handling of PT samples versus routine patient samples. PT samples are usually in different containers, may be labeled differently, and often must be "flagged" in the laboratory's electronic information system to avoid creating a patient record. PT results, of course, are also reported to the PT provider, rather than to an attending physician, house staff, or client physician office or clinic. Also, some PT materials come as lyophilized substances – requiring reconstitution while assay of the analyte in the biological matrix does not.

Electronic (Web-based) reporting of results is now offered in some PT programs, but in many smaller hospitals and reference laboratories, technicians may not have Internet access and must report PT results by mail, courier, or fax using paper reporting forms. Not only does this slow reporting of PT results, it also increases the likelihood of PT failures caused by transcription and other clerical reporting errors. Delays in reporting summaries of results occur less commonly as electronic reporting becomes the norm, but even in 2007 anecdotal reports persist from laboratories that find they must conduct analyses of new PT samples before they receive summary reports from the previous round of PT. If the interval between PT testing and receipt of reports from the PT provider is weeks or more, the results may have little value in helping identify the cause of a failure. For identification of quality problems in time to avoid reporting of incorrect results on patient samples, laboratories rely on daily QC and QA processes.

One of the most conspicuous limitations of the United States' approach under CLIA is the exemption of waived testing from PT requirements. Some laboratorians believe that testing errors with adverse consequences for patient care are more likely to occur in laboratories that perform only waived testing because these laboratories are not routinely inspected and do not have to meet any specific quality requirements other than to follow the manufacturer's instructions for performing the tests. Proponents of extending PT requirements to the currently waived tests argue that PT is the single available indicator of laboratory test accuracy. An example of the potential consequences for patients of errors in performing waived tests is provided by the most commonly performed waived test, blood glucose concentration. Based on an erroneous glucose result, a clinician could administer or withhold diabetic medications, either of which could have serious consequences for the patient.

Attributes of a good PT program

From the point of view of laboratories using PT programs, desirable attributes of such programs can be grouped into managerial considerations, arrangements for receipt of samples, procedures for processing and analysis of PT materials, submission of data, and distribution and presentation of results.

Managerial considerations

For some laboratory managers, particularly for those who run reference laboratories that specialize in less frequently performed or esoteric tests, finding a set of PT offerings that

matches their needs can be a major challenge. Providers typically "bundle" PT offerings into modules, and a given module may include some analytes for which the laboratory does not perform testing. To meet CLIA requirements for regulated analytes, laboratories must use a PT program offered by a CMS-approved PT provider (Table 2-2). These providers also offer PT programs for many analytes that are currently unregulated, as do other providers who are not approved to offer PT for the regulated analytes (e.g., because they are for-profit entities). Thus, for some unregulated analytes, laboratory managers may have a wider array of choices for PT programs and providers.

An important consideration in selection of a PT provider is the number of participants. Larger numbers of participants are generally better, on purely statistical grounds. The larger the group is, the smaller the variance of the group mean. Groups of 20 or more are necessary for stable means.

Turnaround time for PT results is also an important factor for reasons discussed in a preceding section. The value of PT for quality management purposes diminishes asymptotically as turnaround time increases. Ideally, reporting of results should be immediate (within one or two hours of submission), so that the likelihood of identifying the sources of deviant results is maximized. This consideration is a strong argument for electronic reporting of results by PT program participants, electronic dissemination of results by providers, and on-line (Web-based) access to complete performance reports.

Processing and analysis of PT materials

For some tests (e.g., therapeutic drug monitoring), many PT providers use lyophilized materials that require reconstitution. Some laboratorians, given a choice, prefer to avoid PT programs that require reconstitution of materials on the grounds that such programs are even less like routine laboratory procedure than typical PT exercises and introduce multiple additional opportunities for errors in dilution and handling. For some analytes, no alternative to reconstituting test materials exists.

A special consideration for PT users with complex operations (e.g., laboratories with multiple sites that conduct blood glucose testing under a single PT program), is the availability of an extended period (one to two weeks) for submission of PT results. Clearly, this requirement may conflict with the previously discussed desire for prompt turnaround of results by the PT provider. Such special considerations illustrate the desirability of a variety of offerings in PT programs rather than a rigid adherence to a generalized approach.

Submission of data and distribution of PT results to users

As previously noted, Web-based electronic submission of PT results to providers offers several advantages in turnaround time, minimization of transcription errors, and utility of findings for educational and quality management purposes. Important characteristics of

the user interface in electronic submission are simplicity and ease of use. Equally important is a feature that permits the user laboratory to review results before they are sent to the PT provider.

Appropriate graphical display of the user laboratory's performance results in relation to the peer group can substantially enhance their value for quality management and educational purposes. Compared to a simple tabular presentation of results, an appropriate graphical presentation conveys much more information and is more easily understood and acted upon by laboratories.

CHAPTER 3. PT WORKING GROUP TASK REPORT

As discussed in the preceding chapter, PT is one element of external quality assurance, along with accreditation assessments and is an essential component of a comprehensive quality management system. Although PT is mandated by CLIA, the effectiveness of PT programs in meeting the regulatory, quality improvement, and educational intent of the law has not been evaluated from a policy perspective. An underlying, but largely untested, assumption in the CLIA regulatory scheme is that PT performance is a valid predictor of laboratory performance in routine testing of patient samples. From a policy perspective, additional questions might be asked about the uniformity of PT programs and the ability of these programs to stimulate improvements in test performance, to enhance the educational value of PT, and to assure that these programs provide samples that keep up with new laboratory technologies and methods as they emerge.

The charge of the PTWG encompassed five specific tasks:

1. To evaluate PT programs used by U.S. laboratories to determine the extent to which these programs meet quality improvement, regulatory, and educational expectations.
2. To evaluate the need to improve the use of PT programs and materials by laboratories.
3. To evaluate ways to enhance the educational value of PT programs.
4. To evaluate how well PT programs are providing samples to stay abreast with advances in laboratory testing technology.
5. To determine if accreditation to an international standard would improve the uniformity of PT programs.

To address this charge, CDC formulated a series of questions to guide the discussion of the PTWG and to facilitate the collection of input from the larger community of PT stakeholders. These questions appear as subheadings in the following discussion.

TASK 1: TO EVALUATE PT PROGRAMS USED BY U.S. LABORATORIES TO DETERMINE TO WHAT EXTENT THESE PROGRAMS MEET QUALITY IMPROVEMENT, EDUCATIONAL, AND REGULATORY GOALS.

What evidence is available that PT is effective?

Evidence concerning the value of PT programs in improving laboratory performance comes from a broad base of experiential evidence and anecdotal reports that indicate participating in PT improves laboratory performance. There is, however, no single study, much less a body of published evidence, which unequivocally demonstrates that

participating in PT reduces the rate of errors in routine testing of patient samples. Despite years of experience with PT in clinical laboratories, the substantial regulatory apparatus required to administer the system, and the costs of PT, evidence concerning its effectiveness has not been pursued in a systematic fashion. This lack of clear evidence, which is characteristic of many aspects of laboratory medicine, is especially problematic in assessing the benefit of PT to medical care quality or patient safety in relation to the cost of PT to the health care system.

Experience with PT unquestionably reduces PT failure rates. For example, data gathered by CDC after promulgation of the CLIA regulations showed broad improvements in PT performance by regulated laboratories.[5] Other studies also indicate that PT performance improves as laboratories gain experience with PT participation.[6-9]

These improvements are not merely evidence that practice makes perfect; several systemic effects also contributed to the reduced failure rates[7]: (1) elimination of chronic poor performers from the pool of laboratories participating in PT for a given analyte, or correction of chronic problems by laboratories that remained in the pool, (2) improved PT materials and report forms (3) familiarity with program by participants, (4) identification of problems with methods and their correction, (5) adoption of more accurate and reproducible methods, (6) generally improved technical education and technical performance. The relative contributions of these mechanisms are not completely clear. The first three factors (1-3) provide evidence of improvements in PT performance but do not necessarily indicate concomitant improvement in the routine testing of patient samples. In particular, the role of elimination of poorly performing laboratories through self-selection is unclear (i.e., laboratories that cease performing a test for a given analyte rather than accepting the consequences of continued PT failure for that test). The fourth through sixth factors, however, clearly support the notion that PT participation leads to improvements in laboratory quality that are likely to translate into improved performance in routine testing. In laboratories that fail PT challenges or experience "near misses," investigation of PT results that are chronically worse than those of peer laboratories using the same method may turn up a systemic problem that, once corrected, will also improve the accuracy of routine testing. In other cases, chronically poor PT performance may motivate a laboratory manager to adopt a new instrument or method that is intrinsically more accurate or reliable. Improved education and performance may be the result of hiring more qualified personnel but may also result from an effective use of PT results and performance to educate laboratory personnel. A review of performance in CAP PT programs provides some support for this interpretation of the mechanisms by which PT participation improves routine testing. In the analysis, laboratories showed a consistent and statistically significant improvement in performance for the first three to four years of participation – in a fashion consistent with a learning curve for PT per se. Nevertheless, the PT error rates for these laboratories continued to decline, at a slower rate, in a fashion consistent with the more systemic mechanisms described.[10]

True quality improvements may take place in response to PT failures. One CAP study reported that, when laboratory directors reviewed the causes of repeated unacceptable PT results, 50% of the investigations isolated and corrected problems related to

instrumentation, methods, or other technical aspects of testing. Only 4.4% of the failures were attributed to the survey format or materials.[11] A 1994 study by the Wisconsin State Laboratory of Hygiene (WSLH)[12] found that laboratories demonstrated a pattern of improved performance after unsatisfactory PT performance brought attention to correctable problems. Meeting the CLIA PT requirements has been a prime motivator in improving laboratory performance.[13]

Frequently reported sources of analytic errors in clinical chemistry laboratories include system instability resulting in calibration drift and random errors.[13] While improved instrumentation and reagents are important in long-term improvement of testing accuracy, often PT performance can be improved by using current instrumentation according to manufacturers' recommendations. For example, in one study of causes of unsatisfactory performance in PT,[14] more than 50% of the laboratories used an "allowable error" for routine QC of analytical systems that exceeded the manufacturers' recommended threshold error limit for stable instrument performance. Mishandling sample dilutions was another major source of analytic error, a problem easily correctable with additional analyst training. Investigation of PT errors can uncover such inadequacies in the laboratory's QA program and lead to improvements in laboratory PT performance and patient data quality.

Historically, hospital laboratories and independent laboratories as a group have performed substantially better than POLs in PT.[5] Using PT data reported to CMS by PT program providers, a CDC review of the PT performance for laboratories in CLIA-mandated PT programs demonstrated improvement in performance for most of the analytes/tests evaluated over a 12-year period (1994–2005) (unpublished data). In 1998, Stull[15] found disparate PT performance between traditional laboratories (hospital and independent laboratories [HI]) and alternative testing sites (all other testing laboratories [AOT], including POLs and ancillary health care providers).The aggregate rate of satisfactory test event performance for all regulated analytes, tests, and specialties was 97% for the HI group and 91% for the AOT group. In the ensuing decade, a broad and general improvement in PT performance has occurred in all types of laboratories, including smaller hospital laboratories and POLs. A 2007 publication by the American Proficiency Institute describes a 10-year study of PT performance by physician's offices, clinics, and small hospital laboratories.[16] Failure rates for chemistry and hematology analytes declined significantly during the 10-year period. Failure rates for microbiology also declined but remained above 5% in 2004 for certain tests.

Several studies illustrate the value of PT as part of a proactive, integrated approach to laboratory quality management. Johnson[17] reviewed various approaches that can improve laboratory quality and, in turn, improve PT performance. Laboratories can take proactive steps such as narrowing their QC ranges, increasing the frequency of calibration, performing instrument function verification, and examining PT results closely for trends and bias, even when they are deemed acceptable. Several studies indicate that laboratories that incorporate these approaches into their quality management system and that routinely monitor their analytic performance instead of simply correcting PT errors,

not only reduce PT failures but also produce more accurate patient test results over the long run.[7,14,18]

How can we obtain new evidence of the effectiveness of PT?

Despite the broad consensus among laboratorians regarding the effectiveness of PT in evaluating and improving laboratory performance, additional comprehensive and systematic studies are needed that assess directly the linkage between PT performance and performance in routine testing of patient samples. Such studies are not merely of academic interest. Planned, prospective studies of a broad range of laboratories of all types are needed. These studies could track performance of the same group of laboratories over time to identify predictors of PT performance and effective strategies and practices for integrating PT performance into a broad quality management system. Monitoring data for the same group of laboratories would also help distinguish true improvements in PT performance and error rates in routine testing from artificial improvements due to drop outs by poorly performing laboratories. These surveys could identify areas in most need of improvement or additional research and analysis.

New studies should examine several aspects of PT performance. Laboratory performance data should be more specific than simply PT failures, including all unsatisfactory results and perhaps standard deviation indexes or SDIs (result-mean/SD). More specific performance data would allow tracking of marginally performing laboratories, not just failing laboratories. A significant limitation of studies relying on the CMS PT database is that CMS does not track or report the reasons for failure. PT failures that occur because of a poorly designed reporting format cannot be distinguished from "true" analytical errors. From the standpoint of assessing compliance, this distinction may not be important – to the patient whose condition is misdiagnosed, it makes no difference whether the source of error is analytical or clerical. If data are to be used to monitor long-term trends in performance of types of laboratories or PT providers, however, data on the reasons for failure are essential. Any new study should determine the reasons for failure, requiring laboratories to self-investigate and report specific causes of unacceptable results. The CAP Laboratory Accreditation Program (LAP) presently has such requirements in place. LAP identifies four broad error categories: methodological problems, technical problems, clerical errors, and problems with PT materials. Each of these has four to nine subcategories of error that specifically describe the problem that caused the unacceptable value. CAP used these specific error data to detect and correct the major causes of systematic laboratory problems in marginally performing laboratories.[19] In future studies the link between internal QC and external QC (PT) should also be studied to determine the importance of establishing QC ranges in predicting a laboratory's PT performance.

What criteria can be used to evaluate a PT provider's performance?

The value of PT might be enhanced if CMS could expand its oversight of PT programs by requiring and holding PT providers accountable to certain quality standards such as

the International Laboratory Accreditation Cooperation (ILAC) Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes.[20] Some of these requirements would likely demand no more than a formalization and documentation of policies and procedures that many PT providers already have in place as a part of their own programs of good business practices and good laboratory practices. Others may require more extensive changes in present practice. Under the system currently in place in the United States, a presumption of competence among approved PT providers exists, but providers are not required to provide positive, objective evidence of competence. These points are discussed in more detail later in this chapter (see Task 5).

The lack of explicit quality and competency standards for PT providers is an inconsistency in the current U.S. system of clinical laboratory regulation. Clinical laboratories are subject to inspection by CMS or a "deemed" accrediting agency, must participate in PT performance evaluations as a condition of certification by CMS, and must have in place a QA program. PT providers do not have to meet similarly rigorous standards. CLIA is silent on the topic of accreditation, quality standards, and on-site assessments for PT providers, providing neither a requirement nor a mechanism through which the Secretary of HHS could enforce adherence to such requirements. Compulsory adoption of accreditation standards or imposition of specific competency or quality standards for PT providers would require a change in regulation, but PT providers are free to adopt standards (e.g., International Organization for Standardization (ISO) Guide 43) voluntarily. Again, the potential impact of adopting international standards is discussed more extensively later in this chapter (see Task 5).

Some steps to provide PT users with additional information about PT provider performance can be taken without any changes to CLIA law or regulations and with few additional resources. CMS, state programs, accrediting agencies, and PT providers themselves presently receive complaints from PT users, but this information is not formally compiled and reported to PT users. CMS and State programs can adopt explicit policies and procedures for receiving, investigating, and reporting on complaints about PT providers. Such information would help PT users make informed decisions regarding their choice of PT provider. At a minimum, records should be maintained of all complaints and of the investigations and corrective actions taken by the PT provider. These complaint processes should be publicized and communicated to PT users. A publicly accessible database of records of complaints, responses, and corrective actions should be made available to PT users and other interested parties.

Task 1 Recommendations

- Undertake a systematic assessment of the relationship of laboratory performance on PT samples and laboratory error rates in routine testing of patient samples and measures of patient outcomes and quality of care. This assessment should be based on primary data collected over a period of years in a group of collaborating laboratories and health care institutions including acute inpatient hospitals, ambulatory care, emergency departments, and long-term care settings.

- Conduct a study of the existing information in the scientific literature and in current databases regarding reasons for unsatisfactory PT results to identify areas most in need of improvement or additional research/analysis.
- In addition to routine reports to PT participants, PT providers should periodically publish PT results in appropriate independent peer-reviewed journals.
- Develop a process to collect and consolidate all complaints received by CMS, state programs, accrediting organizations, and PT providers about PT; develop an appropriate statistical analysis of collected data to identify trends that may be subject to corrective action; and publicize and disseminate the complaint process for widespread use by all parties.
- Develop and make available a database to collect PT data for characterizing the performance of all laboratories, identifying reasons for unsatisfactory PT results, reviewing acceptance criteria used by PT providers, and identifying a list of analytes that should be regulated.
- Develop a process to assure that all clinical laboratories, including those that perform waived tests, engage in an external quality assessment program that may include PT.

TASK 2: TO EVALUATE THE NEED TO IMPROVE THE USE OF PT PROGRAMS

In addressing this issue, the Working Group considered gaps in current PT program content; potential improvements in the design, integrity, and distribution of PT samples; and possible approaches to improving the evaluation (grading or scoring) of PT results.

What are the gaps and needs in current PT program content?

Program content

The language of the CLIA statute: 42 U.S.C. §236a (f)(3)(A) and (f)(3)(B) appears to require PT for all tests except those for which the Secretary determines that a proficiency test cannot reasonably be developed.* Since the HHS promulgated regulations implementing CLIA in 1992, many tests for new analytes, as well as entirely new analytical technologies, have come into use in clinical laboratories. The list of regulated analytes has not been updated to include these tests. Conversely, some tests that are on the 1992 list are obsolete (e.g., LDH) and are no longer routinely performed to support clinical care. As newer "unregulated" analytes were adopted by clinicians, PT providers began to include the tests in PT programs. PT for these unregulated analytes has achieved wide acceptance. To illustrate the scope of this gap between the current list of routinely performed moderate and high complexity testing and the list of regulated analytes, CAP's LAP in 2005 monitored 423 graded analytes, of which 97 are on the CMS list of "regulated" analytes.[19] Some of these tests are for "critical" analytes that can have substantial impact on patient care decisions. Examples include free calcium, lactate, troponins, myoglobin, "brain-type" natriuretic peptide (BNP) and other cardiac markers, most tumor markers, activated clotting time, and d-dimer, as well as prostate specific antigen (PSA), and Hb A_{1c}. Replacement of some of the analytes on the regulated list with new ones, especially if the tests are critical, is overdue. Current regulations (42 C.R.F. §493.1709 (b)) do require laboratories to assess the accuracy of their tests for unregulated analytes at least twice a year and many laboratories use PT for this purpose.

Laboratories that participate in PT for unregulated analytes do so for a variety of reasons. Some are required to participate in non-regulated PT by their accreditation organizations. For instance, laboratories accredited by CAP agree to enroll in PT for tests they perform that are on the expanded CAP LAP list. Other laboratories enroll in PT for unregulated

* The relevant language is in (A); "The standards shall require that a laboratory issued a certificate under this section be tested **for each examination and procedure conducted within a category of examinations or procedures for which it has received a certificate**, except for examinations and procedures for which the Secretary has determined that a proficiency test cannot reasonably be developed"; and in (B): "The standards established under subparagraph (A) shall include uniform criteria for acceptable performance under a proficiency testing program, based on the available technology and the clinical relevance of the laboratory examination or other procedure subject to such program. The **criteria shall be established for all examinations and procedures** and shall be uniform for each examination and procedure."

analytes independently of explicit requirements by accreditation organizations, as part of their EQA (external quality assessment) or total quality management programs.

The fact that the process for adding regulated analytes is cumbersome may be one reason the list of regulated analytes has not changed. A preferred approach may be to develop a more flexible system that would allow an authoritative body, such as CDC or CLIAC, to review, evaluate, and recommend adjustments to the list annually or biannually. This approach is currently in use for providers of PT to laboratories that conduct tests for environmental contaminants under EPA regulation. An alternative to a governmental agency performing this task would be to have a list of analytes managed and regularly updated by a respected professional body. Although an alternative approach such as this may be advantageous, it would require a change in the CLIA regulations.

More challenging tests

CLIA regulations stipulate that analyte levels in PT samples should reflect clinically abnormal ranges as well as values in the clinically normal range. Non-consensus in PT results is more likely to occur when test values are near the extremes of the dynamic ranges of analytical instruments. PT providers therefore tend to keep sample values well within typical operating ranges of instruments used. This situation becomes problematic when clinically plausible abnormal values coincide with values at one extreme or the other of the usual operating range of the instruments used by a substantial fraction of the members of a peer group.

From the perspective of the clinician attempting to make decisions about how to manage a patient's care, plausible but extreme values may not be particularly relevant. Accuracy in a relatively narrow range around critical decision points may be more important, and in most cases, these decision points are unlikely to fall near the extremes of the dynamic range for the instrument. On the other hand, critical decision points for clinical care may coincide with extremes of the dynamic range of the relevant instrumentation. Such a situation may arise, for example, with blood glucose levels in an unconscious diabetic patient, especially in the case of an infant in a Neonatal Intensive Care Unit (NICU).

PT event frequency and number of samples

Implementation of CLIA requirements resulted in a change in the frequency for most PT testing events from four times to three times per year and an increase in the number of samples from two or three per testing event to five per event. At the time the regulations were formulated, HHS personnel believed that PT providers would eventually have to provide programs for up to 600,000 laboratories, making quarterly PT more costly and less feasible. The number of samples per PT event was increased to improve the statistical comparability and stability of comparisons across peer groups. Although the actual number of subscribing clinical laboratories never came close to the early estimate (approximately 30,000 laboratories currently participate in PT programs offered by approved providers), the number of events has remained at three. While it would now be

possible to increase the frequency, more frequent PT events could still be logistically and financially burdensome to both PT providers and participants.

Even when providers are not bound by CLIA regulations (i.e., for unregulated analytes) the frequency of PT events offered by U.S. PT providers rarely exceeds three per year. The number of samples per event is usually five but it can vary from one to seven. In contrast, a wide range of testing frequency and number of samples is used internationally. One European PT model is to challenge participants more often, up to once per month, but with only a few samples in each event. Shipping costs are sometimes minimized by bundling several samples into a single shipment. For example, the European Research Network for Evaluation and Improvement of Screening, Diagnosis and Treatment of Inherited Disorders of Metabolism (ERNDIM) ships 8 samples for its quantitative scheme for amino acids in January for testing throughout the year (see http://www.erncimqa.nl/InfoFrame.php?PHPSESSID=42321ccele72bac4b7a293cf8a264ad3). Some provinces in Canada require three samples, six times per year. The International Union of Pure and Applied Chemistry (IUPAC) international PT protocol leaves the frequency up to the PT provider and suggests that it will normally be between two and ten testing events per year.

What improvements are needed in the design of PT samples, the integrity of samples being distributed, and PT distribution?

Design of PT samples

PT sample panels generally reflect the make up of testing services offered by laboratories that comprise the customer base of the PT provider. Inevitably, some mismatching occurs between the PT panels offered by providers and the testing services performed by participating laboratories. Some laboratories are obliged to purchase PT panels that include samples for analytes or tests these laboratories do not perform. PT providers must balance comprehensiveness that meets the needs of a wide range of laboratories against the disadvantages of including analytes that only a few potential users need.

In the view of the Working Group, no regulatory prescription to resolve this situation is possible or warranted. Communication between PT users and providers can certainly reduce the number of mismatches. Laboratories should make their panel content needs known to the PT providers, and PT providers should compile complaints and suggestions to improve panel content. PT providers should regularly review the content of their PT panels and adjust their offerings to reflect the current needs of their subscribers.

PT providers can assist subscribers in evaluating their laboratory's precision (consistency of performance) and bias (calibration accuracy) by strategic design of sample concentrations. Precision can be assessed with two samples which are close in concentration (or the same concentration), and calibration can be checked with samples across the dynamic measurement range. Some PT providers are designing PT samples

with these considerations in mind, but do not routinely analyze data from the testing for characteristics other than accuracy.

The materials that are used by PT programs should, as much as possible, behave like patient samples. Most PT materials in clinical chemistry consist of treated human serum that is spiked with one or more analytes of interest to produce the desired concentration. The artificial matrix may cause assays to perform differently with this material than with clinical samples. This difference is referred to as "matrix effect." Also, when the analytes are proteins, such as hormones and tumor markers, the proteins that are secreted or shed into the circulation may exist in more than one molecular form. Spikes containing proteins with a molecular form different from those in the serum sample could result in a "spike bias." The presence and magnitude of a matrix effect or spike bias is typically unknown for any given analyte/method combination. Consequently, the performance of a given method in an external survey may not reflect the results obtained when measuring patient samples. *

CAP has undertaken several studies to evaluate this matrix effect or spike bias.[21-23] In one such study[21], an unadulterated serum sample was fresh frozen and distributed along with PT materials spiked with five tumor markers as a part of a CAP survey. The investigators were able to determine the differences in the precision and bias in the measurements of the two types of proficiency samples. The largest contributor to the total survey imprecision was the long-term within laboratory variation attributable to reagent lot changes and recalibration. The overall bias as reflected in the peer group mean indicated that the PT material did not behave like the fresh frozen serum for three of the five tumor marker assays. Therefore, a matrix effect contributed significantly to the measurement bias. This type of study is an important means of investigating the validity of PT programs and can often uncover problems that would otherwise remain unnoticed. Studies similar to this one should continue to be undertaken not only to improve the quality of PT materials but also manufacturers' control of the quality of reagents and calibrators, both of which ultimately affect the quality of an individual laboratory's clinical laboratory testing results.

Sample integrity

To achieve the aim of PT, results of the PT samples must reflect the quality of the laboratory's test system and the operator's ability to use that system. This aim obviously cannot be achieved if the shipment includes a PT sample of compromised quality. For this reason, PT providers must work closely with their suppliers and with instrument manufacturers to assess sample matrix effects and sample stability. Typically, challenges in maintaining sample integrity and accounting for matrix effects are greatest with whole blood samples, which are stabilized using various preservatives. These preservatives may affect the results of the analysis. PT providers also must ensure that samples are homogeneous and ship samples expediently. This is a constant challenge to the PT supplier and provider. To avoid loss in sample integrity, some PT providers set shorter

*The list of analytes for which spiked samples or artificial matrices are the only practical options for PT samples is growing because human disease-state materials are increasingly difficult to obtain.

reporting deadlines for certain sample matrices (fresh blood or fresh plasma) and analytes that are unstable.

Though rare, problems with PT sample integrity do occur. When they occur, these problems should be promptly resolved by the PT provider, and amended laboratory performance reports should be submitted to the regulatory agency, if needed. All complaints of sample integrity problems should be recorded by the PT provider, every complaint should be investigated, and the outcome of the investigations should be documented. PT providers should maintain records of the frequency of problems with sample and shipping errors.

Many laboratories do not realize that CMS requires PT providers to retain at least five percent of each lot of sample base, so that participating laboratories can request additional samples for retesting as a part of an internal investigation of an unacceptable PT score. Samples are also retained to replace those that are damaged or lost in shipping. Whole blood samples, however, are particularly difficult to preserve and may not be stable enough for retesting; therefore, retaining a fraction of the sample base for retesting may not be particularly useful to laboratories that are attempting to investigate problems with these tests.

How can the evaluation (grading, scoring) of PT results be improved?

Target values and evaluation ranges

The CLIA regulations describe the evaluation criteria used to score regulated analytes:

1. Quantitative analytes: To determine the accuracy of a laboratory's result, the PT program must compare the result for each analyte with the target value that reflects agreement of either 80% of ten or more referee laboratories or 80% or more of all participating laboratories. After the target value has been established, the PT program must determine the correct result for each analyte by the distance of the result from the target value using evaluation criteria. The evaluation criteria are expressed as the target values ± either fixed limits (e.g., ± 6 mg/dL), percentages (e.g., ± 20%), or the number of standard deviations (e.g. ± 3 SD). In some cases, the evaluation criteria consist of two criteria where the criteria that produce the greater acceptable range are used (e.g., ± 20% or ± 1.0 meq/dL, whichever is greater).

2. Qualitative or semi-quantitative analytes: To determine the accuracy of a laboratory's result, the PT program must compare the result for each analyte with the target value that reflects 80% or more of all participating laboratories. The criterion for acceptable performance of qualitative tests is either positive or negative. The criterion for acceptable performance of semi-quantitative tests is the minimum concentration that will be considered as indicating a positive result.

For quantitative tests, the "target value" is defined either as the mean of all participant responses after removal of outliers (those responses greater than three standard deviations from the original mean) or the mean established by definitive or "well accepted reference methods." In instances where definitive or reference methods are not available, or a specific method's results demonstrate bias not observed with actual patient samples as determined by a defensible scientific protocol, a comparative method or a method group (peer group) may be used. If the method group comprises fewer than ten participants, the target value is defined as the overall mean after outlier removal unless acceptable scientific reasons are available to indicate that such an evaluation is not appropriate.

Despite the requirement that peer group means be used for target value determinations only in instances where a matrix effect has been demonstrated by a defensible scientific protocol, this approach has become commonplace. While it is understandable that PT providers prefer this method since it is more likely that they will achieve 80% participant agreement, this practice can mask true analytical bias that may affect patient samples, as well as PT samples. Rej, et al.[24] reported on two cases seen in the New York State Clinical Laboratory Evaluation Program that illustrate the pitfalls of using peer group means instead of an all-participant mean. In one case, a problem in one manufacturer's reagent quality would have been overlooked. In the other case, a problem with another manufacturer's calibrators was uncovered when results for these instruments were compared to the overall mean. In a study of the accuracy of eleven routine chemistry analytes in the CAP Chemistry Survey[23], nearly 80% of inter-laboratory measurement variance arose from peer group calibration errors, not matrix effects.

For reasons such as these, target values for quantitative tests should be calculated from all-participant mean values, as stated in the CLIA regulations. Where available, target values should be established or verified by reference methods. Comparative method or peer group targets should be used more rarely and only when it is shown that specific methods demonstrate a bias with PT samples not observed with patient samples. It is recognized that changing the method of target value assignment from peer group to a reference method value could have disruptive effects both for PT providers and participating laboratories. If an instrument has a problem in providing results comparable to the "true" value as determined by a reference method or reference material, the laboratories using it should not be penalized for their choice of instrumentation. Some accommodation for this possible outcome should be made by providing a transitional period that allows manufacturers to correct problems or laboratories to change instrumentation.

Ideally, target values should be derived from reference methods and reference materials. The PTWG considered a recommendation that CDC identify reference or standard methods for as many CLIA analytes as possible and establish guidelines for manufacturers and PT providers to utilize these methods to assign target values for analytes. The group was unable to reach a consensus in this discussion, and no recommendation was made. The principal impediment to consensus was that reference or standard methods exist for relatively few clinically important analytes, and establishing such methods is often a difficult, time-consuming, and expensive effort. The scientific

bodies that approve reference methods include the International Federation of Clinical Chemistry and Laboratory Medicine (IFFC), National Institute of Standards and Technology (NIST), and the World Health Organization (WHO).

PT providers have pursued matrix-free PT materials and accuracy-based grading. Throughout the 1980s and 1990s, the CAP supported NIST in its efforts to develop comparative methods for key analytes. These comparative methods were then applied to PT results in a system of dual grading within the comprehensive chemistry survey. This practice was halted in 2000, as was CAP's support of the NIST effort because, after several years, reference methods had been assigned for only four analytes, and dual grading created confusion among PT participants.

When the PT peer group includes a small number of participants for a particular analyte or the analytical values reported by peer group members are so diverse that a "consensus value" cannot be computed, a proficiency test may not be graded. When this occurs, CLIA allows laboratories to receive a 100% score for that PT sample. Methods exist to address this problem. An example is a grading scheme developed for small peer groups (fewer than 10 members) for two tests, prothrombin time (PT) and activated partial thromboplastin time (APTT).[25] A modified target value for small peer groups was derived based on the assumption that measurement variability in PTT and APTT testing is more greatly influenced by variations in reagents than by variation among instruments. The scheme developed modified target values that were a weighted average of the mean of the peer group and the mean of all participants using the same reagent. Peer groups with as few as four members were graded provided that the following conditions were met: at least five peer groups were available for a given reagent, at least three of five had more than three members, and the coefficient of variation for the reagent group was less than 10%. When applied to data from a CAP survey, this grading scheme substantially reduced the number of ungraded tests by a range of 42% to 55% for both tests.[25] This approach is not specific to these analytes. A similar approach could be applied to small peer groups for other test methods or analytes that satisfy these statistical assumptions. In addition, robust statistical procedures are available that can process results from peer groups with as few as five participants (ISO 13528, *Statistical methods for use in proficiency testing by inter-laboratory comparison* 2005).

Since many analytes offered by PT providers are not regulated analytes under CLIA, grading of unregulated analytes is more flexible. This means that unregulated analytes do not have standardized acceptance criteria, and the acceptance criteria are left to PT providers to determine. Therefore, these criteria differ among PT providers. Standard criteria should be developed for all analytes so that a laboratory's results are scored equally regardless of PT provider. A process is needed to bring the PT providers together to agree on a consensus grading system that would be consistent across PT programs.

Graphical presentation of PT results

Improvements can be made to the reporting of PT results by using graphical presentations in addition to tabular presentations and pass/fail rates. Histograms showing the

distribution of results from other laboratories in the group have more impact to a reviewer than simple mean and standard deviation statistics. A bar graph plot of the relative distance of a laboratory's results from the target as a percentage of allowed deviation is an effective means of spotting trends and warning of potential problems. In addition, many laboratories would benefit from interpretive materials that exploit graphical representation to improve the interpretability of the information contained in PT results. For example, laboratory directors would be helped by a statistical and graphical representation of data indicating whether an unacceptable result is due to imprecision or bias.

Some Canadian and European PT programs provide highly informative reports. The EQA reports from Finland indicate the laboratory's individual result, the peer-group histogram, and the all-method histogram in the same graphic presentation. This is provided with the usual all-method and peer-group mean, standard deviation, percent coefficient of variation, and number of laboratories reporting. Additionally a Youden plot is presented with all three groups indicated in different colors, data from up to 10 previous PT events, and a performance versus concentration plot. All of these depictions allow participating laboratories to fully compare their performance to others, identify trends, and detect bias and performance weakness across the concentration range. A guideline document that illustrates how results are summarized is available on the Internet and should be considered by PT providers as a model for enhancing program effectiveness by including more instructive graphical presentations: http://www.labquality.org/LQ/(S(bxzg2xi4sdaaq455tfw3psjb))/pdf.aspx?dir=3&path=tulosten_tulkinta_eng.pdf.

Some international PT programs provide early warnings when results approach the limits of acceptance criteria. An analogous warning flag approach may be a useful adjunct for U.S. PT programs and would be valuable in reducing errors and expediting laboratory response to unacceptable PT results. A "pass with warning" result also would alert laboratories of a growing or persistent problem that might be overlooked with a simple pass/fail result.

Timeliness of PT reports and Electronic Reporting

The lack of timeliness of PT reports is a hindrance to effective use of PT in identifying causes of laboratory deficiencies. Slow turnaround can limit the usefulness of PT program data, both for QA and educational goals. PT providers should search for ways to reduce the turnaround time for PT reports. Causes of slow turnaround of PT results include:

- Some PT participants submit reports on paper by mail or fax, rather than electronically.
- Some PT providers must manually enter results submitted on paper forms or by fax.
- PT providers must collate results by test and peer group.

- PT providers must wait until a minimum number of results are received before they can identify and exclude outliers and calculate acceptable ranges for peer groups.
- Some participants receive results only by mail.

PT providers should search for ways to reduce the turnaround time for PT reports. As a first step, all PT providers should provide the option for participants to report results electronically. Eliminating the process of transcribing paper results into PT provider data systems would substantially reduce both processing time and cost and would benefit both providers and participants. It is difficult to overstate the importance of reducing turnaround time for PT results. If PT reveals weakness in a laboratory's system, it is imperative to correct it immediately. If it takes up to 60 days to receive a notification of failure, it is difficult to identify the cause of the failure and to correct it. In the same vein, systems which offer drop-down menus for methods, reagents, and instrumentation make the reporting process more convenient, more rapid, and less prone to transcription and transposition errors, further reducing the chances that participants will fail because of non-analytical (clerical) errors. Reduction of PT failures caused by clerical errors should represent an incentive for laboratories as well as PT providers.

Even when an electronic reporting option is offered by a PT provider, some laboratories insist on submitting handwritten data reporting sheets. In some laboratories, personnel are prohibited from accessing the Internet – an unnecessary limitation since relatively simple technological means exist for laboratory directors to grant limited access to the Internet. Some PT providers, such as the State of New York, are actively encouraging electronic reporting by PT users and plan eventually to require it. All PT providers should follow this example and move expeditiously toward requiring total electronic reporting. Laboratories that impose Internet access restrictions on employees should allow access to specific sites to a limited number of personnel for the purpose of submitting PT results and receiving PT reports.

For many analytes routinely used in clinical medicine, multiple analytical systems are available that report results in different units, often because underlying analytical methods differ fundamentally. When laboratories that use different instruments are part of the same PT peer group, it is common for some laboratories to find that they must convert the PT result in the "native" units of their instruments to "foreign" units used by other instrument manufacturers. The calculations needed to accomplish this conversion are a frequent source of errors in reporting PT results. Many laboratories address this problem by developing a spreadsheet to do the computations automatically before the PT result is reported to the provider. Electronic submission forms can easily address this limitation. This approach substantially reduces opportunities for transcription and calculation errors in reporting.

Ideally, a system should be developed to allow laboratory and clinical information systems to communicate directly (interface) with the PT providers. Such a system would, however, be difficult for both the PT provider and the laboratory to implement as it would require an intermediate standardized platform into which the laboratory would

export results and from which the PT provider would import the results. Health Level Seven (HL7) is a flexible data standard that can enable the exchange and interoperability of electronic health records. In this system, a platform of standardized fields is being established which could potentially be utilized for reporting PT results. The use of this system would still require laboratories to provide software to export data into the standardized fields, but this capability could also be offered by laboratory information system (LIS) providers as a value added option. Because of these hurdles, the use of a system for direct reporting of results is a long term goal for reducing administrative error in PT reporting. Reducing administrative errors could in turn reduce time for reporting results, and could help make participation in PT more similar to the testing of patient samples.

Task 2 Recommendations

- CMS should develop a process to periodically review, update, and publish the requirements of the CLIA PT program, including the list of regulated analytes and allowable limits. Alternatives to the current CLIA PT scoring schemes should be evaluated.
- PT providers should evaluate alternatives to the current CLIA PT frequency and number of challenges.
- PT providers should provide samples that mimic patient samples as much as possible with a minimum of artificial matrix effects.
- PT providers and laboratories participating in PT should collaborate to conduct small adjunct studies using fresh frozen samples from a single patient in conjunction with routine PT to identify and characterize testing problems that would otherwise go undetected.
- PT providers should develop a system for electronic submission and reporting.
- PT providers should report PT results to end users graphically in a manner that is easy to read and understand.
- PT providers should seek ways to provide a faster turnaround time and more detailed feedback on test results.
- PT providers should allow for the reporting of analyte results in various units of measure and be able to convert and report those measures to common units in their reports in accordance with current regulations.

TASK 3: TO EVALUATE WAYS TO ENHANCE THE EDUCATIONAL VALUE OF PT PROGRAMS

The experience of the Wisconsin State Laboratory of Hygiene (WSLH) in administering their PT program provides an example of the educational benefits of PT prior to and after the advent of CLIA regulations. In the 1970s and 1980s, the WSLH provided education opportunities for participants in its PT program. Instrument manufacturers also participated in the Wisconsin program. The educational program effectively brought together instrument vendors and users to solve problems. For example, the program revealed matrix effects for some analytes when using certain analytical platforms and instruments, and also demonstrated that some instruments could not meet generally accepted performance criteria. WSLH worked with instrumentation manufacturers and encouraged these vendors to enroll in WSLH PT programs for purposes of identifying method- or instrument-specific analytical idiosyncrasies. Examples of information obtained through this cooperation with manufacturers include instrument performance with specific matrices, linear ranges and other performance characteristics of instruments, and calibration standards.

A more recent example of an educational PT that was effective in improving laboratory testing quality was a Wisconsin PT educational program that partnered with the CAP during the 1990s to address poor laboratory performance in Lyme disease testing. The program consisted of five case-defined and/or normal serum samples for detection of antibody to *Borrelia burgdorferi.* The results were scored and reported back to the participants. Since the program did not fall under the CLIA regulations, the scores were for informational purposes only. Each PT event summary provided educational material for each specific Lyme disease case study as well as solutions for testing problems identified by participants. This program significantly improved the performance of the individual laboratories and the mean score across the whole participant group, an outcome that was particularly beneficial because laboratory performance on this test may have a significant impact on patient health.

Laboratories must balance the benefits of participating in educational PT and the costs of such activities when they are not required by regulations. Ideally, all laboratories should be interested in participating in educational PT surveys for tests or specialties they regard as critical to patient health. Educational PT represents one of the best means of improving quality in the laboratory. PT programs should go beyond the important role they play in the regulation of laboratories. They should also serve as a vehicle to educate and prepare laboratories, not only for new technology and specialized areas of testing, but also for future PT events that will assess their capability to respond to needs that have not yet been clearly defined.[26] If PT providers are interested in developing samples to keep up with advances in laboratory testing technology, they will need to provide more samples for which grading is done purely for educational purposes.

What are the best ways to enhance educational value of PT and to evaluate effectiveness of any changes?

Unfortunately, opportunities for education often occur after a laboratory is notified of a PT failure. Investigations of PT failure involve gathering and reviewing data to identify the root cause of the failure and then taking appropriate remedial action. The Clinical and Laboratory Standards Institute (CLSI) GP-27-A2 document, "Using Proficiency Testing (PT) to Improve the Clinical Laboratory," provides an approved guideline for remedial actions.[27] The entire investigative process should be documented.

When failure investigations are educational in nature, they are more likely to inspire a positive response among the participants if undertaken as part of a proactive approach to continuous quality improvement. Education can take the form of formal continuing education courses or the inclusion of additional comments or explanations on reports. Presenting PT results in an instructive manner, rather than as a simple pass-fail dichotomy, is also educational. None of these educational options are required by the CLIA regulations. PT providers are not required to offer educational materials along with PT evaluation results. The ILAC guidelines for the requirements for the competence of providers of PT (ILAC-G13)[20] do not mention a requirement for PT providers to offer educational materials, but to add value to their services several PT providers provide educational materials for use in improving laboratories' PT performance. For example, the American Proficiency Institute (API), the American Academy of Family Physicians (AAFP) PT program, and the American College of Physicians' Medical Laboratory Evaluation Program (MLE) offer free continuing education (CE) credit through self-assessment activity to physicians, medical technologists, and technicians enrolled in their PT program. In API's program, up to three CE credits per year can be earned in hematology, chemistry, microbiology, and immunology by each participant. Questions in this self-assessment activity focus on ASCP commentaries related to the samples from each test event. CAP also offers survey education at no charge for staff from laboratories participating in select CAP survey products. CE credits can be earned for online activities that are intended to enhance existing surveys offered throughout the year.

PT reports that are more graphical in nature make them more understandable and provide more educational value. Cumulative graphical reports can provide a visual indication of bias or systematic trends of tests above or below the mean, a warning that a test is trending toward unacceptable limits, or an indication of random error that a test is exhibiting. As an example, CAP provides hard copy and online PT reports that present graphical cumulative reports of the three most recent surveys. The online reports show the results plotted as the relative distance from the mean in color, allowing for rapid interpretation. All PT providers should offer comments on the reports, based on trends or types of identified errors, to provide assistance to laboratories in troubleshooting and preventive measures (see Graphical Presentation of Reports in Task 2). The benefit of such enhanced reporting is well recognized. International reports such as PT programs in Finland and the UK provide examples of more educational reports. Additional examples are available in published standards documents (e.g., the previously cited CLSI GP27[27] and ISO 13528).

PT can be used for evaluating staff competency and training. All laboratory personnel should have the opportunity to analyze PT samples, not just the top performing analysts. As mentioned above, the process of performing PT, reviewing results, and taking corrective action is an educational process in itself. Alternatively, laboratory personnel may reanalyze samples and compare the results to the values that have been published. In this way, new personnel may be introduced to the process, supervisors can assess their competency, and instruction can be applied without risk of regulatory failure. At the very least, the personnel performing PT need to understand why they are doing it. This lack of understanding may be more of a problem in waived laboratories where PT is not required by regulation. A brochure explaining PT and how to evaluate results is particularly needed for analysts in these laboratories.

The educational value of PT is reinforced by the fact that new medical technologists receive little instruction on the practice of PT. For example, students in Arizona State University's program are exposed to PT twice during their course of study. PT is introduced in connection with QC basics (internal QC processes versus external QC). More detail about PT is provided in a unit on CLIA regulations that describes the use of PT to evaluate staff and procedures and explains the rules under which PT operates, including the roles of PT providers and sample handling. A report of an interesting application of PT for analysts' education describes how analytical chemistry students in the Netherlands participated in PT for the purpose of proving and improving their skills.[28] The results were submitted and judged in the customary way with the data, mean value of all participants, and z-scores presented just as they would be in routine PT. The final presentation of the results took place in an annual symposium where awards were presented to the best performing student groups, and a lecture program was presented with subjects in the area of QA and QC. A similar system could be considered for medical technologist training in American universities.

Users of laboratory data, such as patients and clinicians, can also benefit from education regarding PT. Patients rarely have the opportunity to choose the laboratory that tests their samples since laboratories are typically chosen by the clinician or dictated by third-party payers. The choice of laboratory, however, may eventually become an issue that a patient will discuss with his or her clinician as increasingly savvy patients become more discerning about all aspects of their medical care. In addition, because of increased mobility of patients among providers and the increasing prevalence of patients who are being cared for by multiple medical specialists (e.g., a diabetic patient who receives care from an internist, an endocrinologist, a cardiologist, and an orthopedic surgeon), comparability of clinical testing results is becoming increasingly important.

Task 3 Recommendations

- PT providers should consider providing more ungraded samples whose sole purpose is educational.
- PT providers should develop an educational program that teaches personnel working in waived and non-waived testing laboratories to evaluate PT results and to increase PT benefits.

- PT providers should provide training materials that follow guidelines such as those contained in CLSI GP27[27] to help PT users interpret and use their PT results for quality improvement.
- PT providers should consider providing immediate feedback to laboratories before releasing official results when results indicate that PT failure is likely. PT providers should also institute a system that warns a laboratory when the trend of its cumulative results indicates that it is at risk for failing a subsequent PT round.

TASK 4: TO DETERMINE HOW WELL THE PT PROGRAMS ARE PROVIDING CHALLENGES TO KEEP UP WITH ADVANCES IN LABORATORY TESTING TECHNOLOGY

The Working Group considered areas of laboratory practice for which PT currently is not available and alternative approaches to ensure that PT becomes available as soon as possible after a new area of testing or a new technology is introduced. These alternatives include use of PT programs from other countries and formation of specialized PT providers for esoteric testing. Also considered were a potential role of for-profit PT providers and the special requirements of PT for molecular testing.

For what areas of clinical laboratory testing is PT currently not available?

PT providers in the United States cover analytes that make up most of the total annual volume of laboratory tests. The tests for which no PT is available may include those with potentially great clinical significance. One large reference laboratory reported that PT was not available for more than 300 of the tests performed by its personnel. At another large laboratory, PT was not available for more than 200 tests performed. At a third organization, it was not available for 80 tests.

PT is not available for many newly developed tests, new applications for existing tests, and esoteric tests. Gaps in PT availability also exist for tests in which analyte stability is a frequent issue and for tests of analytes in specific matrices, such as hair, saliva, or sweat. Clearly, there are growing unmet PT needs that may affect some laboratories more than others, depending on their test repertoire.

What alternatives exist to assure that PT is available as rapidly as possible in new testing areas or for new technologies?

When PT is not available for tests being used in laboratory medicine, several alternatives have been offered. The CLSI offers guidance for assessing laboratory tests, using a variety of alternative assessment methods, when PT is not available.[29] These alternatives include splitting samples with another laboratory that also performs the test, performing an internal blinded split sample procedure in which the second (or n^{th}) replicate can be tested in the same run or at some future time, and audit sample procedure or analysis of calibration material traceable to a reference material or procedure. Other options include analysis of inter-laboratory QC data submitted from multiple laboratories, analysis of patient data to evaluate test stability and reproducibility, use of PT from another country, and use of government or university inter-laboratory comparison programs. Some of these alternatives are viable for only a limited number of tests.

PT from other countries

For purposes of external QA, alternatives to PT have limitations. When the results of split sample comparisons are widely discrepant, results can be very difficult to interpret – which laboratory is "right?" Laboratories may consider PT from providers in other countries when domestic providers do not offer PT for a particular analyte. CDC has compiled a listing of foreign PT programs (http://wwwn.cdc.gov/mlp/eqa.aspx). The list of individual analytes offered by each program may not always be up to date since such a list is very difficult to maintain. By accessing the individual programs through active links at the website, an interested laboratory can obtain the latest information directly from each program.

Many PT programs offered by providers outside the United States have been in existence for many years, are of high quality, enjoy a wide base of participating laboratories from around the world, and keep pace with advancing technologies. Several prominent examples of well established foreign programs that fill gaps in existing domestic PT programs are listed (Table 3-1). Many other programs, ranging from the analysis of metabolites of industrial chemicals in urine (Finland) to esoteric therapeutic drugs (UK), are available from international providers.

Table 3-1. Examples of PT programs from international providers that fill gaps in offerings of approved U.S. providers

Analyte or analyte class	Provider/Country of origin	Description
Trace metals in blood and urine	Centre de Toxicologie du Quebec/ Canada	Offers surveys for 8 trace metals in blood and serum & 10 metals in urine. Also offers PT for organochlorine pesticides, polychlorinated biphenyls, and polybrominated diphenyl ethers in serum.
Metabolites and molecular markers of inherited metabolic disorders	European Research Network of for Evaluation and Improvement of Screening, Diagnosis and Treatment of Inherited Disorders of Metabolism (ERNIDM)/European Union	Network sponsors several PT programs originating from academic institutions, e.g., Clinical University of Heidelberg in Germany offers a qualitative urinary organic acid analysis QA scheme with 77 participants, including 10 US laboratories.
29 amino acids in serum (quantitative analysis)	University of Basel/Switzerland	182 participating laboratories in 26 countries
Nucleic acid amplification in diagnostic virology	QC for Molecular Diagnostics (QCMD)/Glasgow, Scotland, UK	Specializes in the standardization and QC for molecular diagnostics and genomic technologies; 100 laboratories and 26 different virus detection surveys.

Formation of independent PT providers for esoteric testing

The continuing gap between new technology development and availability of PT materials and programs needs periodic review and attention. This review could be

achieved by establishing a board of professionals working in laboratory medicine with the mission of seeking to keep PT programs in the United States abreast of advances in laboratory testing technology. This is the model that foreign PT programs and a few American programs have followed (e.g., the American College of Medical Genetics (ACMG) CAP collaboration for genetic testing). Most of these providers are academic institutions that have seen the need for EQA programs to improve data that support medical research. As these programs and testing have matured, the testing has become more widespread, leading to increasing participation. Another alternative, not incompatible with such a board, could follow the model of the orphan drug program in which the government would subsidize PT for clinically important tests that have few participants.

When a "sufficient number" of laboratories perform a test, most PT providers will make PT programs and materials available. "Sufficiency" is an elastic concept that depends on many factors including PT providers' calculations of development costs, the nature and extent of idiosyncratic problems in PT sample preparation or handling (sample stability, matrix effects), and the providers' calculation of the eventual utilization of the test by laboratories and clinicians. Developing and evolving testing methods can be a challenge to PT providers. Usually a PT "champion" is needed who will assume the risks of being the first to offer a PT program for a new test. Other providers typically follow as the test is adopted by increasing numbers of laboratories. The supply of PT materials for new technologies often lags behind demand because of technical difficulties and development costs for suppliers. Yet several PT "champions" have worked closely with suppliers of PT materials and participating laboratories to launch very successful programs in new areas. An example is the blood and urine trace metals program in Quebec, Canada (Table 3-1).

For-profit PT providers

CLIA specifies that PT providers for regulated analytes must be non-profit entities. For-profit companies are not allowed to be PT providers for testing of regulated analytes in the United States. A change in this restriction would require a change in the CLIA statute: Title 42, Chapter 6A, Subchapter II, Part F, subpart 2, §263f (3) (C). The exclusion of for-profit organizations from participation as PT providers was originally derived from the perceived conflict of interest that arose when manufacturers of instruments were also providing PT samples to their clients. Some have argued that if for-profit organizations avoided conflicts of interest, they should be allowed to compete with the non-profit PT providers. Conflicts of interests would include activities such as selling instruments, testing kits, QC materials, clinical laboratory methods, or other vested interests that might influence judgment of PT results. For-profit providers may be well-suited to offer PT services for non-regulated analytes in specialized fields such as genomic testing or testing in new areas of laboratory technology. If resource issues are a barrier to development of PT for these new technology areas by non-profit providers, then perhaps the for-profit PT providers could make PT materials available, bearing in mind that if the analyte(s) covered by the program were to become regulated by CMS, they would not be a recognized PT provider.

Molecular Diagnostic Testing - Area for Alternative Approaches to PT?

In recent years, molecular or nucleic acid-based methods have become widely used in clinical laboratories as a result of the advances in human genomic research, the sequencing of an increasing number of organisms, and *in vitro* diagnostic (IVD) technologies. Common molecular genetic technologies, such as nucleic acid amplification, polymerase chain reaction (PCR), nucleic acid sequencing, and probe hybridization have been used in diverse areas of laboratory testing, including not only genetic testing but also microbiology, hematology, clinical chemistry, and other traditional disciplines in laboratory medicine**.**

The accuracy and reliability of molecular diagnostic tests can be influenced by many factors, including the diversity of testing methodology, the rate of technology evolution, the variety of applications, regional differences in the tests offered and the populations tested, low-volume testing for many conditions and genetic targets, the lack of standardization inherent in in-house methods developed by individual laboratories, and other factors. As molecular diagnostic tests have emerged, mechanisms for PT of these tests have become both necessary and challenging.[30] PT programs are currently available for only a small number of analytes and diseases/conditions. The CAP/American College of Medical Genetics Molecular Genetic Survey includes challenges for 21 relatively common genetic diseases in 2008; in contrast, based on information from GeneTests, molecular genetic testing is presently performed for at least 1,000 diseases.[31] For many molecular diagnostic tests, especially newly developed tests, tests evaluating less common or rare conditions or genetic targets, tests performed by only one or a few laboratories, and tests associated with a lack of suitable PT materials, analyte-specific PT programs are not available or considered not feasible or practical. The inability of existing PT or external quality assessment (EQA) programs to keep pace with the growth of molecular diagnostic tests in the field has been recognized as a major concern.

In light of the need for quality assessment and quality improvement of molecular genetic testing, DLS initiated studies in 1999 through 2000, to identify issues critical for establishing quality assurance for performance improvement in molecular genetic testing. One of the five core recommendations was to develop methodology-based performance evaluation approaches to supplement existing PT programs, particularly for diseases and/or methodologies not covered by the existing programs.[32, 33]

Currently, PT is available for many of the most commonly performed molecular genetic tests, such as testing for cystic fibrosis mutations, factor V Leiden, and fragile X CGG repeat expansion, target at specific gene(s), sequence(s), mutations, or genetic variations. Molecular infectious disease tests are often performed to detect and/or quantify specific genes or genomes of pathogenic organisms, while molecular diagnostic methods used in other testing areas, such as molecular oncology testing, mostly evaluate specific gene amplifications, translocations, and other alterations associated with disease. In the meantime, nucleic acid sequencing, which is capable of exact determination of every base within a gene or gene fragment of interest, is evolving rapidly as an important method in molecular diagnostics. Sequencing is well-accepted as the "gold standard" for analytic validation of new DNA-based mutation testing and for unambiguous genotype

determination. Nucleic acid sequencing is routinely used for evaluating genetic susceptibility to cancer and other conditions that are associated with complex gene variations. This technology is especially useful for identification of mutations or clinically significant variations that occur throughout the gene or gene segment with no predominant mutations or variations and for private mutations that are seen in only one or a few families. In addition, sequencing can be used to precisely measure trinucleotide repeat expansions; for viral genotyping and determination of sequences associated with drug resistance; for confirming preliminary findings from scanning methods; for high resolution HLA typing in tissue typing for transplantation; and in pharmacogenetic testing. In addition, PCR technologies are essential procedures for most molecular diagnostic tests. These common methods have been suggested as specific areas for considering the development of methodology-based, or generic performance evaluation programs.

In response to the needs for PT in molecular diagnostic testing, efforts have been made both in the U.S. and internationally to explore alternative approaches to PT. Laboratories and professional organizations have formed networks for PT and other purposes. The ACMG and CAP address generic technological issues in molecular diagnostics using a methodology-based approach to PT, since developing a menu of PT programs for each genetic test would be unfeasible, at least in the short term. This generic, method-based approach examined 5 common procedures in molecular genetic testing: 1) DNA extraction, 2) PCR and amplicon purification, 3) cycle sequencing, 4) electrophoresis, and 5) interpretation and reporting. Participants were sent a PCR primer set along with instructions for amplification conditions. This method-based approach (rather than analyte) allows examination of analytical performance in specific technical steps, for example, the effectiveness of DNA extraction separately from PCR amplification. This generic methods-based approach was discontinued, but needs to be reconsidered.
In Europe, generic, methodology-based PT (EQA) programs have been considered valuable tools for evaluating laboratory performance for molecular diagnostic tests. For example, an EQA survey in Italy was reported by Raggi,[34] in which 39 participants were supplied primers and reference materials to evaluate their competence in three specific areas: DNA extraction, PCR performance, and interpretation of results after electrophoresis. The EQUAL project (Multi-National External Quality Assay Programmes in Clinical Molecular Diagnostics), which is supported by the European Union and the European Communities Confederation of Clinical Chemistry (EC4), consists of a series of 3 different EQA programs for assessment of molecular methods independent of analytes. One of the programs, EQUAL-qual, a continuation of the Italian program, attempts to provide critical assessment of laboratory performance in DNA extraction, PCR amplification, and data interpretation after electrophoresis steps separately. EQUAL-quant assesses quantitative PCR assays, and EQUAL-seq evaluates sequencing-based assays in which only the DNA sequencing step is addressed.[35, 36] The participants identified as having performance issues were invited to participate in an EQUAL training course offered in three European cities. The results of a repeat test performed by the training course attendees showed a significant improvement in performance. Another EQA program, the European Molecular Genetics Quality Network (EMQN), has also developed a scheme for the evaluation of laboratory performance in

nucleic acid sequencing, through assessing the ability of laboratories to detect variants, analyze raw DNA sequence data, and use standard statistical measures.[37] Both the EQUALseq and the EMQN sequencing scheme demonstrated that there is room for improvement and PT is essential for assessing the quality of the molecular genetic tests being performed and the test reports being issued. It was suggested that these pilot programs, by evaluating the methodologic aspects of sequencing methods, are crucial to uncover strengths and weaknesses in techniques and postanalytical proficiency; and the experiences obtained should serve as an approximation to such future EQAs.[37-39]

In this growing field of technology, regulatory staffs and professional laboratorians face difficulty keeping pace with rapid advances by using their EQA programs. Both the expansionist approach of proliferating individual disease challenges and the reductionist approach of generic methods-based challenges have advantages and disadvantages and place unique stress on participating laboratories.[38] Although demand for adding gene-specific schemes to PT programs is strong, the number of gene targets relevant to clinical laboratory testing places a practical limit on the coverage of PT in this format.[37] Eventually, American PT providers will need to reconsider developing generic technical schemes to meet the QA demands of clinical laboratory medicine that is being transformed by molecular genetic technologies.

Task 4 Recommendations

- An independent advisory board should be established with the purpose of identifying new and evolving technologies and analytes in laboratory medicine, developing innovative approaches in PT programs, and alerting PT providers of new opportunities for PT offerings.
- A PT program should be developed for genomic testing based on the process of testing so that it can be used generically for many molecular genetic tests rather than developing a unique test for each member of the vast array of genetic tests.
- A listing of national and international PT programs should be maintained by CDC and should continue to be posted on the CDC website.

TASK 5: TO DETERMINE WHETHER ACCREDITATION OF PT PROGRAMS TO AN INTERNATIONAL STANDARD WOULD INCREASE THE QUALITY OR UNIFORMITY OF PROGRAMS

In considering adoption of international standards for PT programs in the United States, the changes that would occur in the operation and cost of PT should be assessed as well as the potential benefits that could accrue from their adoption.

The International Standards Relating to PT

The ISO/IEC Guide 43:1997, parts 1 and 2 is the main international consensus document referenced for PT and presents guidance on PT for a broad variety of testing and calibration laboratories. It is not, however, a requirements standard. There are other international consensus standards for PT in specific fields of testing or from professional organizations. Guide 43 is the most widely referenced document and is used as the basis of all international agreements.

The most widely used requirements standard is ILAC document G13 – *Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes*. First published in 2000, G13 was revised in 2007 and contains the guidance language from Guide 43-1, rewritten as requirements. It also includes the Management Requirements from ISO/IEC 17025:2005 - *General requirements for the competence of testing and calibration laboratories*, slightly edited for PT providers rather than laboratories. The ISO 17025 management requirements are essentially the same as for ISO/IEC 9001:2000 *Quality management systems – Requirements*; therefore, organizations accredited to ILAC G13 are considered to be in conformity with ISO/IEC 9001 and ISO/IEC Guide 43-1. The 2007 revision includes the experiences of multiple accreditation bodies and PT EQA providers.

G13 is used by most, if not all ILAC members for recognition of the competence of PT providers. Some ILAC members offer accreditation of PT, but all members approve the use of accredited providers (if the program has adequate content). G13 is used for the approval of medical PT or EQA providers in Australia, Canada, China, Hong Kong, Mexico, South Africa, Thailand, and countries in Europe. These countries have different requirements for program content, frequency, and evaluation criteria, but all operate in conformance with the standard. G13 is also used for recognition of PT providers in the United States and internationally in the fields of food, environmental science, calibration, electro-mechanics, mechanical properties, and many others. In the United States, accreditation to G13 is required for all environmental PT providers.

ISO/IEC Guide 43-1 and 43-2 are currently being revised as a requirements document and will be titled ISO/IEC 17043. Work began in 2006, and the document is expected to be published in 2009 or 2010.

A related relevant ISO document is ISO 13528:2005 *Statistical methods for use in proficiency testing by inter-laboratory comparisons*. This is a standards document that describes what the ISO Technical Committee for Statistical Applications (TC69) considers best practices for the analysis of PT data. The robust method for estimating the mean and SD and the use of z-scores* are significantly different than current CLIA practices. ISO 13528 allows use of alternative statistical methods if they are statistically valid and are fully explained to the participants. This standard could raise questions about the 3SD outlier criterion currently used for some analytes in the CLIA PT scheme (i.e., values that are three standard deviation units or more from the population mean are excluded when calculating the sample mean for purposes of scoring PT performance in a peer group).

ISO 13528 methods are being used successfully in a variety of programs including medical PT. They are preferred for smaller group sizes and for identifying samples that could be contaminated or otherwise inappropriate for grading.

Differences between G13 and CLIA

In considering ISO and ILAC standards, it is important to consider that the standards are very general and are meant to apply to a broad range of programs. The standards apply to one-time PT programs, round-robin circulation of measurement artifacts, and one-off measurement audits (individual test samples with known characteristics and a pre-determined acceptance interval), as well as the bulk shipment of identical materials to every participant (as in the CLIA programs). The requirements are intended to apply to all fields of testing and calibration. The very general requirements can sometimes seem obvious, mundane, or not relevant for mature PT programs in an established market, such as exists under CLIA.

From the ISO perspective, CLIA requirements are considered as specifications for the technical content of PT programs, including number of samples, frequency of testing, and evaluation criteria. CLIA includes quality requirements for samples regarding matrix and homogeneity, stability, and operational requirements for handling, storage, and shipping; data processing; and reporting. CLIA requirements and the ISO G13 standard differ in level of detail. For example, CLIA requires homogeneous samples, while ISO G13 has requirements for selection of samples for homogeneity testing, timing of the testing, and criteria for acceptance. CLIA requires that technical assistance and detailed reporting of results by participants be provided, while ISO G13 gives long lists of specific responsibilities for the advisory committee and specific items to be reported to participants. Exceptionally diligent U.S. providers may be meeting both sets of requirements for reporting and documentation, but most do not.

* The z-score is the value from the Gaussian normal distribution with mean μ and standard deviation σ corresponding to a value x, such that $z = (x - \mu)/\sigma$. In the context of PT, x is a value of a PT test result for one laboratory in a population of laboratories participating in a given PT event.

The programs currently operating for CLIA are quite mature and have evolved many of the technical practices advocated in the standard such as those governing design and planning of PT samples, handling, packaging, shipping, and reporting. Although by no means obvious to newcomers to the field of PT, the value of these practices quickly becomes evident as new providers gain experience. Most U.S. clinical PT providers have substantial experience, and their technical operations, both in-house and subcontracted, are largely consistent with international requirements.

ISO and CLIA differ primarily in their management system requirements. These are specific requirements for management practices, document control, staff qualification and training, subcontractor monitoring, internal audits, corrective actions, and continual process improvement. The requirements are similar to those of ISO 9000 and are broadly applicable to assuring consistent levels of quality in service and manufacturing organizations. Since CLIA is silent on management system requirements, and no U.S. PT provider is certified to ISO 9000 or accredited to G13, the extent to which U.S. providers meet ISO management system requirements has not been documented.

In summary, if international accreditation standards were adopted in the United States, currently approved providers of PT for clinical laboratory testing would find themselves at different levels of conformity with the requirements. This level of conformity would depend on how the providers qualify and monitor their vendors and sub-contractors, how they verify the quality of their PT samples, how they train their staff, and whether they follow conventional quality management practices. It would also depend on the extent to which they have documented their management policies and work procedures and their system for keeping records. It is possible that some providers are operating in essential conformity with G13 and would have to make few changes in operations to meet the standard.

It is likely, however, that most providers would have to make at least some operational changes if G13 were required. In decreasing order of extent of current application, those aspects of G13 that would require the most effort to implement are:

1. Documenting the management system and technical procedures
2. Adopting ISO-type quality management practices
3. Verifying the homogeneity and stability of samples

Recognition of PT providers

Presently, U.S. PT program providers must comply with CLIA requirements under 42 C.F.R. §493. PT providers must renew their CLIA applications each year to continue as approved PT providers under CLIA. Specific requirements are delineated in the documents that each provider receives from CMS for the subsequent program year. The application includes information on the planned programs, staff expertise, and technical assistance for the PT participants. The review documents also include information on performance in the past PT activities and information on ungraded samples. Providers are frequently asked to send additional material or correct inadequacies in their submissions,

but the approval process does not entail a site visit. ISO Guide 43, like ISO 9000, requires confirmation of conformance by an on-site audit.

Recognition of conformance with ISO/ILAC standards could occur in several ways. In order of increasing rigor, the options include:

1. Self-declaration – providers adopt the practices and documentation requirements of ISO Guide 43 on their own recognizance, without an audit or inspection by outside reviewers.
2. Review of documentation – providers submit their conformity to the requirements of ISO Guide 43 as they do in current annual submissions to CMS.
3. Audit by CMS staff – in addition to reviewing documents, CMS would conduct site visit(s) and assessments of each provider to ensure that they meet the requirements of the standard.
4. Accreditation by an ILAC member organization – following ISO 17011 and ILAC practices, this option includes an on-site assessment by trained assessors and following a documented procedure for working through the accreditation process.

Costs of accreditation to international standards would depend on the level of recognition of conformance and the differences between current practices and the standard. Self-declaration is likely the least costly option for PT providers and users. Although more difficult to document than costs, implementation of ISO/ILAC standards is likely to yield benefits from improved PT processes, quality, and customer service. Recognition by an ILAC member body would also open new business opportunities for U.S. providers in other countries where accreditation is required. Competition among PT providers in the United States could also increase substantially if accreditation (and content) were the conditions for approval of providers, rather than the country in which the provider is domiciled or the profit/not-for-profit status of the provider.

Comparing CLIA and ISO/ILAC requirements in the abstract is difficult and of limited utility. In general, CLIA requirements are less general and more prescriptive. ISO Guide 43 requirements are less detailed but broader in scope. In addition, because of differences in practices among providers and CMS interpretive guidelines,[†] many providers have implemented practices that go beyond the letter of CLIA requirements. Consequently, it would be useful for individual PT providers to conduct an internal comparison of their specific current practices to the ISO Guide 43 requirements (a crosswalk) to determine how much effort would be needed to meet the requirements.

[†] CLIA Subpart I requirements for laboratories and surveyors from accrediting organizations do not address PT, per se, but nonetheless influence laboratory practices, QA, and documentation requirements regarding, for example, investigation and resolution of PT results that indicate potential problems with analytic precision.

What are the barriers and facilitators for adopting an international standard?

That PT providers and PT suppliers would have to implement changes to their current operations to meet the requirements of the international standard is one difficulty of adopting an international standard. A PT provider would need to establish a clearly defined management system, have fully documented standard operating procedures and policies, and would be subject to on-site assessments. At present, CLIA relies upon review of documentation in evaluating PT provider compliance, but there is no on-site assessment. Furthermore, PT providers would have to ensure that their subcontractors meet the same international standard. A PT provider may incur a potentially significant cost, depending upon the status of the organization making the application, if they are not already performing these procedures. There would also be monetary costs of implementing ISO standards and gaining accreditation fees for application, annual fees, audits, and expenses for training staff. Nevertheless, many organizations that have adopted ISO quality management practices have documented cost savings as a result of increased efficiency, reduced waste, and improved customer satisfaction.

Even though CLIA regulations do not require PT providers to meet international standards such as ISO Guide 43, there is nothing in the regulations or governing law that prevents PT providers from voluntarily becoming accredited to ISO standards. The identification of new opportunities in international markets might influence some PT providers to expand and to seek accreditation to an international standard. At present, most PT providers and PT users in the United States see little or no significant opportunity for collaboration with the international community, and therefore have little incentive to adopt an international standard.

Workgroup members commented that U.S. requirements for PT providers are considerably less rigorous than are requirements for laboratories. CLIA includes specific requirements that laboratories demonstrate competence, implement quality management practices, and meet specific standards of personnel training. They also point out that non-U.S. laboratories that provide services to the U.S. medical community must meet CLIA requirements, be CLIA-accredited, and utilize a CLIA-approved PT provider. These international laboratories also are required, in most cases, to obtain accreditation to the more comprehensive international standards. Moreover, these critics note the irony that providers of PT services for laboratories that conduct analyses of environmental contaminants in the United States must meet ISO/ILAC requirements, while PT providers for clinical laboratories do not. Although adoption of an international standard may increase costs of PT, many in the clinical laboratory community are likely to see the move as strengthening current PT programs.

One definite benefit of adopting ISO standards is an improved sense of trust in the results of PT by laboratories and accreditation bodies. In the international arena and in U.S. non-medical areas, demand for accreditation comes primarily from the PT providers themselves, who are interested in demonstrating competence.

How can the other stakeholders be assured that the PT providers are operating programs that meet their needs?

The most reliable source of trust is experience, but absent that, users rely on third-party or government oversight for content and quality. Currently, stakeholders rely on the CMS review of provider documents to ensure appropriate and reliable PT. Laboratories can rely on their experience with various providers and the different program content, schedule, price, and customer service. State regulators can implement their own rules for program content and reporting, but they typically accept decisions by CMS and the laboratories. The internationally preferred process involves competent third party assessment against consensus standards for competence, in addition to review for technical content.

Task 5 Recommendations

- U.S. PT providers should be encouraged to assess the use of internationally recognized PT standards, such as ILAC-G13:2007 or ISO Guide 43-1: 1997.
- PT providers and CMS should assess the benefits and costs of adopting a standard that requires PT providers to be audited by a qualified third party.

CHAPTER 4. RECOMMENDATIONS

The recommendations that appear at the end of each subsection in the previous chapter are here categorized and presented as follows:

- Recommendations 1-5 involve *Data Collection and Analysis*
- Recommendations 6-17 involve *Process Improvement*
- Recommendations 18-19 involve *Process Evaluation*
- Recommendations 20-21 involve *Education*

If adopted, these recommendations can help improve the state of PT. Although implementing them presents many challenges to the laboratory medicine community, most are feasible. Most recommendations require resources (i.e., funding and/or personnel) to do the job. In each case the key question to be resolved by CDC, CMS, and the larger laboratory community is whether the potential benefit from implementation of the recommendation warrants the additional costs. Some recommendations require changes in CLIA regulations or the statute. Some recommendations may result in increased costs to PT providers, PT users, and perhaps ultimately to the patient – a challenge to our health care system where medical costs are already deemed too high. The intangible savings in terms of improved quality and confidence in laboratory results, however, make these recommendations worth striving for. Although the costs of laboratory tests constitute a relatively small fraction of the costs of medical care, laboratory testing has a pervasive influence on clinical decision making that affects patient care; these recommendations therefore warrant careful consideration by regulators, PT providers, and the clinical laboratory community at large.

Data Collection and Analysis

22. Conduct a study of the existing information in the scientific literature and current databases regarding reasons for unsatisfactory PT results in order to identify areas most in need of improvement or additional research/analysis.
    a. Desired result – Direction and priorities to achieve meaningful improvements in the quality and reliability of laboratory medicine results.
    b. Who should do this? – CDC should lead this effort in partnership with stakeholders.
    c. Feasibility – The principal obstacles include garnering cooperation of stakeholders who control data sources and costs of the study.
23. Develop and make available a database to collect PT data for characterizing the performance of all laboratories, for identifying reasons for unsatisfactory PT results, for reviewing acceptance criteria used by PT providers, and for identifying a list of analytes that should be regulated.
    a. Desired result – An ongoing, up-to-date database that can be used to assess the current state of PT and how it might be improved.
    b. Who should do this? CDC should establish and maintain this database, analyze these data, and provide input to CLIAC and other stakeholders for recommended improvements.
    c. Feasibility – Principal obstacles include concern among stakeholders that data can be misused (e.g., by the news media) if access is not controlled.

24. Develop a process to collect, consolidate, analyze, and summarize all complaints received by CMS, state health programs, accrediting organizations, and PT providers about PT. This process includes developing appropriate statistical analyses of data to identify correctable trends and the publication and dissemination of the complaint process for widespread use by all parties.
    a. Desired result – Provide information for the development of strategies to identify and address systemic problems and issues regarding PT so that improvements can be accomplished and confidence in laboratory medicine can be maintained.
    b. Who should do this? –CDC should establish relationships with involved stakeholders, develop this process, and perform on-going analyses of data.
    c. Feasibility – CLIA regulations allow collection of these data. Principal obstacles include garnering cooperation from stakeholders who control data sources (who may see this information as proprietary) and the associated costs.
25. PT providers should publish scientifically credible PT results on a regular basis in peer-reviewed journals.
    a. Desired result – To make information available to the PT user community that could help in future directions of laboratory medicine.
    b. Who should do this? – PT providers would have to take the initiative to accomplish this.
    c. Feasibility – Providers have PT results, which can be tabulated and presented in a format suitable for submission to a suitable scientific journal. Principal obstacles include persuading PT providers that this information is needed and valuable to the laboratory medicine community.
26. CDC should continue to maintain and update the listing of national and international PT programs on their website.
    a. Desired result – To provide a resource for the PT user community that lists current sources of PT programs and materials.
    b. Who should do this? – CDC has already established this webpage, but it needs to be updated and to include all possible PT providers.
    c. Feasibility – CDC should assign responsibility and resources to ensure that the listing is updated regularly. Principal obstacles include obtaining the resources to update and maintain this listing.

Process Improvement

27. Develop a process to assure that all clinical laboratories, including those that perform waived tests, participate in PT. This recommendation requires a change in the CLIA statute (law) (Public Health Service Act: Section 353 [263a] [d] [2] [C]) that specifically exempts waived laboratories from standards (i.e., QC programs, PT, and inspections). *(Statute change)*
    a. Desired result – To ensure all laboratories produce reliable results that the patient and clinician can trust.
    b. Who should do this? – CMS would likely be responsible for this, but this process would have to be initiated by HHS or any interested party interacting with Congressional staff.
    c. Feasibility –A change in the statute would have to be initiated by the appropriate Congressional committees. Principal obstacle is the lack of political support for a statutory change. In addition, POLs performing only waived testing may oppose this process because of additional costs and regulatory burden. The large number of POLs could present substantial compliance problems unless CMS is given additional resources.

28. Develop a process to periodically review, update, and publish the requirements of the CLIA PT program, including the list of regulated analytes and allowable limits. (*Regulatory change)*
    a. Desired result – To maintain a current list of regulated analytes that reflects current technology and practice and keeps abreast of new developments in laboratory medicine.
    b. Who should do this? – CMS, possibly in partnership with CDC and CLIAC.
    c. Feasibility – This is a widely-supported change sought by a broad base of stakeholders. Principal obstacles include the onerous process of making regulatory changes and the grading criteria and selection of analytes could be controversial.
29. PT providers should seek ways to provide for faster turnaround time for PT results, including developing a system(s) for electronic submission.
    a. Desired result – To reduce the turnaround time for reporting PT results so that PT may be used in evaluating and solving problems.
    b. Who should do this? – All PT providers and users.
    c. Feasibility – Presently some but not all PT providers have electronic reporting systems. Principal obstacles include persuading some PT providers and users that this system is needed and will be valuable to clinical laboratory testing.
30. Before releasing official results, PT providers should consider providing immediate feedback to laboratories when results indicate that PT failure is likely. PT providers should also institute a system that gives warning to laboratories that trends of cumulative results are moving toward PT failure.
    a. Desired result – To institute a system that rapidly assists PT users in recognizing trends that lead to PT failure and to enable laboratories to make more timely corrections to reduce the effect of errors on patient results.
    b. Who should do this? – PT providers.
    c. Feasibility – Principal obstacles include slow turnaround times; this obstacle could be eliminated by electronic reporting of results requiring shorter time for a laboratory to report PT results upon receiving the PT sample.
31. PT providers should allow for the reporting of analyte results in various units of measure, be able to convert those measures to common units, and evaluate them in accordance with current regulations.
    a. Desired result – To reduce calculation errors introduced when laboratories have to convert units to something other than what they use for patient testing
    b. Who should do this? – PT providers would have to take the initiative to accomplish this. CDC could serve as a consultant in this endeavor.
    c. Feasibility – PT providers could accomplish this relatively easily. Principal obstacles include persuading PT providers that this process is needed and valuable to participants. This process may very well be automated for Internet-based PT results reporting systems.
32. PT providers should summarize PT results graphically for end users in a manner that is easy to read and understand.
    a. Desired result – To promote a better understanding of PT results
    b. Who should do this? – PT providers; CDC could serve as a consultant in this endeavor.
    c. Feasibility – Some providers already provide graphic reports. There are no serious obstacles to this recommendation other than the time, cost and effort required to develop new report formats.
33. PT Providers should provide samples that mimic patient samples as much as possible with a minimum of artificial matrix effects.
    a. Desired result – To reduce or eliminate complications due to undesired errors from artificial matrices.
    b. Who should do this? – PT providers.

c. Feasibility – Principal obstacles include the nature of some analytes, which may require PT sample matrices that are different from patient samples; source material may not be available; and cost may be an additional consideration.

34. Small adjunct studies using fresh frozen samples from a single patient should be conducted in conjunction with routine PT to identify and characterize unrecognized testing problems.
    a. Desired result – To assist PT users in solving problems with analytical processes.
    b. Who should do this? – PT providers.
    c. Feasibility – Such studies are done occasionally and costs are known. Principal obstacles include the cost of preparing fresh frozen samples and the possibly of limited market for these samples.
35. An independent advisory board should be established for the purpose of identifying new and evolving technologies and analytes in laboratory medicine, to develop innovative approaches in PT programs, and to alert PT providers of new opportunities for PT offerings.
    a. Desired result – To keep PT abreast with the changing technology in laboratory medicine.
    b. Who should do this? – CDC in partnership with stakeholders and CLIAC.
    c. Feasibility – CDC could identify experts to serve on a board that makes recommendations for PT providers. Principal obstacles include funding an expert advisory board and ensuring stakeholders are represented.
36. Rather than developing a unique test for each of the thousands of clinically relevant molecular genetic tests, develop a process-based approach for PT that can be used to assess proficiency in processes common to many molecular genetic tests (e.g., nucleic acid sequencing, PCR amplification and purification, electrophoresis and interpretation).
    a. Desired result – To have a process for conducting external quality assessment for numerous genomic tests.
    b. Who should do this? – PT program providers, perhaps with the advice of an independent advisory board.
    c. Feasibility – CDC could identify experts to serve on a board that would make recommendations for PT providers. PT providers would have to implement these recommendations. Principal obstacles include the technical difficulty and cost of developing widely applicable PT processes and materials.
37. U.S. PT providers should be encouraged to assess the use of internationally recognized PT standards, such as ILAC-G13:2006 or ISO Guide 43-1: 1997, in evaluating their quality management systems.
    a. Desired result – To bring uniformity to PT providers and assurance that PT providers are meeting minimal consensus standards.
    b. Who should do this? – PT providers or CMS.
    c. Feasibility – PT providers would voluntarily meet these international standards or CMS would need to amend CLIA regulations to require PT providers to meet international standards. Principal obstacles include costs to PT providers including changes to adapt to the international standard, application, and assessment costs.
38. Benefits and costs of adopting a standard that requires PT providers to be audited by a qualified third party should be assessed. *(Regulatory change)*
    a. Desired result – To ensure that PT providers meet quality standards.
    b. Who should do this? – Potential accrediting bodies that comply with ISO 17011. CMS?
    c. Feasibility: If an international standard were adopted, accrediting bodies could be created that comply with ISO 17011. Principal obstacles include costs to PT providers for assessment expense; these costs are likely to be passed through to laboratories, and ultimately to payers.

Process Evaluation

39. Alternatives to the current CLIA PT testing and sample frequency should be evaluated. *(Regulatory change)*
    a. Desired result – To ensure that PT testing provides adequate external assessment of the quality of laboratory testing.
    b. Who should do this? – CDC in cooperation with stakeholders and CLIAC.
    c. Feasibility: CDC, in collaboration with PT providers, could identify experts to serve on a board to make recommendations, which would ultimately have to be implemented by CMS. Principal obstacles include costs to establish a board to consider and propose potential changes.
40. Alternatives to the current CLIA PT scoring schemes should be evaluated. *(Regulatory change)*
    a. Desired result – To ensure that CLIA PT scoring is adequate to assess the quality of laboratories.
    b. Who should do this? – CDC in collaboration with PT providers, stakeholders and CLIAC.
    c. Feasibility: Principal obstacles include obtaining necessary data and costs to establish an expert panel to consider proposed changes.

Education

41. An educational program should be developed that teaches laboratory personnel how to evaluate PT results to increase the benefits of PT participation.
    a. Desired result – To ensure that all laboratorians understand PT and interpretation of PT results.
    b. Who should do this? – CDC, working with PT providers and with input from stakeholders.
    c. Feasibility: At least three PT providers have such a program in place. Principal obstacles include costs; educating laboratory managers about the need for continuing education and training in PT.
42. PT providers should provide training materials on interpretation and use of their PT results for quality improvement using an approach such as the one described in CLSI GP27[26].
    a. Desired result – To ensure that PT users clearly understand the results from PT providers.
    b. Who should do this? – PT providers, possibly in partnership with deemed accreditation agencies (e.g., CAP, COLA, AAFP) and CLSI.
    c. Feasibility – Principal obstacle include cost and constraints on time available for training of laboratory personnel.

REFERENCES

1. Belk WP, Sunderman, FW. A survey of the accuracy of chemical analyses in clinical laboratories. *Am J Clin Pathol.* 1947;17: 853-861.

2. Department of Health, Education and Welfare, Public Health Service, Center for Disease Control, Clinical Laboratory Improvement Amendments of 1967. Part F, Title III, PHS §353.

3. Centers for Medicare & Medicaid Services Medicare, Medicaid, and Clinical and Laboratory Standards Institute Programs. Laboratory requirements relating to quality systems and certain personnel qualifications. *Federal Register.* 2007;68(16):3639-3714.

4. Elevitch, Franklin D., Noce, Peter S. (Editors), Data ReCAP 1970-1980: A Compilation of Data from the College of American Pathologists' Clinical Laboratory Improvement Programs. Skokie, Ill, College of American Pathologists, 1981

5. Centers for Disease Control and Prevention. Clinical laboratory performance on proficiency testing samples--United States, 1994. *MMWR*. 1996;45(9):193-196.

6. Leeber JC. Role of external quality assurance schemes in assessing and improving quality in medical laboratories. *Clin Chim Acta.* 2001;309(2):173-177.

7. Reilly AA, Salkin IF, McGinnis MR, Gromadzki S, Pasarell L, Kemna M, Higgins N, Salfinger M. Evaluation of mycology laboratory proficiency testing. *J Clin Microbiol.* 1999;37(7): 2297-2305.

8. Tholen DW. Improvements in performance in medical diagnostics tests documented by interlaboratory comparison programs. *Accred Qual Assur.* 2002;7:146-152.

9. Ehrmeyer SS, Laessig RH. Has compliance with CLIA requirements really improved quality in US clinical laboratories? *Clin Chim Acta.* 2004;346(1):37-43.

10. Tholen DW, Lawson NS, Cohen T, Gilmore B. Proficiency test performance and experience with College of American Pathologists' programs. *Arch Pathol Lab Med.* 1995;119(4):307-311.

11. Hoeltge GA, Duckworth JK. Review of proficiency testing performance of laboratories accredited by the College of American Pathologists. *Arch Pathol Lab Med.* 1987;111(11):1011-1014.

12. Ehrmeyer SS, Burmeister BJ, Laessig RH, Hassemer DJ. Laboratory performance in a state proficiency testing program: what can a laboratorian take home? *J Clin Immunoassay.* 1994;17:223-230.

13. Ehrmeyer SS, Laessig RH. Effect of legislation (CLIA'88) on setting quality specifications for US laboratories. *Scand J Clin Lab Invest.* 1999;59(7):563-567.

14. Jenny RW, Jackson-Tarentino KY. Causes of unsatisfactory performance in proficiency testing. *Clin Chem.* 2000;46(1):89-99.

15. Stull TM, Hearn TL, Hancock JS, Handsfield JH, Collins CL. Variation in proficiency testing performance by testing site. *JAMA.* 1998;279(6):463-467.

16. Edson DC, Massey LD. Proficiency Testing Performance in Physician's Office, and Small Hospital Laboratories, 1994-2004. *Labmedicine.* 2007;38(4):237-239.

17. Johnson PR. The contribution of proficiency testing to improving laboratory performance and ensuring quality patient care. *Clin Leadersh Manag Rev.* 2004;18(6):335-341.

18. Kroll MH, Styer PE, Vasquez DA. Calibration verification performance relates to proficiency testing performance. *Arch Pathol Lab Med.* 2004;128(5):544-548.

19. Hoeltge GA, Phillips MG, Styer PE, Mockridge P. Detection and correction of systematic laboratory problems by analysis of clustered proficiency testing failures. *Arch Pathol Lab Med.* 2005;129(2):186-189.

20. ILAC Proficiency Testing Consultative Group. ILAC Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes ILAC-G13. Silverwater, New South Wales, Australia: International Laboratory Accreditation Cooperation; 2006.

21. Schreiber WE, Endres DB, McDowell GA, Palomaki GE, Elin RJ, Klee GG, Wang E. Comparison of fresh frozen serum to proficiency testing material in College of American Pathologists surveys: alpha-fetoprotein, carcinoembryonic antigen, human chorionic gonadotropin, and prostate-specific antigen. *Arch Pathol Lab Med.* 2005;129(3):331-337.

22. Steele BW, Wang E, Palomaki G, Klee GG, Elin RJ, Witte DL. Sources of variability: a College of American Pathologists Therapeutic Drug Monitoring study. *Arch Pathol Lab Med.* 2001;125(2):183-190.

23. Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine chemistry analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. *Arch Pathol Lab Med.* 1998;122(7):587-608.

24. Rej R, Norton-Wenzel CS, Cao Z. Target values and method evaluation in proficiency testing programs. *Clin Chem.* 2001;47(12):2185-2186.

25. Cunningham MT, Praestgaard J, Styer PE, Brandt JT, Fairweather RB, Laposata M, Olson JD, Sosolik RC, Triplett DA. A method for proficiency testing of small peer groups in the College of American Pathologists Coagulation Surveys. *Arch Pathol Lab Med.* 1999;123(3):199-205.

26. Hassemer DJ. Interlaboratory surveys to assess performance: more than proficiency testing. *MLO Med Lab Obs.* 2006;38(1):20-25.

27. Tholen DW, Berte LM, Boone DJ, Cooper WG, Gun-Munro J, Noble MA, Sarewitz SJ, Williams, MW. GP27-A2, Using Proficiency Testing to Improve the Clinical Laboratory; Approved Guideline. 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2007.

28. te Nijenhuis, NB, Heutinck K, Van Schijndel B. The proof of analytical chemical education lies in proficiency testing: a proficiency testing programme in polytechnic academia and vocational colleges in The Netherlands. *Fresenius J Anal Chem.* 2001;369(7-8):741-743.

29. Sarewitz SJ, Ball DJ, George H, Hortin GL, Johnson SR, Mei J, Miller WG, Rabinovitvh A, Smith JM, Stokeld TD, Tholen DW. GP29-A, Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline. Wayne, PA: Clinical and Laboratory Standards Institute; 2002.

30. Richards CS, Grody WW. Alternative approaches to proficiency testing in molecular genetics. *Clin Chem.* 2003;49(5):717-718.

31. GeneTests: Medical Genetics Information Resource (database online). Copyright, University of Washington, Seattle. 1993-2007. Available at http://www.genetests.org. GeneTests . 2007.

32. Williams LO, Cole EC, Lubin IM, Iglesias NI, Jordan RL, Elliott LE. Quality assurance in human molecular genetics testing: status and recommendations. *Arch Pathol Lab Med* 2003;127:1353-1358.

33. Williams LO, Cole EC. General Recommendations for Quality Assurance Programs for Laboratory Molecular Genetic Tests, Final Monograph . Available at: http://www.phppo.cdc.gov/dls/pdf/genetics/dyncor.pdf . 1999. Atlanta, GA, Division of Laboratory Systems, Public Health Practice Program Office, Centers for Disease Control and Prevention.

34. Raggi CC, Pinzani P, Paradiso A, Pazzagli M, Orlando C. External quality assurance program for PCR amplification of genomic DNA: an Italian experience. *Clin Chem.* 2003;49(5):782-791.

35. Ramsden SC, Daly S, Geilenkeuser WJ, Duncan G, Hermitte F, Marubini E, Neumaier M, Orlando C, Palicka V, Paradiso A, Pazzagli M, Pizzamiglio, S, Verderio P. EQUAL-quant: an international external quality assessment scheme for real-time PCR. *Clin Chem.* 2006;52(8):1584-1591.

36. Orlando C, Verderio P, Maatman R, Danneberg J, Ramsden S, Neumaier M, Taruscio D, Falbo V, Jansen R, Casini-Raggi C, Malentacchi F, Marubini E, Pizzamiglio S, Vernelen K, Libeer JC, Palicka V, Pazzagli M. EQUAL-qual: A European Program for External Quality Assessment of Genomic DNA Extraction and PCR Amplification. *Clin Chem.* 2007;53(7):1349-1357.

37. Patton SJ, Wallace AJ, Elles R. Benchmark for evaluating the quality of DNA sequencing: proposal from an international external quality assessment scheme 2. *Clin Chem.* 2006;*52*(4):728-736.

38. Bakker E. Is the DNA Sequence the Gold Standard in Genetic Testing? Quality of Molecular Genetic Tests Assessed. *Clin Chem.* 2000;52(3): 557-558.

39. Ahmad-Nejad, P. et al. Methodologic European External Quality Assurance for DNA Sequencing: The EQUALseq Program. *Clin Chem.* 2006;52(4): 716-727.