Performance in Proficiency Testing: An Indicator of Laboratory Quality?

Robert Rej, Ph.D.
Wadsworth Center for Laboratories and Research
New York State Department of Health
School of Public Health - State University of New York at Albany
Albany, New York

Abstract: Performance in external quality control or proficiency testing schemes is often cited as a measure of the quality of clinical laboratory testing throughout the world. There are significant differences which exist between routine testing of clinical specimens and samples tested for proficiency assessment. The demonstration of equivalent performance in proficiency testing and routine testing is a difficult association to establish. Differences in the mode of testing may include: requestor of laboratory services; characteristics of the specimens; sample transport; specimen identity to laboratory; pre-analytical variables; processing and accessioning; interferences and matrix effects; analytical phase; calculation of results; mechanism or reporting results; reference values; and application of test results. Only in the analytical phase and calculation of results are the processes nearly identical. It would be unexpected that performance in proficiency testing would be identical to routine performance unless these two phases contributed the largest source of error to the process. Nonetheless, split-specimen, or audit sample, testing for cholesterol and theophylline has demonstrated a significant correlation between routine performance and that based upon proficiency testing results.

Introduction

The ultimate objective of proficiency testing is the monitoring and improvement of health care through improving laboratory performance. Laboratory-improvement agencies typically rely on results of proficiency testing, along with on-site inspections, and regulations that specify educational requirements for staff for accreditation. There is no definitive information, however, describing which management attributes are of primary importance when related to performance on laboratory proficiency and which are of secondary importance.

In 1980, Peddecord and Cada[1] examined the effect of several variables on laboratory proficiency and concluded, at least for clinical chemistry and a few other branches of laboratory medicine, that enrollment in an external inspection and accreditation program is related to better performance. A 5-year review by the Centers for Disease Control and Prevention[2] confirmed that an overall program of inspection and accreditation generally improved laboratory performance over time. This review showed that the average number of major deficiencies (those which may have a direct effect on the quality of patient care or could affect the health and safety of hospital or laboratory personnel and must be corrected before accreditation can be extended to the laboratory) decreased from 16 to 6 over the 5-year period.

Improvement in laboratories cannot always be measured in objective or direct terms since factors intermingle and overlap to the point that it would be inappropriate to suggest that laboratory improvement is

solely due to the analytical proficiency testing component of the accreditation process. Several empirical studies, however, have suggested that continued participation in proficiency testing programs is related to improved performance.[3-10]

The Clinical Laboratory Improvement Amendments of 1988 (CLIA'88) have helped the laboratory community in the United States to renew interest in defining the true role of proficiency testing. The CLIA'88 legislation itself calls for assessing of "validity, reliability, and accuracy of proficiency testing."[11] This is a charge to evaluate the effectiveness of proficiency testing. Several questions, while straight-forward at first, are rather complex and difficult to address. Does accuracy of proficiency testing entail that results are exact predictors of those that would be obtained on patient testing or that results are correlated with the quality of patient testing? This review examines the strengths and limitations of proficiency testing as an evaluator of laboratory performance.

Differences in the Process: Patient vs. Proficiency Testing

The clinical laboratory testing process is comprised of several phases and components. One classification scheme may be: the requestor of laboratory services; characteristics of the specimens; sample transport; specimen identity to laboratory; pre-analytical variables; processing and accessioning; interferences and matrix effects; analytical phase; calculation of results; mechanism or reporting results; reference values; and application of test results. This is shown schematically in Table 1, with some potential differences and similarities between the proficiency- and the patient-testing processes. The highlighted area note the two phases where, in my view, considerable similarity and overlap exist. Key areas are reviewed here:

Analytical specificity, interferences, and matrix effects

Are the specimens used in proficiency testing similar to authentic patient samples encountered in routine laboratory testing and therefore a realistic challenge of performance? It is known that so-called matrix effects may give rise to artificially induced errors in proficiency testing. Methods that are exquisitely sensitive to matrix effects, however, are similarly sensitive to alterations in patient sera, and such factors can be assessed.[12-17]

Lyophilization may introduce errors not normally encountered with processing patient specimens.[12] To examine the extent of proficiency test specimen matrix effects, we distributed three specimens differing primarily in their matrix composition in a single proficiency testing event of laboratories enrolled in the New York State survey for clinical chemistry.[15] These were: pooled normally clotted liquid human serum with minimal supplementation, liquid serum prepared by re-calcification of pooled human plasma, and commercially lyophilized serum prepared by re-calcification of pooled human plasma. Wherever possible, analyte concentrations were adjusted to be comparable in all three specimens. Twelve chemistry analytes were selected for comparison including lipid, enzyme, substrate, and ionic constituents. Significant differences in the inter-method behavior (commutability) were found amongst the three types of specimens for HDL-cholesterol with inter-laboratory coefficients of variation (CV) of 11.4% (liquid serum), 28.7% (liquid re-calcified

Specimen Type:	**A Proficiency Sample**	**A Patient Sample**
Requestor:	Ordered by HCFA, proficiency test provider	Ordered by physician or health provider
Sample Characteristics:	Sample obtained from large pool	Client sample obtained from individual
Sample Transport	Transport in mail;	Transport within/among institution(s);
Specimen Identity to Laboratory:	Usually identified; Unique vial or tube	Relatively anonymous - usually one of many
Preanalytical Variables:	Reconstitution errors	Patient preparation; specimen collection; sample collection device; sample pretreatment and centrifugation
Entry into Process:	Enter process at a later stage	Enter process at earliest stage
Accession:	May require special accessioning to avoid creating patient record	Usually routine
Interferences:	Matrix effects due to lyophilization or preparation not seen with patients	Drugs and metabolite effects usually not seen with proficiency specimens
Analysis:	Should be routine; may require special handling due to sample characteristics or analyte level	Usually routine; may require special handling due to analyte level
Calculation of Results:	Should be routine; may require special calculation due to dilution of specimen	Usually may require special calculation due to dilution of specimen
Mechanism or Reporting Results:	Extraordinary reporting (usually manual)	Routine reporting (usually electronic)
Reference Values:	May differ amongst laboratories	Usually uniform
Application of test results:	Result used for laboratory evaluation and/or accreditation	Result used for patient care

Table 1. Differences and Similarities between the Proficiency Testing Process and Routine Clinical Laboratory Analyses.

plasma), and 52.1% (lyophilized). For the other 11 analytes, liquid serum and liquid re-calcified plasma demonstrated similar commutability, while in nearly each case lyophilization introduced considerable matrix effects. With a specimen of liquid origin, a single normal distribution was found for total creatine kinase (mean = 149 U/L, CV = 11.2%, Figure 1 shaded bars), while an apparent bimodal distribution was observed for a lyophilized material using identical analytical methods (mean= 122 U/L, CV = 33.1% Figure 1 open bars). Inter-laboratory coefficients of variation were considerably larger with the lyophilized material for most, but not all analytes, indicating that errors in reconstitution/filling were not the predominant source of variation. CVs (%) were liquid and lyophilized materials were, respectively: glucose 6.9, 6.9; sodium 2.0, 2.0; chloride 3.0, 5.5; cholesterol 3.9, 5.9; creatinine 10.4, 44.9; calcium 3.6, 6.9.

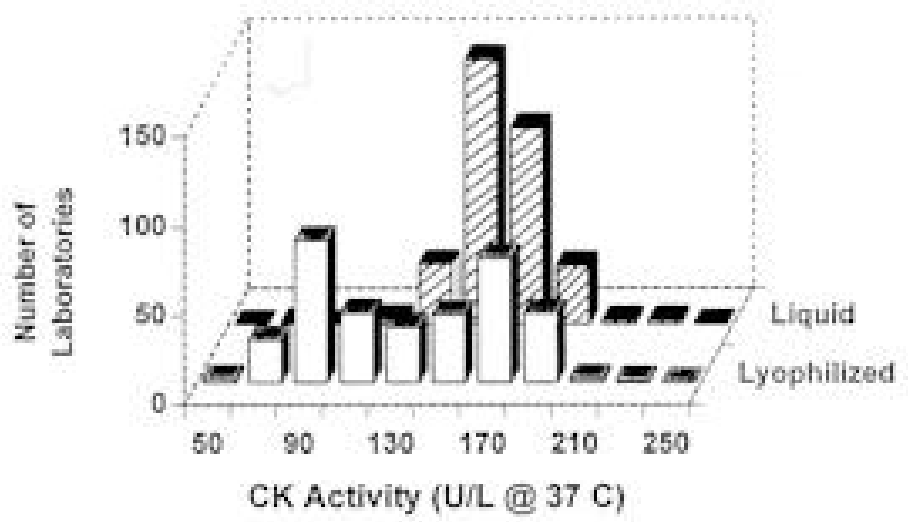

Figure 1. CK Activity by Number of Laboratories, liquid and lyophilized specimens.

There is also the potential for a "reverse matrix effect," whereby interferences in some authentic patient specimens (metabolites and/or drugs) are absent from proficiency specimens, and proficiency testing samples do not provide the range of interferences encountered in routine analysis.[18]

The Analytical Phase

This is the focal point of the clinical laboratory process and one where the proficiency and patient specimen can undergo identical processing. Although this would be ideal, a survey of laboratorians in hematology and clinical chemistry suggests that some special treatment of proficiency test specimens was commonplace a decade ago[19]. It has also been observed that special treatment of external quality control specimens can result in improved performance.[20]

In reviewing proficiency test records and laboratory inspections over the past two decades, it has been my experience that "special treatment" can also lead to poorer performance in a proficiency survey due to the fact that it is out-of-the-routine. The entire range of analyte concentration will inevitably be larger in the proficiency testing specimens for most analytes since these can usually be supplemented at concentrations far outside physiological ranges.

	First Testing	Second Testing
Average Score (LH)	80.1%	91.4%
Number of Laboratories Failing (LH)	19	6
Average Score (FSH)	85.3%	96.6%
Number of Laboratories Failing (FSH)	10	1

Table 2. Performance of Participant Laboratories Upon Test Introduction for LH and FSH in the New York State Endocrinology Proficiency Testing Program

Furthermore, differences amongst the type of laboratories will also affect distribution of analyte concentration (with smaller variations being observed in large reference and university hospitals while larger variations being observed in physician office laboratories). Substandard performance in a proficiency test at extremes of analyte ranges will not provide data that allow projection to the performance likely to be found within usual reference values. Special handling (dilution of samples with elevated concentrations of analyte) or method of presentation to the instrument (e.g., syringe injection or aspiration of blood gas specimens) may be required for proficiency test specimens.

Reference values and application of test results

Although uniform criteria of evaluation are provided by approved CLIA'88 proficiencytesting providers, use of peer group evaluation may perpetuate use of procedures that cause personnel to perform in a clinically unacceptable manner. Although evaluation by peer groups should be performed only for specimens that demonstrate a "matrix effect," criteria for establishing peer groups are vague, and subtle interactions between a method group and a given proficiency testing material may well be treated differently amongst proficiency testing providers. Data on the mechanism used to establish a target value should be available, and if overall participant mean, peer-group mean, or reference method value was used in establishing the target.

Influence of the Proficiency Testing Process Itself

Proficiency testing is usually not a passive barometer that merely monitors laboratory performance. Participating in a proficiency testing program is interactive and genuine poor performance is examined and corrected in most laboratories. The actual performance of laboratories not involved in an external quality assessment scheme is difficult to estimate. This might be gauged, however, by examining the performance of laboratories that are newly enrolled in a

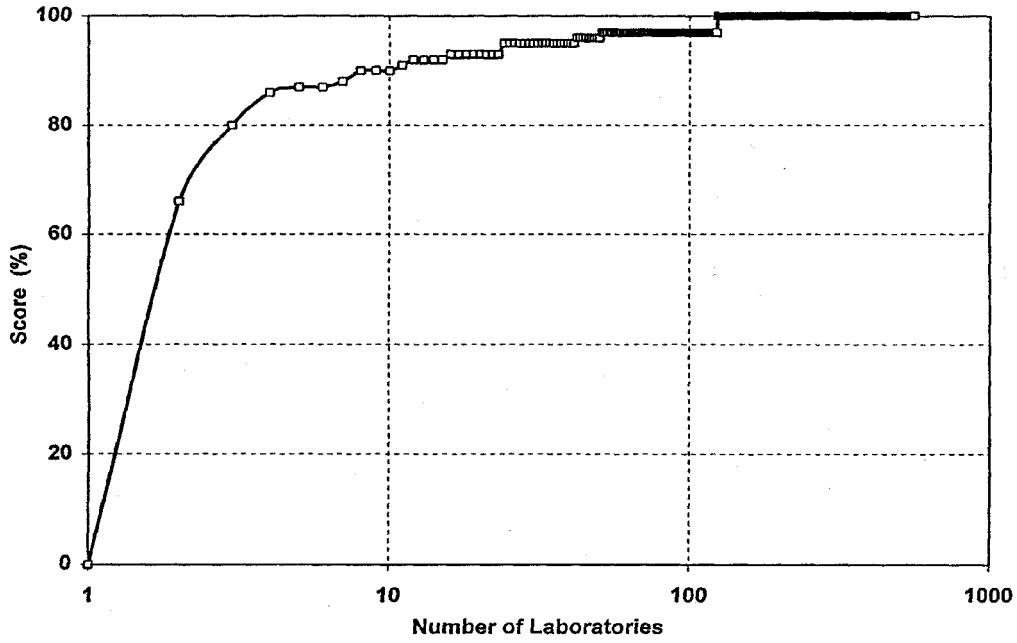

Figure 2. Improvement in proficiency scores by laboratories.

proficiency testing program or with established laboratories when a new analyte is introduced. Table 2 demonstrates that in the 4-month period intervening between the first and second testing in the New York State proficiency testing program, a dramatic improvement in performance can be found immediately after the introduction of proficiency testing for lutropin (LH) and follitropin (FSH). This improvement in performance, both in improved average scores in proficiency tests and reduction in numbers of failing laboratories, was due to two factors: voluntary withdrawal of testing for these analytes by some laboratories and improved performance by those remaining in the program.

Improvement in performance as evidenced by analysis of laboratory proficiency testing results has been demonstrated in the Regional Quality Assurance Programs in the U.S., elsewhere in North America, and throughout the world. Accordingly, examination of laboratories regularly participating in proficiency testing for several proficiency cycles will likely result in examination of reasonably fine distinctions amongst laboratories.[21] This is demonstrated in
Figure 2. Of approximately 600 laboratories participating in the New York State Hematology proficiency test, 10 laboratories failed to achieve scores > 90% (using the CLIA'88 grading schemes), and most (78%) attained a score of 100%. Only two laboratories failed to achieve an overall passing score of 80%.

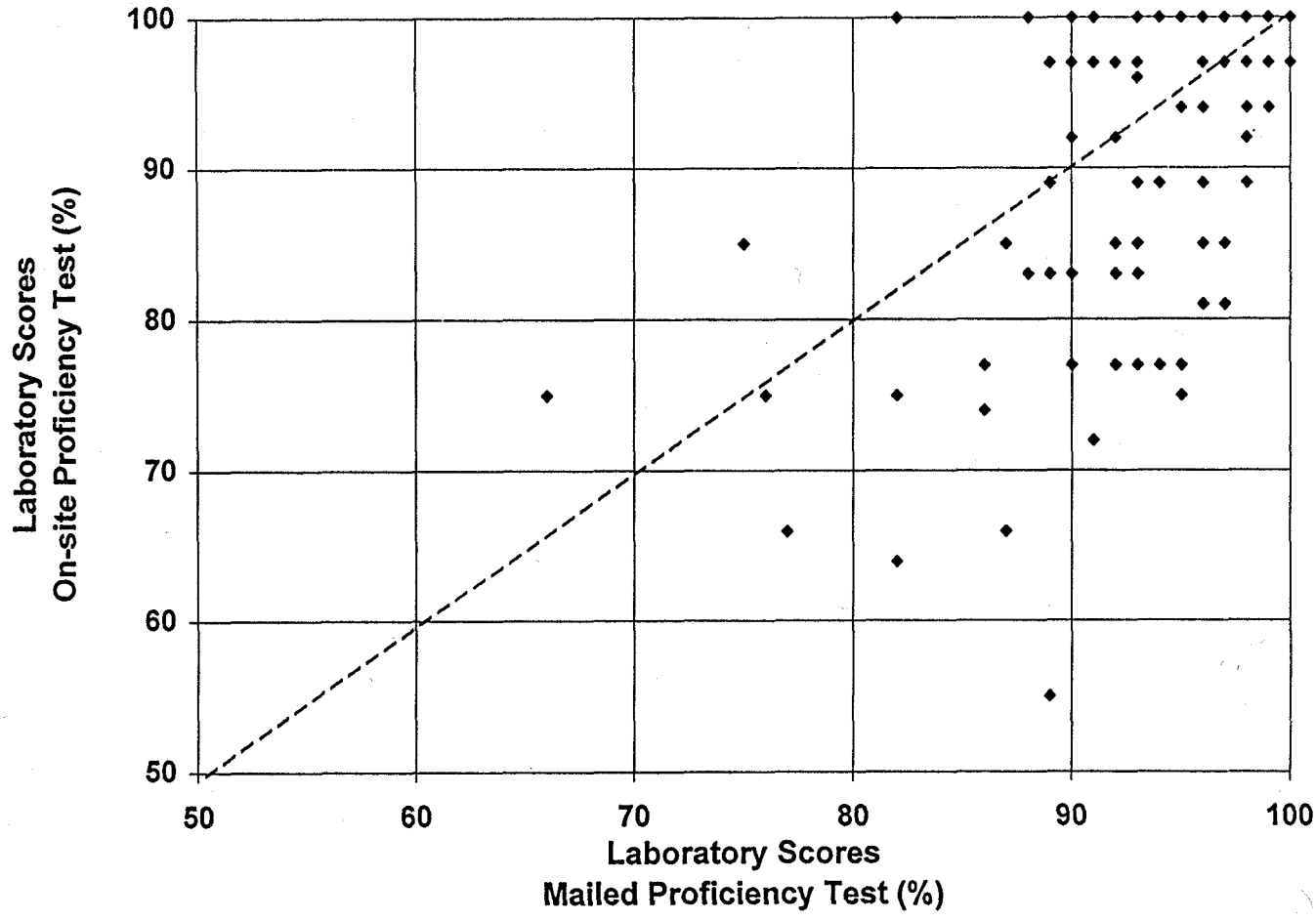

Figure 3. Comparison of laboratory scores, mailed proficiency testing vs. on-site proficiency testing.

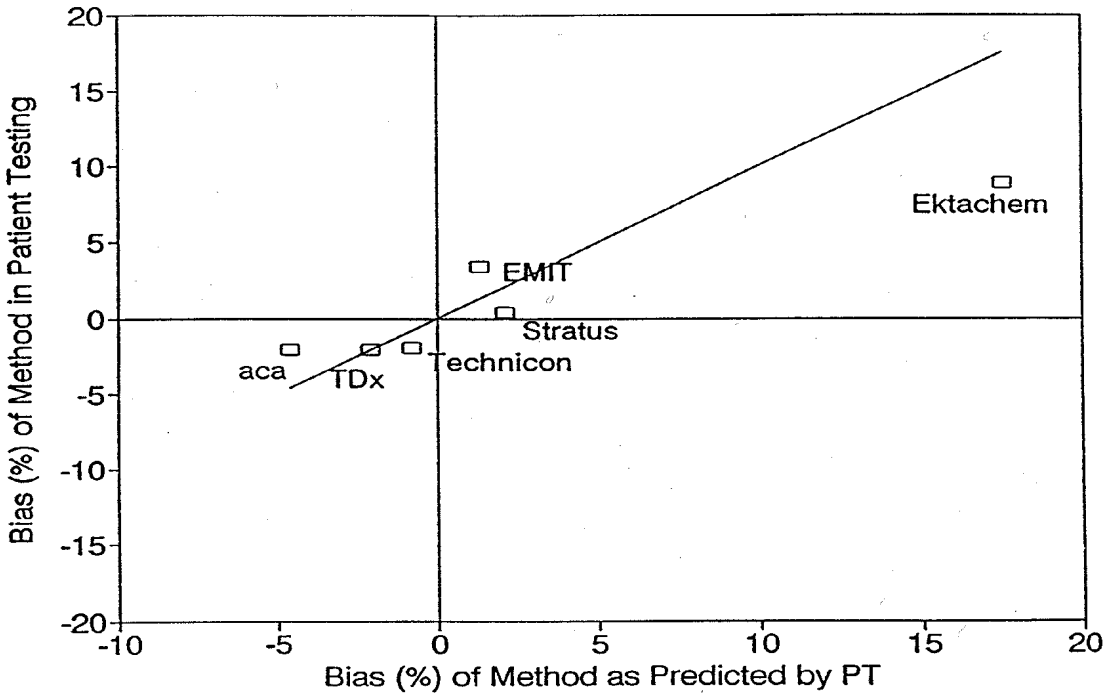

Figure 4. Comparison of bias (%) of method predicted by PT vs. bias (%) of method in patient testing.

Auditing Proficiency Testing

A number of mechanisms have been devised as audits of results oabtained by conventional proficiency testing. Three are reviewed here:

"Blind Submission" of Proficiency Samples

In this schemem, samples used in proficiency tests are distributed to laboratories disguised as "patient" specimens.[22] Although this mechanism may circumvent some special treatment of proficiency test samples, it may not provide the same information, because even if they are treated in an identical fashion, pre-analytical and post-analytical processing may differ (Table 1).

Overt proficiency testing samples enter the system at the analytical phase and are subject to extraordinary reporting, whereas blind proficiency testing samples enter the system at an early phase and are subject to routine reporting.

On-site Proficiency Testing

Some information may be gained from examining routine proficiency testing distributed by scheduled mailing and that presented to laboratories during inspection. In the area of blood pH and gases, the proficiency test program organized by the Wadsworth Center presented specimens in both manners; four sets (3 vials) by scheduled mail; one set (3 vials) presented at the time of unannounced inspection. Blood

gas measurements may represent an analysis where special treatment (increased calibration, replacement of electrodes, etc.) can be effected for routine proficiency testing but not possible at the time of inspection. Results are shown in Figure 3. A high degree of correlation was found between the results of the scores obtained (r^2 = 0.49). A slight, but statistically significant (P< 0.01) by paired t-test), difference was observed between scores obtained by on-site (mean = 87.6%) and mailed (mean = 91.6%) testing routes.

Split-specimen patient testing

To better examine routine laboratory performance, at the time of annual laboratory inspections conducted by the New York State Health Department, we obtained aliquots of each of two sera that had been analyzed for cholesterol or theophylline by the inspected laboratory.[23,24] These aliquots were mailed to our laboratory; we also obtained the clinical results determined and reported for those specimens by the laboratory. Specimens were stored at < -60° C and analyzed by reference methods (CDC modified Abell-Kendall for cholesterol and HPLC for theophylline). Results were obtained for > 200 laboratories. We found that the predictive value of proficiency testing performance in assessing quality of routine testing was high; for theophylline, 100% for predicting substandard reliability of routine patient testing and 94% for excluding substandard reliability of patient testing. Significant correlation was found between analytical bias observed in proficiency tests and that found for patient testing (Fig. 4). For cholesterol, the average difference between participant performance and the reference method was a positive bias of 1.47%; this is equivalent to the overall population bias measured by our routine proficiency testing. Most reported results (88%) were within $\pm$ 10% of the value determined by the reference method. This is similar to performance determined by our proficiency program, where ca. 15 % of results were beyond $\pm$10% of the reference method target value. Using NIH guidelines for risk assessment (200 and 240 mg/dL), 13 specimens (4.4%) were misclassified to a lower risk; 7 specimens (2.4%) were misclassified to a category of higher risk.

We found this manner of auditing laboratory performance effective in that true patient specimen results and the results reported are used in the evaluation process. A similar split- specimen testing study is under way for calcium analysis, using atomic absorption as the reference technique. This analyte meets many of the criteria shared by cholesterol and theophylline (stability, availability of reference methods, a wide variety of analytical procedures) and is an analyte where analytical goals are stricter that current performance ability.

References

1. Peddecord KM, Cada RL. Clinical laboratory proficiency test performance. Its relationship to structural, process and environmental variables. *Am J Clin Pathol*. 1980;73:380-5.

2. Duckworth JK. Voluntary inspection and accreditation and improvement of laboratory performance. The private sector view. In: Quality Assurance in Health Care: A Critical Appraisal of Clinical Chemistry. RN Rand, RJ Eilers, NS

Lawson, A Broughton (Eds.), American Association for Clinical Chemistry, Washington, DC, 1980, pp. 381-90.

3. Copeland BE, Rosenbaum JM. Organization, planning and results of the Massachusetts Society of Pathologists regional quality control program. *Am J Clin Pathol*. 1972;57:676-88.

4. Vincent WF. The Perspective of Connecticut Department of Health. In: Proceeding of Second National Conference on Proficiency Testing. M Davis (Ed.), Information Services, Inc., Bethesda, MD, 1975, p 18.

5. Pinkerton PH, Wood DE, Burnie KL, et al. Proficiency testing in immunohematology in Ontario, Canada, 1975-1977. *Am J Clin Pathol*. 1979;72:559-63.

6. Shephard MDS, Penberthy LA, Fraser CG. Evolution of a national urine quality assurance programme: the Australasian experience, 1981-1983. *J Clin Pathol*. 1984;37:415-23.

7. De Leenheer AP, Thienpont LMR. External quality assessment of laboratory performance in haematology in Belgium: analysis of two and a half year's experience. *Clin Chim Acta*. 1984;144:95-103.

8. Whitehead TP, Woodford FP. External quality assessment of clinical laboratories in the United Kingdom. *J Clin Pathol*. 1981;34:947-57.

9. Jansen RTP, Jansen ADP. A coupled external/internal quality control program for clinical laboratories in The Netherlands. *Clin Chim Acta*. 1980;107:185-201.

10. Hill PG, Kanagasabapathy AS. Improvement in laboratory performance by an interlaboratory quality control programme. *Indian J Med Res.* 1979;6:853-64.

11. Public Law 100-578 (HR 5471) Clinical Laboratory Improvement Amendments of 1988; Section 4-a-1. Congressional Record 134, October 31, 1988.

12. Rej R, Jenny RW, Bretaudiere JP. Quality control in clinical chemistry: characterization of reference materials. *Talanta* 1984;31:851-62.

13. Rej R. Accurate enzyme activity measurements. Two decades of development in the commutability of enzyme quality control materials. Arch *Pathol Lab Med*. 1993;117:352-64.

14. Rej R. Proficiency testing, matrix effects, and method evaluation. *Clin Chem*. 1994; 40:345-6 [Editorial]

15. Rej R, Norton-Wenzel C. Matrix effects and external assessment of laboratory measurements using liquid and lyophilized specimens. *Clin Chem*. 1994;40:1000 [Abstract].

16. Bretaudiere JP, Rej R, Drake P, Vassault A, Bailly M. Suitability of control materials for determination of alpha-amylase activity. *Clin Chem*. 1981;27:806-15.

17. Bretaudiere JP, Dumont G, Rej R, Bailly M. Suitability of control materials. General principles and methods of investigation. *Clin Chem*. 1981;27:798-805.

18. Jenny RW. Interlaboratory evaluation of salicylate interference in colorimetric acetaminophen methods and its clinical significance. *Clin Chem*. 1985;31:1158-62.

19. Rowan RM, Laker MF, Alberti KG. The implications of assaying external quality control sera under "special conditions." *Ann Clin Biochem*. 1984;21:64-8.

20. Cembrowski GS, Vanderlinde RE. Survey of special practices associated with College of American Pathologists proficiency testing in the Commonwealth of Pennsylvania. *Arch Pathol Lab Med*. 1988;112:374-6.

21. Rej R, Jenny RW. How good are clinical laboratories? An assessment of current performance. *Clin Chem*. 1992;38:1210-17; discussion 1218-22.

22. Hansen HJ, Caudill SP, Boone DJ. Crisis in drug testing. Results of CDC blind study. *JAMA* 1985;253:2382-87.

23. Jenny RW, Jackson KY. Proficiency test performance as a predictor of accuracy of routine patient testing for theophylline. *Clin Chem*. 1993;39:76-81.

24. Rej R, Norton CS. External assessment of laboratory cholesterol measurements using patient specimens. *Clin Chem*. 1989;35:1069 [Abstract].