Optimal Evaluation of Proficiency Testing in the Hematology Laboratory : A Systematic Approach

Carolyn D. O'HARA, MD, BSc*, Gisele CHENARD, MLT*2 and George S. CEMBROWSKI, MD, PhD*3

* *Department of Laboratory Medicine and Pathology, University of Alberta Hospital, 4B1.24 Walter C. Mackenzie Centre, 8440-112 Street, Edmonton, Alberta, Canada.*

*2 *Supervisor, Core Laboratory, University of Alberta Hospital, Edmonton, Alberta, Canada.*

*3 *Regional Director of Clinical Biochemistry, University of Alberta Hospital, Edmonton, Alberta, Canada.*

With increasing government regulation of clinical laboratory testing in the United States and other countries, more emphasis is being placed on government monitoring of proficiency testing (PT) results. In the United States, the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88) requires that 5 unknowns be periodically analyzed. Modern hematology analyzers are very stable, and the broad CLIA PT limits are rarely violated. However, the analysis of these 5 unknowns can provide highly accurate information to the laboratorian about the presence of random and systematic error, information that is extremely useful for quality improvement and error minimization. This article describes an almost effortless system for reviewing PT results which consists of a screening rule and 3 control rules used in conjunction with a proficiency results review form. The system is simple to incorporate into the hematology laboratory and will result in the maximization of information about analyzer performance while minimizing effort and time allocated for PT review.

(Sysmex J Int 8 : 72～78, 1998)

Key Words Clinical Laboratory Improvement Amendment of 1988 (CLIA '88), Proficiency Testing, Automated Hematology Analyzer, Quality Control, Quality Improvement, Quality Assurance

INTRODUCTION

Proficiency testing (PT) programs have had similar origins worldwide. Most early PT programs were initiated by one or two enthusiastic individuals who usually sent out a single survey to a small number of laboratories[1]. Later, surveys were instituted on a regular basis, usually by professional societies. One of the very first PT programs was started in 1946 by Belk and Sunderman[2]. They found tremendous inter-laboratory differences among laboratories in Pennsylvania, New Jersey, and Delaware; their findings resulted in the establishment of the first voluntary national PT program in the United States under the jurisdiction of the College of American Pathologists (CAP). Initial PT practices were poorly regulated, and "special treatment" of the PT specimens was prevalent, including the analysis of PT specimens in replicate, reporting the average of PT results, and getting the section's best technologist to run PT specimens[3].

PT, also known as external quality assessment, is now mandatory in the United States and Canada and has become an integral component of clinical laboratory practice. In the United States, the Clinical Laboratory Improvement Amendments of 1988[4] (CLIA '88) made PT performance the most important surrogate indicator of laboratory quality. Currently, for CLIA-specified analytes, the PT provider ships 5 unknowns every four months to the participating laboratory, which analyzes the specimens and then returns analyte concentrations as well as any relevant interpretation. Until now, the focus of the PT provider has been to provide a simplistic report card both to the participating laboratory and the government. Often the report card just indicates acceptability of performance; some providers do not even provide deviations of participant results from peer means expressed as standard deviation indexes (SDI).

We have observed that the review of proficiency results can be very inconsistent, even by the same individual for the same set of results. There is also significant inter-laboratorian variability. Those laboratorians employed in the quality departments of today's commercial reference laboratories can be obsessive when reviewing PT results, and will go to great lengths to characterize even minor analytical error. Other laboratorians have found the assessment and follow-up of PT results time-consuming and tedious, and in fact may passively wait for frank PT failures before making changes that bring their results in line with their peers. We propose a systematic approach that can be used by all laboratorians, whether they are from highly analytical backgrounds such as clinical chemistry, or more qualitative backgrounds such as clinical hematology. This approach can be used to quickly evaluate PT results. It incorporates 4 quality control (QC) rules together with a form that standardizes the evaluation of PT results. The combination of rules and form will enable laboratorians to maximize the information provided by their PT organization as well as alert them to potential problems.

Hematology proficiency testing

CLIA approved hematology PT programs survey 8 different analytes - leukocyte (WBC) count, erythrocyte (RBC) count, hemoglobin (HGB), hematocrit (HCT), platelet (PLT) count, prothrombin time, activated partial thromboplastin time, and fibrinogen. Five commercially prepared test samples are distributed per test event, with 3 test events per year. The test samples are usually chosen to reflect clinically relevant concentrations. Laboratories are expected to process the specimens in the same way that they routinely test patient specimens, with the results returned to the PT program for grading. Acceptable ranges are established using fixed criteria specified by CLIA (***Table 1***), which are expressed in percentages of the target[5]. The goal for the participating laboratories is to have results which fall into the acceptable range; results falling outside of the acceptable range are deemed unsatisfactory.

Laboratories are required to obtain a score of 80% or greater in two of three consecutive test events, and must maintain a cumulative average score of 80% or greater in the most current and the two immediately preceding test events. Two types of scores are calculated[5]. The first, the analyte score, is determined for each testing event and cell identifications, leukocyte differentials, and the other eight analytes. It is calculated from the following formula:

$$\text{Analyte score} = \frac{\text{(Number of acceptable responses for the analyte)}}{\text{(Total number of challenges for the analyte)}} \times 100$$

The second type of score is the testing event score and is calculated for all of the regulated hematology analytes and all of the challenges that it performs in the current testing period. It is calculated from the following formula:

$$\text{Event score} = \frac{\text{(Number of acceptable responses for all challenges)}}{\text{(Total number of challenges for the event)}} \times 100$$

If a laboratory receives a failing score, hematology personnel must take the necessary actions to find, correct, and document any problems in the testing performance. Failing laboratories are usually subject to three principle sanctions - suspension, limitation, or revocation of their CLIA certificate. There may also be alternative sanctions, such as onsite monitoring, monetary penalties, or civil or criminal sanctions. Choice of sanctions is based on multiple criteria and are often determined on a case-by-case basis.

Assuring timely and complete PT analysis and results submission to the PT provider

In a review of PT failures at a large national reference laboratory, only one third of the failures were due to analytic errors[6]. One third were clerical errors, and the remaining third were classified as miscellaneous and included variables such as the lack of a proper peer group, inability of the PT sample to perform as a patient sample, and error on the part of the PT provider. The authors stressed that laboratories should handle PT specimens the same way they handle their routine patient samples, including assigning bar code labels, analysis, and reporting. We recommend dividing the PT process into three components - pre-analytical, analytical, and post-analytical[5] in order to ensure accurate PT analysis and evaluation.

Table 1 *Comparison of CLIA PT error limits to typical within-instrument CVs of the Sysmex SE series* and the Sysmex CA-6000* The magnitude of error is the size of the systematic error that will cause 100% of the PT results to be outside the CLIA limits (assuming no random error is present).*

** These data were obtained with permission from the Sysmex Corporation.*

Analyte	CLIA PT error limits	Typical within-instrument CV (near midpoint of normal range)	Magnitude of error
HGB	Mean + 7%	1%	7 CVs
WBC count	Mean + 15%	2%	8 CVs
PLT count	Mean + 25%	2%	12 CVs
HCT	Mean + 6%	1%	6 CVs
RBC count	Mean + 6%	1%	6 CVs
Prothrombin time	Mean + 15%	3%	5 CVs
Activated partical thromboplastin time	Mean + 15%	3%	5 CVs
Fibrinogen	Mean + 12%	5%	2 CVs

Pre-analytical

CLIA '88 requires that government mandated PT only be done on the primary instrument[4]. If a number of hematology instruments are used in a laboratory, optimal PT performance can be achieved by using the most accurate instrument, provided it is routinely used. Significant deviations from the PT mean can occur if a laboratory utilizes alternate methods of calibration, such as using reference methods or whole blood specimens analyzed on another instrument[5]. It is thus important to use the manufacturers' calibration materials in the manner recommended by the manufacturer, thus minimizing inter-instrument variation.

Virtually all hematology laboratories use three level quality control products, which are usually provided by the instrument manufacturer. Many laboratories do not effectively utilize the QC data and may miss opportunities to detect systematic errors. Hackney, et al.[7] have outlined a useful multi-rule control procedure for hematology.

Analytical

There will always be some pre-selection of time, operator, and instrument for PT analysis; the laboratorian needs to balance the need for successful performance with what is ethically acceptable[5]. We recommend that laboratories do not miss legitimate opportunities to optimize their PT, such as reporting the results of a second instrument to the PT provider, *e.g.*, the CAP. The analysis of PT products by alternate systems will decrease the

Survey _____ Scheduled Ship Date (from PT provider): _____
Receipt Date: _____
Due Date (3 days before required mailing date): _____

Notify supervisor if period between ship date and receipt date exceeds 7 days.

Systems involved in this survey: _____

ACTION	DATE	TECH.
Make a copy or copies of the PT report form		
Check all tests to be performed		
Check to make sure there will be no scheduling problems or scheduled maintenance		
Review QC		
Schedule dates to perform tests (consider sample stability after reconstitution)		
Check sample integrity: unlabeled, missing, broken, etc.		
Read all survey instructions: test handling, precautions, reconstitution, stability, pretreatment, calculations, reporting, etc.		
Reconstitute according to instructions; date, initial and include the expiration date on all survey samples		
Compare sample volume to test volume requirements		
Run PT specimens with routine patients		
Refrigerate whole blood specimen		
Freeze aliquot of coagulation specimen		
Record results on the report form copy according to reporting instructions		
Record date and initial on report form copy when tests are performed		
Attach a copy of the instrument printout and QC chart		
Have another technologist check transcription		
Return the report form copy to the supervisor or survey coordinator before or by due date posted		
Transcribe data from the report form copy to the original PT report form and have transcription checked		
Complete the additional information requested on the report form, including method and instrument codes		
Sign the CLIA-required attestation statement		
Keep copies of all reports for your files		
Mail the original report form and document return date		

Fig. 1 *Proficiency survey checklist for PT sample processing, analysis, and reporting*

reporting of incorrect results. Similarly, if the hematology morphology review protocol indicates the need for pathologist or specialist review, their effort should be enlisted to more accurately identify the morphologic abnormality. If either the critical value or linearity policy indicates the need for re-analysis, then that policy should be invoked to reanalyze any PT specimen with outlying values (replicate analysis reduces random error). Cembrowski, et al.[5] recommend laboratories use a proficiency survey checklist (***Fig. 1***) for PT samples; this form provides consistency in PT processing and analysis and helps to prevent unethical special handling practices.
Today's hematology analyzers are very stable, as indicated by their small long term CVs. On the other hand, the CLIA limits used for PT acceptance are very broad (***Table 1***)[4]. For example, the typical within instrument CV for HGB is 1%, whereas the CLIA error limits are +7%. If the CLIA PT limits are divided by the instrument's CV, the resulting ratio represents the magnitude of the shift that must be encountered to cause an unacceptable result. Using the hemoglobin example, a 7 CV shift is required to consistently cause erroneous results. The use of these broad CLIA limits is of little benefit to most laboratories.

Post-analytical

Post-analytical activities involve transferring the PT results to the data input forms of the PT provider. The data transfer step can be compromised by clerical error, and various checks should be implemented to minimize this error. At the very least, a supervisor should double-check all results transcribed onto the PT data forms.

Evaluation of the PT results report

When a laboratory director or supervisor receives PT results, s/he often inspects the report looking only for failures. Thus there is the potential to miss significant, potentially correctable errors. This superficial review is due to at least 3 different factors: 1) emphasis placed on "passing" PT by the regulators and PT organizations, 2) laboratory directors and supervisors lacking the knowledge to analyze the results effectively, and 3) PT results being delayed so long that investigating PT results is no longer beneficial. Significant analytical error can be present in the PT data and can be uncovered by the use of our system. These errors can be easily characterized, investigated, corrected or minimized and will result in higher quality testing and a lower prevalence of clinically significant errors.

Algorithm

The rule system outlined in ***Fig. 2*** was developed by Cembrowski, et al.[8] using computer simulations of CAP clinical chemistry PT data. In these simulations, the probability of detecting PT problems was calculated for varying magnitudes of systematic error (shifts) or random error (increased imprecision). The algorithm has been slightly modified since its original description in order to increase its specificity and decrease sensitivity of detecting small, usually unimportant errors. The method involves the sequential application of several QC rules analogous to Westgard's multi-rule QC procedure[9]. The algorithm begins with a simple screening rule applied to the sets of 5 proficiency results. If the screening rule is violated, the data are further tested for significant analytical shifts and random error. The method requires that the PT results be reported or converted into standard deviation indexes (SDIs), which are calculated as SDI = (result − mean)/SD. This index represents the number of SDs each result is from the mean. If your PT program doesn't offer SDI results, then cajole them into including SDIs in their reports (if they are unwilling, consider changing PT providers). PT organizations usually generate two sets of SDI values, depending on the laboratory's peer group mean or the all-method mean. We recommend using the peer group results, as most laboratories can do no better than obtaining the same results as those laboratories with identical instruments and reagents.

The rules follow and are shown graphically in ***Figs 3- 6*** [10];

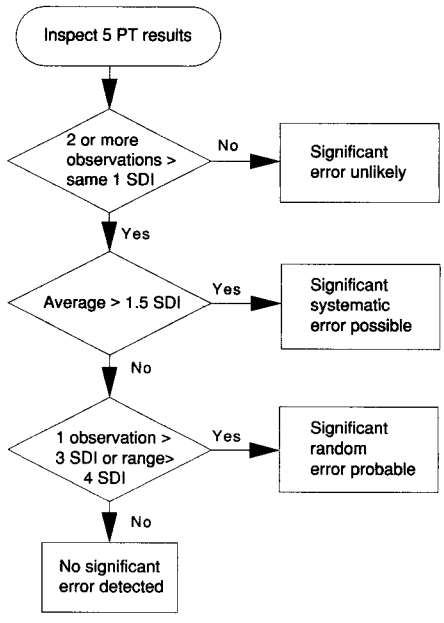

Fig. 2 *Algorithm used to review groups of five PT results*

1) Screening rule: 2/5 > ± 1.0 SDI

If two or more observations are outside the same +1.0 or −1.0 SDI limit, the screening rule is violated, and the PT data are tested further by rules specific for systematic and random error.

2) Mean rule: |Mean| > 1.5 SDI

If the average of the five observations exceeds +1.5 SDI or is less than −1.5 SDI, then significant systematic error is present.

3) 1−3SDI rule

If one or more observations are outside either the +3.0 SDI or the −3.0 SDI limits, there is a high probability of random error.

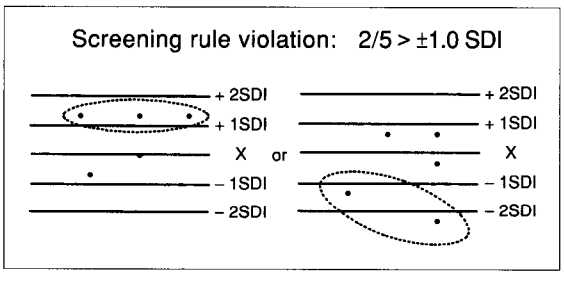

Fig. 3 *Example of the screening rule violation for detecting systematic and random error*

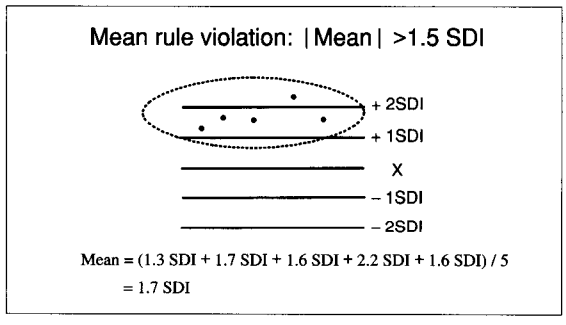

Fig. 4 *Example of the mean rule violation for detecting systematic error*

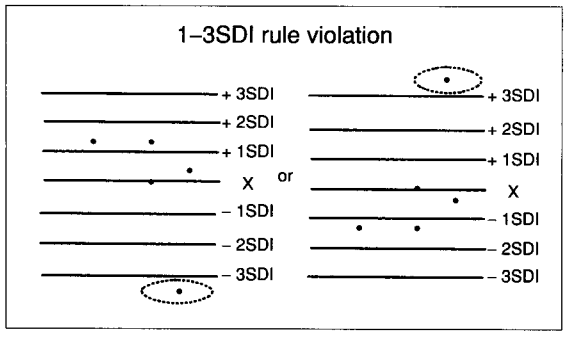

Fig. 5 *Example of the 1−3SDI rule for detecting random error*

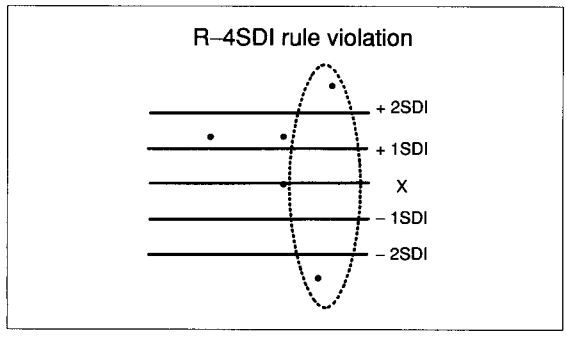

Fig. 6 *Example of the R−4SDI rule for detecting random error*

4)R–4SDI rule
If the range or difference between the largest and the smallest PT results exceeds 4.0 SDI, there is likely random error present.

To summarize, the process involves applying a simple screening rule to the PT data sets. If the data do not violate the screening rule, there is likely no significant analytical error present. If the screening rule is violated, the data are then applied to rules specific for systematic error (mean rule) and random error (1–3SDI and R–4SDI rules). If the screening rule is violated and no systematic or random error is discovered, the SDI values of the next analyte are inspected.

Proficiency evaluation form
An attendee at one of our early workshops on PT evaluation suggested that the algorithm be used in conjunction with a form which summarized the data, outlined the plan to evaluate systematic and random error, and documented follow-up, including letters from regulatory agencies. We designed a simple form (***Fig. 7***), that outlines the essential elements needed to interpret PT data. The form summarizes the algorithm and provides space for investigative and corrective comments. The form worked so well with diverse chemistry supervisory staff that we are now applying it to our hematology PT results.
The staff involved found the system made them more confident in the quality of the PT results, since the system is sensitive to errors that may not be flagged by the PT organizations. Another benefit of implementing the system is that it makes the PT process more educational, thus addressing a limitation of PT described in a recent review[11]. The system has enabled our laboratory staff to more confidently compare our laboratory performance with others. If our PT results are very biased compared to that of the peer group, then our laboratory staff are confident that an investigation will demonstrate a real problem. Another advantage is that the system can be used as a "check" of the PT organization. On occasion, we find that PT data that the algorithm classifies as having a low probability of error are characterized as erroneous by our PT organization. The most important advantage of the system is its consistency. It deters staff from searching for only flagged results, and the improved documentation increases the consistency of follow-up and investigation of potential problems. Thus, even if our laboratory receives a letter from a regulatory body

HEMATOLOGY PROFICIENCY RESULTS REVIEW
Five Specimens Per Analyte

LAB: _____ Date of Analysis: _____
Proficiency Sample Name: _____ Sample ID No.: _____
Lab Accn. No.: _____

Analyte	Rule Violated Screening Rule: 2/5 > ± 1.0 SDI Systematic Error: Mean > 1.5 SDI Random Error: 1-3 SDI or R-4 SDI	Mean (SDIs)	Investigation Check QC [including QC during dates of PT Challenge]	Action	Pathologist Comments

Reviewed By:
Head Hematology Tech _____ Date: _____ Letter from Accrediting Agency ❑ Yes Dated: _____
Pathologist _____ Date: _____ Analyte(s): _____
Laboratory Director: _____ Date: _____ Report Sent: ❑ Lab Manager ❑ Hospital Administrator

Fig. 7 *The hematology proficiency testing review form*

about a failure, the problem may have already been solved or at the very least is being investigated.
Efficient use of the form requires training of supervisory staff and instrument operators in the interpretation of PT results and the investigation of outliers. The operators should be encouraged to have the "first crack" at PT evaluation. This practice is a useful education tool and encourages laboratory employees to suggest and implement possible "quick fixes". It is imperative that the PT surveys get to those who will be interpreting and acting on the reports. PT surveys are of little use under stacks of paper on a laboratory manager's desk. It is important for the laboratory director to schedule a meeting with the head technologist within a week of PT receipt to review any investigations and identify potential trouble spots. In this way, all levels of laboratory personnel are actively involved in the PT process and are well informed of any changes, from minor tinkering to radical system changes.

Application of the system to real examples

The following examples demonstrate the use of the system. Portions of actual data from CAP reports are presented (***Tables 2 - 5***).

Example 1: Problems with the PLT count?
Table 2 shows the data from a CAP survey for *PLT* count. Applying the algorithm, it is apparent that specimens 1, 3, and 5 violate the 2/5 > ± 1.0 SDI screening rule. When the 5 observations are averaged, the result is 1.2 SDI, which implies that significant systematic error is not present. The data are then checked for random error with the 1–3SDI and R–4SDI rules, which are not violated. The conclusion: there is likely no significant error, and no further investigation is necessary. Future PT surveys revealed no problems with the platelet count data.

Example 2: The RBC count
Table 3 illustrates CAP PT survey data for the RBC count, and shows a violation of the 2/5 > ± 1.0 SDI screening rule. All 5 of the specimens are greater than 1.0 SDI. The average of all observations (the bias) is 1.9 SDI, which suggests that significant systematic error is possible. None of the rules for random error are violated. Investigation of the QC data indicated a long-term positive bias with some out of control data. Investigation showed no obvious error, and the red blood cell channel was re-calibrated.

Example 3: WBC count data showing a positive bias
The data shown in ***Table 4*** shows a violation of the screening rule, with all five observations exceeding + 1.0 SDI. The average bias is + 1.7 SDI, indicating systematic error is likely present. When we investigated the QC data at the time of reporting the WBC data, we observed a slight positive shift but no violations of our control rules. The WBC count was re-calibrated downward, and the next PT survey results were in line.

Example 4: The hazards of living in a northern climate
We use this particular example to illustrate the process of investigating some very interesting PT results. ***Table 5*** displays PT summary data for the WBC count. There is violation of the 2/5 > ±1.0 SDI screening rule, with samples 1 and 3 exceeding the +1.0 SDI limits. When the 5 observations are averaged to calculate the bias, the result is +1.2 SDI, which indicates no significant systematic error is present. When checked for random error, specimens 1 and 3 violate the 1–3SDI rule, and the R–4SDI rule is violated by specimens 1 and 4. This indicates the presence of extreme random error, with significant excursions in both directions.

When results for additional analytes were examined, many were aberrant, with most exhibiting severe random error and multiple failures. After further investigation, it was discovered that the samples were sent and analyzed in February, which usually is the coldest month of the year in northern Alberta. We consequently assumed that the samples had frozen somewhere between Chicago and Edmonton, giving us multiple outlying results. The take home message: be extra suspicious on very cold or hot days, especially if at a great distance from the CAP, such as laboratories in Japan and Canada.

Table 2 *PLT count PT data for example 1*

Specimen	Your result	Mean	SD	SDI
1	160	152.6	4.8	+ 1.5
2	78	75.3	3.9	+ 0.7
3	515	483.3	18.2	+ 1.7
4	86	83.2	3.6	+ 0.8
5	207	197.8	7.3	+ 1.3

Table 3 *RBC count PT data for example 2*

Specimen	Your result	Mean	SD	SDI
1	4.96	4.844	0.065	+ 1.8
2	3.95	3.852	0.052	+ 1.9
3	4.76	4.600	0.065	+ 2.5
4	4.04	3.959	0.059	+ 1.4
5	2.09	2.042	0.029	+ 1.7

Table 4 *WBC count PT data for example 3*

Specimen	Your result	Mean	SD	SDI
1	5.7	5.49	0.18	+ 1.2
2	18.9	18.19	0.41	+ 1.7
3	4.9	4.66	0.16	+ 1.5
4	10.0	9.50	0.26	+ 1.9
5	3.4	3.08	0.15	+ 2.1

Table 5 *WBC count PT data for example 4*

Specimen	Your result	Mean	SD	SDI
1	8.8	6.75	0.56	+ 3.7
2	10.7	10.52	0.25	+ 0.7
3	8.0	7.37	0.19	+ 3.3
4	20.1	20.99	0.43	− 2.1
5	27.1	26.80	0.57	+ 0.5

CONCLUSION

This article illustrates a simple yet effective approach that can be utilized when evaluating hematology PT results. Implementing the four rules in combination with the form yields a rational, productive effort that can more readily detect potential errors and/or demonstrate the need for improvement. Most importantly, laboratory staff will be better trained, far more proactive and more confident in their abilities for improvement. No longer do they need to wait for the quality department to issue decrees about the quality of their PT results, nor do they need to passively wait for an admonishing letter from a regulator. Rather, the laboratorians can assess and investigate their PT results as promptly as the day of receipt.

References

1) *Buttolph MA : Evolution of proficiency assessment programs in Europe. In proceedings of the second national conference on proficiency testing. pp. 10, Washington D.C., October 1975.*

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

3) *Cembrowski GS, Vanderlinde RE : Survey of special practices associated with CAP proficiency testing in the Commonwealth of Pennsylvania. Arch Pathol Lab Med, 112 : 374-376, 1988.*

4) *Department of Health and Human Services : Health care financing administration. Clinical Laboratory Improvement Amendments of 1988; Final Rule. Federal register, 57 : 7001-7288, 1992.*

5) *Cembrowski GS, et al. : A systems approach to assure optimal proficiency testing in the hematology laboratory. Clinics in Laboratory Medicine, 13 (4) : 973-985, 1993.*

6) *Gambino R, Mallon P, Woodrow G : Managing for total quality in a large laboratory : Some examples. Arch Pathol Lab Med, 114 : 1145-1148, 1990.*

7) *Hackney JR, Cembrowski GS, Carey RN : Quality control in hematology, eds. Cembrowski GS, Carey RN : Laboratory Quality Management : QC and QA. pp. 186-212, Chicago, ASCP Press, 1989.*

8) *Cembrowski GS, Hackney JR, Carey RN : The detection of problem analytes in a single proficiency test challenge in the absence of the health care financing administration rule violations. Arch Pathol Lab Med, 117 : 437-443, 1993.*

9) *Westgard JO, et al. : A multirule Shewhart chart for quality control in clinical chemistry. Clin Chem, 27 : 493-501, 1981.*

10) *Cembrowski GS, et al. : Pump up your PT IQ. pp. 46-52, Medical Laboratory Observer, January, 1996.*

11) *Shahangian S : Proficiency testing in laboratory medicine : Uses and limitations. Arch Pathol Lab Med, 122 : 15-30, 1998.*