

# Proficiency testing versus QC-data-comparison programs

October 2003



**Hans Bjarne Kristensen**

Radiometer Medical A/S  
Åkandevvej 21  
2700 Brønshøj  
Denmark

Evaluation of the quality of the total testing process of a laboratory is very important to assure that the results supplied by the laboratory are correct and can be used for decisions by the personnel ordering the tests. In this paper Proficiency Testing (PT) or External Quality Assessment (EQA) is described.

PT is very important in evaluating the quality of the analytical phase of the testing process. Furthermore, QC-data comparison is described. Participation in a QC-data-comparison program is a very important supplement to PT participation.

A QC-data comparison will very often give additional information that is not obtained in a PT program. This could, for example, be better information regarding the imprecision parameters, repeatability and reproducibility.

Both PT and QC-data comparison will mostly focus on the analytical phase. Other phases of the testing process, the preanalytical phase and the postanalytical phase, must be evaluated using different tools. These tools are not described in this paper.

## Introduction

All laboratories are involved in a number of activities in relation to quality assurance and quality assurance improvements, as these activities are essential to the general testing quality.

Results that are not reliable are generally useless, a waste of time and money and can lead to wrong decisions. Many laboratories, therefore, participate in a PT program and some laboratories also in a QC-data-comparison program.

The purpose of this paper is to describe the general features of PT and QC-data comparison and to document that these two activities can supplement each other, so that a strong surveillance of the quality of the analytical phase is obtained.

## Proficiency testing

### Purpose of Proficiency Testing

Proficiency Testing (PT) or External Quality Assessment (EQA) is a program in which multiple specimens are periodically sent to a group of laboratories for analysis.

The purpose of such a program is to evaluate the laboratory performance with regard to the testing quality of patient samples.

This evaluation is performed by comparison of the results of the PT samples within a group of similar methods, a so-called peer group, and from this comparison determine the performance of the individual laboratory with regard to imprecision, systematic error and human error concerning the PT samples.

Imprecision is measured as repeatability and reproducibility. Systematic error is measured as bias. From the performance on PT samples the performance on patient samples are deduced.

A PT program will normally only evaluate the analytical phase of patient sample testing. The preanalytical - and postanalytical - phase has to be evaluated using other tools.

### **Procedure for Proficiency Testing**

A number of organizations offer PT programs. Among these are regulatory and accrediting bodies. In the USA, the College of American Pathologists (CAP) offers numerous PT programs. In Europe, a number of national EQA organizations offer EQA programs, for example WEQAS in the UK. In Australia, the Royal College of Pathologists of Australia (RCPA) offers PT programs.

The general procedure is that the PT provider, at regular intervals, distributes PT samples to the laboratories. The laboratories determine the value of the analyte in question and report the result back to the PT provider. The PT provider performs statistical analysis of all the results and sends a report to each laboratory from which the laboratory can evaluate its own results in comparison with other laboratories.

Furthermore, the PT provider sets up the PT acceptance criteria. Most commonly, the PT results are grouped by method, and the mean and standard deviations (SD) are calculated.

An acceptance criterion can then be that a laboratory's result must be within the mean  $\pm$  3 SD of the results

from the laboratories using the same method (the peer group). In other cases, fixed range grading is used where the successful value must be within fixed limits in relation to the mean value.

The material used for PT programs will usually be prepared as a large homogeneous lot. The material will generally be an aqueous material or a lyophilized material; in these cases the "matrix" effects of the PT samples should be considered.

In some cases, in which the PT participants cover a very narrow geographical area, the material can be a whole-blood product or a tonometered hemolysate product.

### **Which information can be derived from the PT results?**

Shahangian [1] has reviewed a large amount of literature on PT, published in the period 1987-1997, and draws a number of interesting conclusions, some of which are mentioned in the following.

Participation in a PT program is a very useful tool in the evaluation of the performance of a laboratory; however, this has to be supplemented by other methods for the evaluation of the testing quality of the total testing process of the laboratory.

PT appears to be mainly a measure of the analytical performance. Concerning PT as a measure of accuracy, a number of investigations indicate positive correlation between accuracy in testing biological specimens and that obtained in PT programs.

However, a PT program has a number of inherent limitations; among these is the low number of PT samples tested, making it difficult to establish the correct "true value".

Furthermore, repeatability is difficult to measure if there is not sufficient PT material for repetitions. Furthermore, matrix effects from using aqueous material or lyophilized materials can distort the correlation between accuracy and imprecision calculated from PT samples and from real biological samples.

An EQA study performed by A. Thomas [2] evaluates the matrix effect of protein-based aqueous material and a hemolysate blood product.

This matrix effect is very different in blood gas analyzers from different companies and also very different from analyte to analyte.

This study emphasizes the importance of the grouping of methods in a PT program, so that laboratories using the same method are compared.

The NCCLS Guideline on PT [3] describes how to investigate unacceptable PT results. This is very important as a quality improvement tool for the total testing process.

In conclusion, the information that can normally be derived from participation in a PT program is the following:

- A bias in relation to the mean value of the peer group.
- By evaluating this bias over several occurrences of the PT program, the stability of the method can be measured.
- In some cases repeatability can be derived.
- Evaluation of unacceptable PT results can lead to quality improvements.

Benefits and limitations of participation in a PT program.

Participating in a PT program has the following benefits:

- An independent evaluation of the general performance of the laboratory.
- A reasonably good estimate of the laboratory's bias for a particular analyte in relation to a peer group.
- A possibility of evaluating the long-term stability of the method in relation to the peer group.
- If the PT acceptance criteria are not met, the investigation to reveal the cause of this will often result in a quality improvement affecting the real biological sample testing.
- The importance of meeting the PT acceptance criteria will focus the laboratory on quality

assurance issues such as daily QC measurements, QC-data comparison, training of personnel, standard operating procedures, and maintenance of equipment and will generally improve the quality of the testing process.

The limitations of participation in a PT program.

- The interval between the PT occurrences is normally from one to six months, thus giving a relatively weak surveillance of the short-term testing quality of the laboratory.
- The number of PT samples is often low, limiting the possibilities of the evaluation of repeatability.
- The "matrix" of the PT samples is normally different from the "matrix" of real biological samples, thus limiting comparison between different methods. This also implies that true values for the PT samples are seldom determined using a reference method, as the true value will be of limited use because of the "matrix" effects.

Besides this, the price of participating and the resources used for the testing of PT samples can be a limiting factor for the participation in PT programs.

## QC-data comparison

### What is QC-data comparison?

A QC-data comparison program is very similar to a PT program. The difference is that the PT program is based on PT samples that are distributed to the participants, whereas the QC-data-comparison program is based on the daily QC measurements that the laboratory performs.

These are then evaluated by the QC-data-comparison provider and reported back to the laboratory.

### Purpose of QC-data comparison

The purpose of a QC-data-comparison program is to evaluate and improve the quality of the analytical phase of the testing process.

This is done by monitoring the long-term stability of the analytical test process, so that timely corrective actions can be implemented.

This monitoring of the long-term stability is a very important supplement to the monitoring of the short-term stability done by performing daily QC measurements.

Furthermore, the purpose is to document the quality of the analytical phase in reports. Such documentation is always useful and could be necessary in connection with accreditation of the laboratory.

Participation in a QC-data-comparison program will also be a very useful supplement to participation in a PT program, as additional information will be obtained from the QC-data-comparison program.

**Procedure for QC-data comparison**

A number of companies producing QC samples for daily measurements offer participation in a QC-data-comparison program.

A procedure could generally be as described in the following. The company supplies the laboratory with QC samples used for daily QC measurements. The laboratory performs daily QC measurements on the method used and collects the results.

These results are then sent to the QC-data comparison provider on a regular basis, for example once a month. The results can be sent either as a paper copy, electronically stored on a diskette, or as a file via the Internet.

The QC-comparison provider receives the data and performs statistical calculations on the data from the laboratory in question and also compares the data with all the data from other laboratories using the same methods, the peer group.

The calculation results in a number of reports, and from these the laboratory can choose the ones that fit the need of the laboratory. The reports are either mailed to the laboratory by mail or e-mail or can be available online on the Internet.

When the laboratory receives the reports, the laboratory will evaluate these with regard to information, whereupon corrective actions are performed, if required.

**Which information can be derived from the QC-data-comparison reports?**

A QC-data-comparison report will normally give the calculated mean value and SD of the QC results for the laboratory and for the peer group. In Figs. 1 and 2, examples of a report from a QC-data-comparison system are given.

These reports are used to demonstrate the information that can be obtained from a QC-data-comparison system:

S7735	Lot 42	Mean	1 SD	SDI	N	Month comp. to prev. history - 2 SD M + 2 SD
pH	Your month	7.107	0.006	-0.5	32	
	Your history	7.109	0.004		315	
	Peer 1 hist.	7.108	0.004	-0.5	8660	
	Peer 2 hist.	7.108	0.004	-0.5	136	
pCO <sub>2</sub> mmHg	Your month	63.7	0.9	-0.6	33	
	Your history	64.4	1.2		319	
	Peer 1 hist.	64.9	1.0	-1.3	8647	
	Peer 2 hist.	64.6	1.0	-1.4	136	

FIG 1. Please note that some of the report information has been left out due to graphical limitations.

From Fig. 1 and the pH parameter, the following information is available:

- Mean value, SD and number of results for the current month (N) for the laboratory's own method.
- Mean value, SD, Standard Deviation Index (SDI), and total number of results (N) for the laboratory's own method for the whole period that the laboratory has participated in the program. Histogram showing mean value for current month, compared with the 95 % confidence interval for results for the whole period except current month.
- For two different peer groups (Peer 1 and Peer 2), mean value, SD, SDI, and total number of monthly mean values (N). Histogram showing the laboratory's mean value for the current month (the black dot), compared with a 95 % confidence interval for monthly mean values of the peer group. The 95 % confidence interval is calculated as mean value  $\pm$  2 SD.
- From these data, own mean value for the current month can be compared with own mean value for the whole period, except the current month. This can be used for evaluation of the long-term stability of the method.
- Similarly, own mean value for the current month can be compared with the mean value of the monthly mean values of the peer group. This can be used to evaluate if the bias of the methods is close to the bias of comparable methods and thus to conclude if the performance of the method is within natural statistical variation compared with the peer group.
- The SD of own data for the current month can be compared with the SD of own data for the whole period to evaluate if the repeatability of the method is stable, and the SD of own data for the current month can also be compared with the SD of the data of the entire peer group. This will evaluate if the repeatability of the method is comparable to the general repeatability of the whole peer group.

In Fig. 2 and the pH parameter, the monthly mean value (black dot) for the laboratory's own results is plotted as a function of time (month). Similarly, the 95 % confidence interval for monthly mean values of the peer group (gray bars) is plotted as function of time.

This indicates the reproducibility of the method compared with the general reproducibility of the peer group (the gray bars).

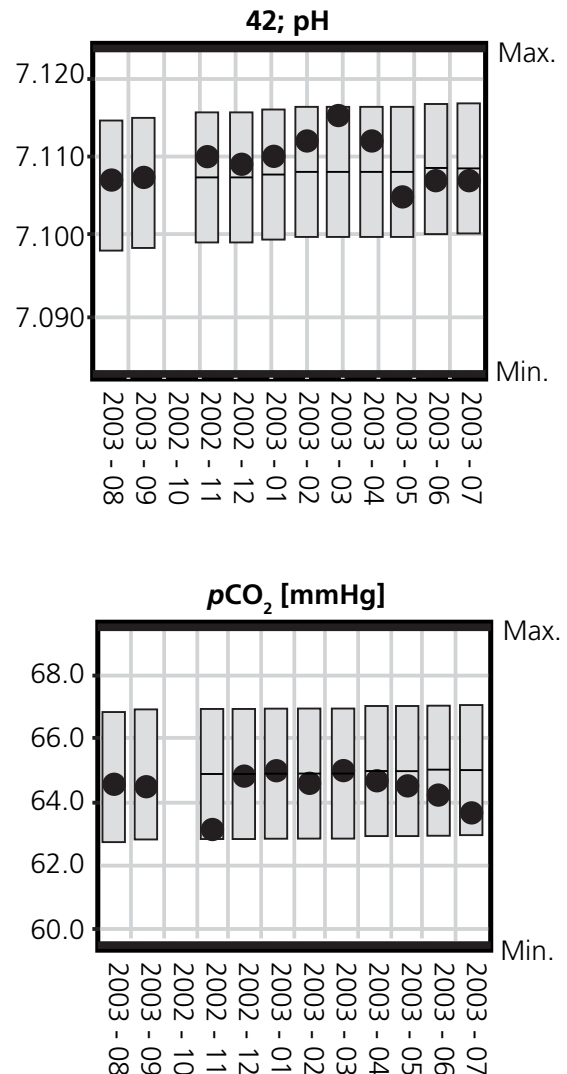


FIG 2.

### Benefits and limitations of participation in a QC-data-comparison program

Participating in a QC-data comparison has the following benefits :

- A very good estimate of the laboratory's bias, repeatability and reproducibility compared with a peer group.
- Easy method for detecting trends, so that corrective actions can be performed in a timely way.
- From the behavior of the method in the QC-data-

comparison program, the behavior of the method in a coming PT occurrence can often be deduced, if the QC material and the PT material has similar "matrix".

- As the QC-data-comparison program is effective in the evaluation of the method compared with a peer group, it means that if the laboratory has several analyzers using the same method, and if these analyzers are both participating in PT and QC-data comparison, a less frequent participation in PT occurrences could be implemented, as the analyzers' long-term stability is monitored by the QC-data-comparison program.
- An effective documentation of the performance of the laboratory's method for the entire period of participation is obtained from the reports.

The limitations of participating in a QC-data-comparison program are:

- The QC-data-comparison provider must offer large peer groups, otherwise the results of the comparison will be relatively uncertain.
- The "matrix" of the QC samples is normally different from the "matrix" of the real biological samples, thus limiting the possibility of comparing different methods.

## Conclusion

Participating in a PT program, performing daily QC measurements, and participating in a QC-data-comparison program are important elements in the quality assurance of the analytical phase of a total testing process. The two elements, a PT program and a QC-data-comparison program, will in many cases supplement each other so that optimal information regarding the method will be obtained by participating in both programs.

## References

1. Shahangian, S. Proficiency testing in laboratory medicine. Arch Pathol Lab Med 1998; 122: 15-30.
2. Thomas MA. An evaluation of the performance of blood gas analysis in the UK using tonometered haemolysate material. Clin Chim Acta 2001; 307: 125-33.
3. Using Proficiency Testing (PT) to Improve the Clinical Laboratory; Approved guideline. NCCLS Document GP27-A. Villanova, Pa.: NCCLS, 1999; 19,15: 1-22. [ISBN 1-56238-381-7].