

Quality control of multi-profile blood gas testing

June 2001



Jesper H. Wandrup

Radiometer Medical A/S
Åkandevvej 21
DK-2700 Brønshøj
Denmark

Whether performed by one “traditional” or several “point-of-care” devices, all Multi-Profile Blood Gas (MPBG) results must be accurate and precise if they are to be effective in managing critically ill patients.

Optimal analytical quality control of multi-profile blood gas analysis requires a good quality assurance planning process covering aspects of both analytical and clinical significance in the service provided.

Introduction

Much theoretical and practical work regarding analytical Quality Assurance (QA) in clinical chemistry has been done within routine serum and plasma analysis. Thus very nice general tools and guidelines for quality planning are available through various sources.

Further guidance and details can be found in both government regulation programs, accreditation programs, and professional consensus documents from NCCLS (National Committee for Clinical Laboratory Standards) and IFCC (International Federation of Clinical Chemistry).

The various analytical quality assurance systems and programs in the literature seem to work very nice in stable analytical batch-testing systems, where analytical processes and stable quality control materials are available.

Thus analytical acceptance and reporting of patient results in these kinds of systems are closely linked to analyzing and evaluating quality control results in a very well-defined analytical run environment.

Here, an analytical run is defined as an interval (i.e., a period of time or series of measurements) within which inaccuracy and imprecision of the measuring system are expected to be stable.

However, the nature of a multi-profile laboratory blood gas service as well as the analytical system of whole-blood analysis seem to create practical quality planning difficulties different from those found elsewhere in routine clinical chemistry testing.

Nonetheless, it is still important that blood gas results

are accurate and precise, if they are to be effective in managing critically ill patients.

The present paper tries to address some of the difficulties and problems encountered when setting up an optimal not too complicated QA system for multi-profile blood gas analysis using the state-of-the-art approach as for example described by QA pioneers like Jim Westgard [1-3], Callum Fraser [4-6] and [8], George Cembrowski [7] and Per Hyltoft Petersen [8], as well as many other great QA specialists [9-11].

As a rookie within this highly sophisticated scientific laboratory field my first goal has been to read up on the different QA theories and increase my own knowledge on this interesting subject.

Secondly, I have put this QA knowledge into perspective with my experience with analytical characteristics of multi-profile blood gas technology available on the market today.

Three key points in qa planning

According to the approved NCCLS [12] guideline C24-A2 on "Statistical Quality Control for Quantitative Measurements: Principles and Definitions" the first three important steps recommended when setting up an optimal quality assurance program for multi-profile blood gas services are:

1. Define the quality requirements for the test
2. Determine analytical performance characteristics of the measurement
3. Identify candidate statistic quality control strategies

In conclusion, setting up MPBG services demands that the laboratory first of all defines clinical needs and requirements and sets up analytical goals as well as a QC procedure that maintain such goals in the short and long term for the quality that one wants to create.

Define the quality requirements for the test

Analytical goals are performance characteristics of methods or devices that facilitate optimal patient care.

It incorporates defining maximum allowable analytical imprecision and bias estimates for the devices that one wants to provide the analytical service with.

The choice of MPBG devices is based on a blend of clinical needs, analytical requirements, cost-effectiveness, and ease of use.

If no immediate medical use will be made of the results, then a delay in reporting the results to gain the best analytical data will not be a significant loss.

If the laboratory results are to be used immediately and are related to significant medical decisions, then both extreme analytical performance characteristics and quick delivery of results are important issues to consider.

The concept of understanding the clinical impact when choosing devices for general STAT laboratory services is an important task to address when providing a multi-profile blood gas service.

In the above view, manufacturers are supposed to specify analytical performance characteristics for their instruments. However, reading and comparing analytical specifications of different devices seem not an easy task for end users.

Thus, some manufacturers specifying analytical characteristics of their STAT products only give SD for within-run precision of the same instrument. Some claims are for QC-materials, some are for whole blood. Instrument-to-instrument variations, sensor-to-sensor variation for a statistically significant number of devices are very often not given.

All in all, there seems to be no accepted standard on how manufacturers specify important analytical characteristics of their devices. Very often, STAT laboratory

services consist of several devices from the same vendor or several devices from different vendors.

In this case, the analytical characteristics and performance specifications such as within-run variation, between-day variation, as well as between-instrument variations should be specified for analyzing whole blood, micromodes, etc. From such analytical performance characteristics the end users would be able to judge clinical impact and acceptability in single-point as well as in serial-monitoring testing for implementing a device(s) for STAT laboratory services.

Performance specifications of analytical tests are a set of quantitative and experimentally derived characteristics consisting of:

- Technical details
- Imprecision
- Inaccuracy
- Specificity
- Detection limits
- Turnaround time
- Laboratory safety
- Cost

Describing defined objective quality performance characteristics requires a thorough understanding of the overall analytical process and the methodologies involved.

Clinicians' ability to initiate efficient and correct treatment is highly dependent on this analytical understanding.

Thus, there is a close relationship between the required quality of the laboratory service provided and the ability to use the data in the clinical surroundings.

The optimal approach of delivering blood gas services within a hospital secures thorough knowledge of the above factors and optimal control to maintain the quality of that service.

However, even that approach has to be secured by some

kind of instrument-specific peer-group surveillance in order to maintain analytical performance over time.

Very often, hospitals seem to use not one brand of device but two or more different brands of devices.

This non-conformity approach may cause unknown difficulties with regard to the clinical ability to make specific decisions, as there might be very large differences in analytical performance characteristics when measuring samples on instruments from different manufacturers.

Even from the same manufacturer, analytical results may have inherent biases and different precision profiles from instrument to instrument.

If more than one device (same/different manufacturer) are used to obtain the STAT service, analytical goals for allowable imprecision (random errors) and inaccuracy (systematic errors) using such an approach must be properly understood.

Furthermore, when those analytical goals are established, laboratory QC systems and procedures maintaining those goals must be implemented in order to secure short- and long-term analytical consistency.

If one does not follow and define analytical goals for acceptability and implement QC systems to maintain those goals, the use of different analytical devices could very well lead to random clinical judgment and decisions in the clinical setting.

Regarding long-term analytical reliability and stability of a given device, one also has to set up rules for judging acceptability.

Those would be determined by frequency of outliers, failure rates, and analytical errors inherent in the device that give misleading clinical directions for evaluating the analytical results.

In 1987 a committee under the American Association of Clinical Chemistry (13) published Guidelines for Providing Quality STAT Laboratory Services.

This document gives excellent analytical goals for acceptable imprecision and bias for many important STAT analytes and describes some organizational aspects of providing STAT laboratory services.

However, it seems to need an update and some clarification, as well as new guidance regarding quality assurance planning, in order to address issues regarding new technology of point-of-care devices.

Some POC devices available on the market might have difficulty passing some of the requirements of this document.

Criteria for acceptability would be more informative in this document if given as CVs (coefficient of variations). Acceptable standard deviation (0.04 mmol/L) for analytical imprecision of ionized calcium seems too big.

On the other hand, acceptable bias seems too small (0.02 mmol/L). Statistical requirements that set limits for both the imprecision and bias that are tolerable in a single measurement or single test result should also be defined in this document.

Total allowable error for a given method should be expressed as:

$$TE_A = \text{numeric value of bias} + z \times CV_A$$

where TE_A is the total allowable analytical error, bias is the inaccuracy of the method towards a well-defined reference method or reference material (if that is available), z is the statistical confidence factor (1.65 \approx 90 %, 1.96 \approx 95 %, and 2.33 \approx 99 %), CV_A is the imprecision of the method (including day-to-day variability and instrument-to-instrument variability) given as a coefficient of variation.

Thus, in routine multi-profile blood gas testing most established analytical quality requirement strategies seem to have some limitations.

The major issue being that most variabilities in multi-profile blood gas testing of whole blood probably do not

relate to difficulties in the analytical process but rather lie in difficulties in controlling preanalytical factors such as sample handling, issues regarding metabolism, as well as change of analytical specimens during processing.

It is very important for the users to be able to separate preanalytical variability from analytical variability in order to focus and optimize the total process of the blood gas service they deliver.

Although it might be difficult to define exact analytical goals in multi-profile blood gas testing that all laboratorians might stick to, it is, however, important to define clear local criteria for what acceptable imprecision and bias mean when setting up your blood gas service for your hospital.

That approach, though not universal, would at least create consistency through the overall service no matter what device is included in the service.

Too many published scientific papers on methodological multi-profile blood gas testing in different highly estimated analytical and clinical journals make very weak definitions of analytical acceptability for imprecision and bias.

With more and more new point-of-care (POC) devices coming to the market, analytical acceptability criteria seem to have had a tendency to become wider and wider.

Either traditional blood gas analyzers seems to have been too good in the past or most point-of-care devices are not acceptable.

Clinically, my point here is simply that analytical acceptability should be the same for traditional as well as for POC devices.

Determine analytical performance characteristics of the measurement

Variability of measurements is a well-known phenomenon in analytical work. The random variation

observed by replicate measurements of the same sample is very often assumed to be Gaussian and expressed as imprecision of the device.

The imprecision can be expressed by the coefficient of variation, ($CV_A = SD_A/\text{mean}$) or in percentage terms ($100 \times (SD_A/\text{mean})$).

A constant mean difference between two methods in a service is called a bias or inaccuracy of the methods.

Thus, analytical performance characteristics of all devices can be described by two important analytical statistics:

- Imprecision
- Inaccuracy

For blood gas instruments, studies in the literature very often describe imprecision and inaccuracy from comparison of two different instruments and use of statistical linear regression analysis.

However, to objectively characterize the quality of an overall blood gas service provided by a given laboratory seems to demand more than a line and correlation coefficient.

To adapt the present total concept all manufacturers should specify errors of measurements as best- and worst-case analytical scenarios for the operation of at least 5-10 analyzers.

Furthermore, such an approach would demand analytical specifications for within-run analytical variation, between-days analytical variation, and between-instrument variation, as well as inherent inaccuracy of the device in comparison with a reference method.

Because many manufacturers only specify their blood gas analyzers according to a best-case scenario, such an approach seems difficult to apply directly.

In blood gas measurements (pO_2 and pCO_2), inaccuracy can only be established if the end user has access to a tonometer and masters that technique.

Inaccuracy of most other analytes of the multi-profile

can, however, be established from commercially available standards, e.g., NIST SRM 965a (USA), or pH buffers with NIST-assigned values.

Though, locally, the absolute inaccuracy of instruments may not be as important as biases between instruments from different manufacturers that are actually in use in a specific laboratory service.

Far too often 'the line of stupidity' (regression analysis) is used to prove analytical acceptability, comparing the performance characteristics of two instruments of the laboratory service.

Though a Bland Altman difference plot (Instrument (A-B) vs. mean readings of Instrument A and Instrument B) gives far more information than focusing on getting a line that is actually an inherent statistical assumption of this analysis.

Many laboratories seem today to be investing far too little time in doing a thorough method comparison and technical study of the devices that they implement into their laboratory services.

However, very often such an insufficient approach runs into serious methodological problems down the road that laboratories might have a hard time resolving in the middle of delivering clinically requested results on the same devices.

Poor methodological studies and selection of poor products that cannot meet analytical and clinical requirements and specifications may be a very bad investment in the long run.

Identify candidate statistic quality control strategies

The purpose of quality control of general chemistry tests are to validate the analyzer's performance by evaluating inaccuracy and imprecision.

Theoretically, the user should determine the proper location of control samples within a run, keeping in

mind the principle that quality control results should aid the technologist with regard to acceptability of reporting patient results to the ward.

To establish such a quality assurance program, a control material must be found that has as many as possible of the following characteristics:

- Matrix similar to whole blood
- Available in quantity
- Stability for a long period (one year)
- Open vial stability during period of use
- Immediate availability in an emergency
- Low vial-to-vial variability
- Available in concentrations spanning the medically significant range

No such product is available for multi-profile blood gas analysis. Also, the parameters measured are so labile and subject to room-air contamination that it is difficult to achieve "open vial" stability.

Performing quality control routinely will help ensure that results from actual patient samples are accurate. A well-developed quality control system evaluates various errors comparing results measured on the control solutions with their predetermined values.

Quality control measurements must be performed daily in order to serve their purpose. Calibration measurements can never take the place of quality control in assuring predetermined analytical requirements.

Ideally, all control solutions of a quality control system should be performed at the beginning of each shift that patient samples are measured. The level of quality control materials should be selected where important clinical decisions are to be made.

High, normal, and low levels are the primary choice of most multi-profile quantities. A quality control strategy is defined by what control materials are used, how many control samples are analyzed, where these control samples are located, what quality control rules are applied to the control sample measurements, etc.

Well-functioning Westgard rules for multi-profile blood gas analysis are, for example, 1_{2s} , 1_{3s} , $(2 \text{ or } 3)_{2s}$, 2_{2s} , 6_{1s} and 10_x .

Individual and more optimal rules might be selected for each parameter using the OPSpecs program by Westgard [14], optimizing and lowering the theoretical probability of false rejection (P_{fr}) as well as the theoretical probability for error detection (P_{ed}).

The appropriateness of an optimal QC strategy seems very much to depend on the quality required, as well as the expected instability of the analytical method (e.g., type, magnitude, and frequency of errors).

It is also important to define the practical run length and stability of the multi-profile blood gas technology chosen for the service in order to estimate how many controls have to be run on the systems included in the service.

The state-of-the-art multi-profile blood gas analyzers on the market have generally a maximum stability and run length of eight hours.

Because of a known run stability of a given state-of-the-art multi-profile blood gas system, the following simple quality control system is recommended as a minimum for such a system:

- One level is tested at the beginning of every shift
- All levels are tested over the course of a day
- Additional quality control should be performed after any troubleshooting or preventive maintenance, which might alter performance
- In each shift duplicate analysis on different instruments should be performed on one or two high-quality samples of whole blood

The approach of analyzing patient samples in duplicate provides a way of controlling estimating performance through the overall multi-profile blood gas service provided to the clinicians.

It permits early detection of clots, electrode drift, and faulty calibrations.

In all quality assurance programs for blood gas analytical services there must also be well-defined criteria for judging in-control and out-of-control signals of the control system, as well as what actions to undertake when the system is out of control.

Thus, in statistical quality control the stable quality control materials are measured many times over time, and the observed results compared with limits that describe the variation expected, when the measurement method is working properly.

A control chart is prepared to display the results on the y-axis vs. time on the x-axis. From these charts in-control and out-of-control results may easily be judged. This is also the minimum CLIA requirement in the USA for this kind of testing.

In Europe and Japan, this is far from the actual implemented QA program, and one can wonder what they really control in these parts of the world with reference to multi-profile blood gas testing.

In Europe, laboratories seem to have very strong theoretical biologically derived analytical requirements regarding imprecision and bias for blood gas measurements based on biologically derived variations, but when it comes to assuring that these requirements are fulfilled on the floor when implementing a practical quality system that actually maintains or fulfills these analytical goals, there seems to be some gap between theory and practice.

In that perspective real quality and quality assurance in my opinion seem very much created on the laboratory floor and not through the implementation of some sophisticated theoretical QA system.

Conclusion

Clinical chemists and clinicians making choices of devices for their analytical multi-profile blood gas services must be familiar with the basic methodological concepts on how to depict analytical impact on clinical decisions of their choice of device(s) and how to set up a QC system

that secures and maintains the quality of their analytical approach and choices.

Current trends towards total quality assurance in laboratory medicine suggest that objective quality goals must be consistent and very well defined.

A step-by-step quality planning approach using CLIA requirements for maximum allowable imprecision and bias and OPSpecs put forward by Dr Westgard seems the most optimal approach for setting up a practical QA program maintaining a high-quality multi-profile blood gas laboratory service.

Before trying to set up a QA system, however, a thorough methodological study and evaluation in respect to the clinical outcome using the laboratory service must be understood in great detail.

Unless patients are measured during critical illness by the same peer group of instruments providing this service, the ability to detect small differences in consecutive samples from the same patient seems limited.

It is of utmost importance not to mix preanalytical variability and analytical variability when optimizing the multi-profile blood gas service. In that perspective, analytical variability is independent of preanalytical variability.

The best laboratory blood gas service scenario seems to be a choice of technically aligned devices with little imprecision and a minimum bias or no bias between the devices.

Training verification of laboratory personnel on both preanalytical and analytical aspects, as well as regular proficiency and skills evaluations are very important in providing an adequate multi-profile blood gas service.

Such a laboratory service approach does not only lead to analytical coherency but also makes patient therapy, monitoring, and clinically important decision-making in critically ill patients consistent throughout the whole hospital.

References

1. Westgard JG, Barry PL. Cost-effective quality control: Managing the quality and productivity of analytical processes. Washington DC: AACC Press, 1986.
2. Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. Clin Chem 1990; 36: 1629-32.
3. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974; 20: 825-33.
4. Fraser CG, Hyltoft Petersen P, Lytken Larsen M. Setting analytical goals for random analytical error in specific clinical monitoring situations. Clin Chem 1990; 36: 1625-28.
5. Fraser CG. Analytical goals are applicable to all. JIFCC 1990; 2: 84-86.
6. Fraser CG. Goals for clinical biochemistry analytical imprecision: A graphic approach. Pathology 1980; 12: 209-18.
7. Cembrowski GS, Carey RN. Laboratory quality management QC/QA. ASCP Press, 1989.
8. Hyltoft Petersen P, Ricos C, Stöckl D, Libeer JC, Baadenhuijsen H, Fraser C, Thienpont L. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. Eur J Clin Chem 1991; 37: 1938-44.
9. Tonks DB. A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. Clin Chem 1963; 9: 217-23.
10. Elevitch FR, ed. Analytical goals in clinical chemistry. Proc, CAP Aspen Conference, 1976. Skokie, IL: College of American Pathologists, 1977.
11. Elion-Gerritzen WE. Analytic precision in clinic chemistry and medical decisions. Am J Clin Pathol 1980; 73: 183-95.
12. Statistical quality for quantitative measurements: principles and definitions: approved guideline - Second edition. NCCLS document C24-A2. Villanova, Pa: NCCLS, 1995; 19, 5.
13. Guidelines for providing quality STAT laboratory services. The Laboratory Quality Assurance Committee of American Association for Clinical Chemistry. Washington, DC: AACC Press, 1987.
14. Westgard JO. Planning and validating QC procedures. Workshop manual 1994. Westgard® Quality Corporation.