Standardization of laboratory equipment

P. M. G. BROUGHTON

Department of Chemical Pathology, University of Leeds, England

The majority of errors which arise in clinical chemistry can be traced to four main causes:

- (1) The analytical method chosen.
- (2) The equipment used.
- (3) Impurities and instability of chemicals and reagents.
- (4) Human errors such as poor technique.

Increasing attention is now being given to the selection of analytical methods with a sufficiently high standard of accuracy, precision and reliability for routine use. A similar critical appraisal of equipment is equally necessary, but basic facts and principles need to be established before any attempt at standardization can be made. This paper is concerned with the errors which can arise from equipment and the principles for selection of instruments which are suitable for use in clinical chemistry. Some examples are also given of the types of error which can result from equipment malfunction, and the application of quality control methods for detecting these.

Analytical methods and equipment

The outstanding success of the AutoAnalyzer (*) has been largely due to the skill with which analytical methods have been modified to fit the capabilities of the instrument. However, some instruments and analytical systems impose a limitation on the methods which can be used with them. For example, the SMA 12/30 (*) (now withdrawn) contained no provision for a sample blank, so that errors due to turbidity could occur in the methods used for the determination of bilirubin, total protein and albumin (¹). Similarly, some discrete analytical systems contain no facilities for the removal

^(*) Technicon Instruments Company.

of protein from the solution analysed. Consequently, the use of such equipment is restricted to those tests for which there is a suitable method. With most automated systems the availability of adequate methods is of major importance in the selection of equipment. In many cases, therefore, the equipment must be assessed with the methods provided by, or recommended by, the manufacturer. This may make it difficult to distinguish between the inadequacies of the equipment and the method. Some general principles, and the types of error likely to be found, can be established by first considering individual instruments.

Inherent errors of equipment

The commonest operation in any analytical procedure is the use of a pipette to deliver a known amount of reagent. Which type of instrument is best? Broughton *et al.* (*) tested 35 commercially available dispensing pipettes and compared their accuracy, precision, ease and speed of use. The majority were not accurate unless calibrated by the operator and, with some, reliance on the manufacturers' calibration could result in significant errors. The precision of repetitive deliveries made with conventional bulb pipettes was entirely dependent on the operator's technique and with many so-called « automatic » pipettes the technique used also influenced precision. Only with mechanically operated instruments was a consistent precision obtained, independent of technique. From the results found with different instruments the authors suggested that an acceptable standard of accuracy for an automatic pipette was 1 %, with a precision (coefficient of variation) of 0.1 %. Several instruments were capable of this performance, but price was usually a poor guide in the selection.

The majority of analytical methods used in clinical chemistry depend on the use of colorimeters. Two comparative assessments, involving nine commercially available instruments, have been described (^{3,4}). The stability, sensitivity, linearity, versatility, precision and ease of use of each instrument were examined by a series of specially devised tests. Many instruments were unstable so that the reading changed or drifted with time, resulting in a gradually changing accuracy or poor precision. Some colorimetric procedures gave non-linear calibration curves with some instruments, but linear calibration with others, so that inaccurate results would be obtained if Beer's Law were assumed to be obeyed. On some occasions, two models of the same instrument gave different readings. The precision of making repetitive readings on the same solution depends on the absorbance and hence the concentration of the solution. This varied widely among the instruments tested and similar effects can be shown with more expensive equipment such as spectrophotometers and colorimeters with print-out or

Ann. Ist. Super. Sanità (1971) 7, 257-264

2

recorder attachments (Fig. 1). This change of precision with concentration is particularly important when evaluating or comparing the precision of either equipment or methods. Precision data should always include a statement of the concentration (5).

Broughton and Dawson (⁶) have recently made a similar comparison of five commercial flame photometers capable of measuring sodium and potassium simultaneously in the same diluted sample. Their speed of analysis,



B (open circles) - colorimeter with recorder output

C (crosses) - colorimeter with print-out

stability, sensitivity, linearity, precision, ease of use and safety were examined. The readings obtained with some instruments drifted with time, due to either electronic or flame instability. All used precalibrated scales which were set with two standards, but in some instruments the response was found to be non-linear.

These evaluation studies have shown that the most expensive instruments are not necessarily the best for a specific purpose, and in many cases the manufacturers' claims for their equipment could not be substantiated when the instrument was tested in the laboratory. Objective tests of performance are a better guide to the purchaser of equipment, but these require that the factors which are of critical importance in an instrument first be defined. Methods for testing these factors can then be devised and the results provide a basis for selecting the best instrument.

Automatic systems

The introduction of work simplification procedures, involving simple semi-automatic equipment, and automation undoubtedly results in better precision of routine analyses (7). Numerous surveys [e.g. (8)] have shown that the performance of laboratories using the AutoAnalyzer is on the whole better than those using manual methods. This is at least partly due to the greater dependence of manual methods on the skill and technique of the operator. The precision of manual methods can deteriorate from overwork, tiredness and boredom whereas the precision of AutoAnalyzer procedures is less affected by these factors (9).

Even with automatic equipment human errors can be important, as in the 'rounding off' of results read from AutoAnalyzer charts, where some operators may show a preference for even numbers (¹⁰). The increasing number of specimens analysed also increases the risk of errors from faulty sample identification, which at present is largely done by the operator. Human intervention must be regarded as an unpredictable source of error and therefore best reduced to a minimum by further automation, including automatic sample identification.

An automatic or semi automatic system contains a number of instruments or modules, each of which has an inherent error contributing to the total error of the analysis. If the total error is large, it may be useful to examine each module or stage in the analysis in order to identify the dominant error. With many automatic systems, interaction between successive samples (« carryover ») may be large and give rise to significant errors in the complete analysis. The amount of interaction depends on the design of the instrument but may be influenced by the speed at which it is used, the use of water wash solutions between samples and the cleanliness and other characteristics of surfaces in contact with the solutions. Minor modifications to the equipment or in the technique used can often reduce excessive sample interaction and, if necessary, a correction factor can be applied to results affected by it.

In Britain, a schedule of tests has been recommended (¹¹) for the evaluation of new instruments for automatic analysis in clinical biochemistry. This describes in detail the tests which should be carried out to provide a complete and independent assessment of such equipment. Similar tests using these principles can be applied to other types of equipment, such as instruments used in haematology (¹²). Each type of instrument or analytical system will have its own individual source of error, and when used in quantitative analysis will result in either poor accuracy or poor precision. Excessive sample interaction, for example, will give poor precision, whereas a non-linear response of an instrument may lead to inaccurate results at concentrations between the standards.

Instrumental requirements

When selecting an instrument to perform a specific task the user should consider the following basic factors:

Accuracy and precision. These can both be measured by suitable tests, and a list can be made of the minimum standard attainable by all instruments. The less satisfactory instruments can then be rejected, but unless there is an objective specification, it is impossible to decide whether the best instruments are good enough. New instruments are invariably judged in comparison with the performance of existing ones, instead of using criteria based on the fundamental requirements of clinical chemistry.

Speed. The time required for the complete analysis of different numbers of specimens is important, not only in the organisation of the laboratory but in determining the cost of labour. The assessment should therefore include the operator's time in setting up the instrument, standardising and adjusting, as well as time taken in calculations and maintenance.

Sensitivity. This determines the volume of specimen required for the analysis. In many laboratories it is preferable that the same procedure be used for both adult and paediatric patients.

Future needs. When purchasing an instrument, the user should reasonably expect it to fulfill his needs, for, say, 5 years. It may also be useful to consider whether the instrument can be extended by the addition of modules such as an automatic sampler and chart recorder.

Cost. This should include both capital and running costs. Broughton and Dawson (⁶) amortised the capital cost of flame photometers over 5 years and calculated the capital cost per day. The labour costs were calculated from the time required for the analysis of different numbers of specimens per day. With flame photometers reagent costs were negligible. They found that with large numbers of analyses, the more expensive instruments were cheaper to run as they included some automation, whereas with small batches the total cost was less with cheaper instruments. However, cheaper flame photometers usually had the less satisfactory performance.

The user must therefore decide what standard of performance he requires from an instrument, and then examine the cheapest method of achieving this with his anticipated work load. The same principles apply with both simple equipment and complex analytical systems. With multichannel analysers the first requirement is to define the tests which it is necessary to perform and the cost-effectiveness of different machines can then be assessed.

Ann. Ist. Super. Sanità (1971) 7, 257-264

Instrumental malfunction

When an instrument has been installed, faults may arise during its use, leading to analytical errors. The cause of the error may be in the environment (e. g. sunlight, cigarette smoke, vibration or mains voltage variation) or due to a fault, sometimes temporary, within the equipment. Whitehead and Morris (¹⁰) found that 18 % of the errors detected by quality control were due to equipment. The type of fault will depend on the instrument and Fig. 2 shows an example which developed when using the AutoAnalyzer. Two independent quality control methods indicated that low potassium results were being obtained, but several days were necessary to identify



Fig. 2. — A quality control problem with plasma potassium: The upper graph (closed circles) shows the daily result for a control serum. The lower graph (open circles) shows the mean values of all patients' plasmas analysed each day (excluding those over 6.5 mmol/1). The dotted lines indicate the confidence limits (\pm 2 standard deviations) established from previous analyses. During the period indicated (black shaft) a different (and presumably faulty) batch of dialyser membranes was in use with the AutoAnalyzer.

and cure the fault, and in this time many low results were being reported. Most quality control procedures will detect large errors immediately, but consistent small ones may be overlooked. Improved methods of quality control are therefore still needed and these could include alarm signals, built into equipment, which would immediately signal an error or fault. It is often useful to compile a list of « trouble shooting» tests which can be instituted to identify and cure the fault immediately it is detected.

BROUGHTON

Conclusions

The clinical chemist can only achieve the standard of accuracy and precision that his instruments allow. At present, this is very variable due partly to the inadequacies of equipment and partly to dependence on the operator's technique. The shortage of skilled laboratory staff and increasing work loads in clinical chemistry will lead to a further extension of automation in the future. However, there is a natural tendency to believe that the machine must give the right results, particularly if it is automatic and gives a printed output. Quality control is therefore essential with all types of equipment and new procedures may be needed for use with automatic instruments operating at high speeds. Poor equipment, and good equipment badly used, can be a major source of error and a more critical approach to the selection of instruments is advocated. This requires that the clinical chemist defines his requirements in exact terms, particularly in relation to performance, where the primary factors are accuracy, precision, speed and sensitivity. An analysis of both capital and running costs can then provide the purchaser with sufficient information to select an instrument with the performance he requires at a price he can afford.

Summary. — In clinical chemistry, the overall standard of accuracy and precision depends not only on the choice of an appropriate analytical procedure but also on the adequacy of analytical equipment and on the individual operator's technique. Poor equipment, and good equipment badly used, can ben a major source of error.

When automatic equipment is used, there is a tendency to believe that the machine must give the right results, especially if it gives printed data: this can easily lead to error. A critical approach to the selection of instruments, and especially of automatic equipment, is therefore required.

New quality control procedures, including alarm systems built into analytical equipment, are needed for the timely detection of errors, especially when high speed automatic instruments are used. The clinical chemist must define his requirements in exact terms, particularly in relation to overall performance, which includes such factors as accuracy, precision, sensitivity and speed.

Riassunto (Standardizzazione delle attrezzature del laboratorio). — In chimica clinica lo standard globale di accuratezza e di precisione dipende non solo dalle scelte di un adatto procedimento analitico, ma anche dalle adeguatezze delle attrezzature analitiche e dalla preparazione tecnica dei singoli operatori. Attrezzature scadenti, o ottime attrezzature malamente usate possono costituire una importante fonte di errore. STANDARDIZZAZIONE E CONTROLLO DI QUALITÀ

Quando vengono usati apparecchi automatici, c'è la tendenza a credere che la macchina debba dare risultati corretti, specialmente se fornisce dati stampati : questo può facilmente portare ad errore. Si richiede pertanto di affrontare criticamente la scelta degli strumenti, in particolar modo di quelli automatici.

Sono necessari, per rivelare in tempo gli errori, nuovi dispositivi per il controllo di qualità, quali un sistema di allarme incorporato nella apparecchiatura analitica, specialmente quando si usano strumenti automatici ad alte velocità. Il chimico clinico deve definire in termini esatti le sue esigenze, particolarmente in relazione alle caratteristiche generali dell'analisi, il che comprende fattori quali l'accuratezza, la precisione, la sensibilità e la velocità.

REFERENCES

- (¹) BROUGHTON, P. M. G., D. SIMPSON, F. L. MITCHELL, C. TOOTHILL & L. G. WHITBY. Clin. Chim. Acta., 19, 297 (1968).
- (2) BROUGHTON, P. M. G., A. H. GOWENLOCK, G. M. WIDDOWSON & K. A. AHLQUIST. Automatic dispensing pipettes: An assessment of 35 commercial instruments. Scientific Report No. 3, Association of Clinical Biochemists, London, 1967.
- (3) BROUGHTON, P. M. G. & C. RILEY. Colorimeters with Flow Through Cells: A critical assessment of four instruments. Scientific Report No. 1, Association of Clinical Biochemists, London, 1965.
- (4) BROUCHTON, P. M. G., C. RILEY, J. G. H. COOK, P. G. SANDERS & H. BRAUNSBERG. Colorimeters: A critical assessment of five commercial instruments. Scientific Report No. 2, Association of Clinical Biochemists, London, 1966.
- (5) BROUGHTON, P. M. G. Ann. Clin. Biochem., 6, 109 (1969).
- (6) BROUGHTON, P. M. G. & J. B. DAWSON. An evaluation of five commercial flame photometers. Scientific Report No. 4, Association of Clinical Biochemists, London, 1970.
- (7) WHITBY, L. G., F. L. MITCHELL & D. W. Moss. Advan. Clin. Chem., 10, 65 (1967).
- (8) GOWENLOCK, A. H. Ann. Clin. Biochem. 6, 126 (1969).
- (*) ROBINSON, R. Lancet, ii, 744 (1966).
- (10) WHITEHEAD, T. P. & L. O. MORRIS. Ann. Clin. Biochem., 6, 94 (1969).
- (¹¹) BROUGHTON, P. M. G., M. A. BUTTOLPH, A. H. GOWENLOCK, D. W. NEILL & R. G. SKENTELBERY. J. Clin. Pathol., 22, 278 (1969).
- (12) LAPPIN, T. R. J., A. LAMONT & M. G. NELSON. J. Clin. Pathol., 22, 11 (1969).

Ann. Ist. Super. Sanità (1971) 7, 257-264

The implementation of quality control and factors affecting its success

F. L. MITCHELL and I. J. L. GOLDBERG

The Medical Research Council, Clinical Research Centre, Harrow and St. Mary's Hospital, London, England

During the course of the Congress, many of the techniques which are now available to us for the control of quality in our laboratories will be considered. We believe it is of equal, if not of greater importance, also to consider problems encountered during the implementation of these techniques and the factors which affect the attainment and maintenance of high quality.

Firstly, we will deal with the implementation of quality control, and before doing this it is most important to understand fully what it will do for us and what it will not do.

As it is used at present all quality control does, is to tell us how well we can reproduce what could be bad in accurate results.

It is a statistical system for measuring only reproducibility or precision. It does not measure the true accuracy of a determination, this is something which is inherent in the technique used. So, before we think of introducing quality control we must check the performance of our methodology for accuracy. There are well known ways of doing this (1), but the best system is not yet available and we believe it is most important that we make it available as soon as possible. It is to have for every assay we do, a control serum or urine which has been assayed by a technique, internationally agreed, if possible, perhaps long and laborious, perhaps entailing the use of extremely expensive equipment such as mass spectrometers or x-ray fluorescence machines, but one which will give a result as near to perfection as is humanly possible. Controls assayed by such techniques are not yet available but we should like to suggest that we make them available and call them referendary controls. We choose the word referendary because according to the Oxford dictionary it defines an arbitrator to whom a dispute is referred for a decision. The need for such controls was particularly evident in November 1969 and in 1968 when representatives of forty European and American Jaboratories met in Geneva and we each took along the results we had obtained on the same standard serum. Practically all the results were different and up to ten different methods had been used for each assay. Who was to say which was correct? The need for some form of enlightened arbitration was most apparent.

Now when we come to instituting a programme of quality control it is all too easy to organise a system which is complex and efficient but, unless everything is under computer control, it is quite a different matter to keep that scheme going year in year out. After the first results have come in and one has found how good or bad one is, it is universal experience that interest flags and there is a great tendency to slip back to that utopia where ignorance is bliss, and it is folly to be wise.

There are several ways of preventing this.

Firstly, a scheme should not be started which is too elaborate. It is better to have a simple slightly imperfect manageable scheme, rather than a complex, perfect system which takes half the laboratory staff to run. We think it is fair to say that unless computer assistance is available, it is not normally practical to run many of the elaborate types of system which are available. It is possible to choose a very simple or a very complicated programme depending upon the circumstances of the laboratory, but choose some sort of programme every responsible laboratory head must. One hears remarks such as «We are so busy, it is absolutely impossible to take on any extra work for quality control». This situation should never obtain under any circumstances. Even if work has to be refused by a laboratory, a quality control programme must be instituted. We must never forget that the results we produce, if they are wrong, can give a great deal of unnecessary suffering and can be lethal.

Secondly, one person should be designated as quality control officer, and in a large laboratory this can be almost a full-time job. That person then often becomes dedicated and he or she will make it their personal responsibility that the programme is strictly adhered to. He or she can of course delegate responsibility but they must be responsible for the stopping of results coming from a method which has gone out of control.

Thirdly, issue the wards at regular intervals with a list of standard deviations giving the laboratory error for each technique. Clinicians rapidly begin to find this of immense value and if it ceases to appear regularly or if they have cause to suspect it, they will soon want to know the reason why.

Fourthly, do not hide the quality control charts in a corner, they should be available for all to see, including visitors to the laboratory, particularly the clinicians. It gives them a degree of confidence in the laboratory which is difficult to achieve in any other way. It is a good idea for the laboratory

266

director to have up-to-date copies of quality control graphs on wall charts in his office. It is most important that he should follow them from day to day and have them easily available for reference.

Lastly, nothing succeeds like success. The importance of a quality control programme is very quickly realised by the laboratory staff. Any improvements instituted in the laboratory working are rapidly reflected in the quality control results, and this leads to an increased pride in the work and an element of competition, particularly if inter-laboratory controls are included. Once this atmosphere has been created, the staff themselves will see to it that the programme does not lapse.

We do not think many of us need reminding of the quotation Lord Kelvin made nearly one hundred years ago, « When you can measure what you are speaking about, and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind, it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science».

Until recently, quality control in clinical chemistry was almost unknown. We had not troubled to develop means of measuring it and we took it all for granted. Now, thanks to our friends the statisticians, we have yardsticks for measuring our quality, we can express it in numbers and we can now therefore investigate the factors which affect it. Because of this and also because of the concurrent developments of automation, we have undoubtedly seen the dawn of a new era in our specialty.

It is most important that all the laboratory staff should know the influencing factors, and then something can be done about them.

We should like to spend the rest of the time available dealing with them in detail as they affect all laboratories: first and foremost, laboratory management. Responsibility for the efficiency of a laboratory lies fairly and squarely in the hands of the director. If his quality control results are bad, he should worry until they are improved, if he has no quality control, he should be even more worried. If, as is all too often the case, the quality of his output is low because he is overloaded with work and cannot, for various reasons, increase his staff, he has the choice of alternatives, either increasing efficiency, or leaving out the less important items of his load and concentrating upon the important work.

The wisdom of this is shown in Table 1. Quality control was introduced in a laboratory known to be under considerable strain, and most alarming situations immediately became evident. Calcium results with a laboratory standard deviation of 1.4 are not only useless, they are dangerously misleading. It is better to stop doing them at all. The introduction of the less laborious Trinder technique more than halved the error, automation reduced

TABLE 1

Quality control results for the assay of serum calcium

Results from Mitchell et al. (2)

												Standard deviation mg/100 ml (Reference value 10.0 mg/100 ml)
Laboratory staff under strain	i											
Clark and Collip method (3)		1	ų	1	•	2	Ā					1.4
Trinder method (4)	•	•		•	•)	ł	÷	•	1940	*	•	0.6
Automated, first month .				٠	•					•		0.4
Automated, second month	191				a n e		*	1	•	2	×	0.23

it again and as we all know, we can now, with automatic methodology, obtain a standard deviation approaching 0.1 mg/100ml. This is all largely a result of being able to express quality in numbers.

Under the heading of « management» we should like to draw attention to two very important factors which, between them, can make the most dramatic effect on quality control results. These are: staff training and automation. Their effects on the quality of results are shown in Table 2.

In the results for the manual assay of glutamic-oxalacetate transaminase (GOT) as might be expected, the senior technician produced a better

TABLE 2

Reproducibility of results from the assay of serum GOT Results from Mitchell et al. (2)

	COEFFICIENT	S OF VARIATION	VARIANCE RATIOS					
TECHNICIAN	Manual techniques (M)	Work simplified techniques (WS)	(M/WS)	Senior/student (M)	Senior/student (WS)			
Senior Student	2.00 4.01	2.15 1.96	1.12 4.12*	4.(12*	1.16			

• Variance ratio significant at the 5% level.

coefficient of variation than the student but with the institution of work simplification, the student halved his coefficient of variation, but the senior technician did not improve further.

The necessity for a happy atmosphere in a laboratory cannot be stressed too highly. No worker, junior or senior, can do good work if his mind is on things other than the job immediately in hand. He cannot concentrate. A good laboratory head should, we believe, always be on the lookout for undercurrents and trouble. Fatherly advice or sympathy in many domestic difficulties can work wonders far out of proportion to the effort involved. Respect counts for so much more than strict discipline. Courses in management invariably start by pointing out that a soldier will die for an officer he respects but not for one he despises, and this applies equally well to loyalty and efficiency in a laboratory.

Before we leave management, could we consider five important points whereby we can affect the one attribute of an employee that always remains his own, that is, his attitude or his desire (⁵).

Firstly:

1. Stress the importance of perfection in the task.

2. Make it obvious that nothing less than perfection is acceptable.

3. Let the required standards be known.

4. Keep a check on the quality of every individual's work and see that they know how they stand.

Finally:

5. Give recognition for good work, and let it be known what is being recognised. We so often criticise and scold but do we equally often give praise when praise is due.

If good management is the foundation stone of good quality, the physical layout and organisation of the laboratory is the next building block. With modern materials, our laboratories should be approaching the ultimate in the efficiency of their layout. It is sometimes said that research workers perform best under slightly crowded conditions, but time and motion studies have shown that the staff in service clinical chemistry laboratories give of their best when working as comfortably as possible under uncramped conditions with the apparatus they require near at hand. The general requirement, layout and appearance of service and research laboratories must of necessity be different.

Dirty glassware is one of the major reasons for poor quality control results. Detergents are difficult to remove and can have very damaging effects. When the manufacturer of a detergent specifies one measure of detergent for a gallon of water, the user often feels that the job will be done twice as well with two. If glassware must be used, washing is generally best done by machine or by a technician washing his or her own.

Wherever possible, however, it is advisable to use disposable containers. A change to disposable tubes and pipettes will usually make a most marked improvement in the results of quality control.

It is important to ensure that instruments are always working at peak efficiency, and the only way to do this is to have a system of regular inspection and maintenance. The wave-length calibration of a spectrophotometer can slowly and imperceptably drift, or slackness can develop in the cam drive of an automatic diluter, dramatically affecting its precision. Broken glass in a centrifuge bucket or worn rubber cushions, can cause repeated breakages of sample tubes. It is remarkable how often this can occur before the reason for the trouble is realised.

The calibration of glassware cannot be taken for granted, the quality control of a manufacturer can slip just as surely as that of a clinical chemist. All of us, for instance, have had experience of occasional wildly inaccurate rogue pipettes.

The working life and making up of reagents must be closely watched and also their handling; for instance, peroxidase used for glucose assay by the glucose oxidase method can be rendered almost completely inactive by repeated removal from, and returning to, a deep freeze. This can seriously affect a glucose assay method and the reason for the trouble is not readily apparent.

Personal comfort is of very great importance, and all five senses need to be considered. Bright sunlight or bad lighting can have deleterious effects on staff, consciously and unconsciously, and also on instruments. Direct sunlight should generally be avoided and usually good artificial light is more satisfactory than natural light. Don't shine bright lights on optical instruments. Excessive noise can be very wearing on all grades of staff.

Posture and easy access to frequently used articles is important; the human race is by no means uniform in size and this should always be taken into account in the fixing of shelves and the purchasing of stools, etc.

With regard to temperature, fortunately, there are not many of us without central heating in our laboratories, but it is necessary to pay attention to more than the setting of the thermostat. In an experiment on factory production some time ago, it was shown that while keeping temperature and humidity constant, the changing of wall colours to a light blue caused production to drop considerably, since the majority of workers felt cold and even stayed away from work. Changing the wall colours back to a rosy tan brought the level of output above that recorded before the experiment started. Draughts should be avoided as they can have an important influence on the reproducibility of results from instruments, particularly AutoAnalysers.

Lastly, the sense of smell: it is not a big factor but ladies' perfume can be most disturbing if it is, as they say, out of taste, which leads one to a factor which can be important. Having attractive girls around can liven up the atmosphere and increase efficiency, but when the attraction is accentuated so that the poor boys can think of nothing else, a time comes when a line must be drawn.

270

A rotation of technicians is important to relieve the monotony of doing the same task for long periods. There is no hard and fast rule for this, since training has to be fitted in, and what is boring to one technician, is interesting to another.

Perhaps another important quality-conscious decision a laboratory can take is to employ and train university graduates rather than technicians. A small number of technicians will probably always be needed, but sophisticated modern equipment and techniques call for skilled analysts in many instances and cannot be competently handled by technicians trained in the old ways of rotation through the various pathology disciplines.

If one does decide to recruit graduate analysts then it is very important to take active measures to prevent boredom and to exploit their training and potential to the full. We have found that it is helpful to rotate all the training staff at, for example, monthly intervals through all the units of the laboratory. If the laboratory consists of an automated laboratory, an emergency laboratory and units for protein work, enzymes, hormones and lipids, for example, they would have six stations to work in and spend only two months in every year at a particular set of tasks. This should go a long way towards solving the problem of boredom. Each graduate should be given a specific branch of the specialty to study and work on in depth.

The organisation of suitable breaks must not be forgotten, to avoid what has come to be known in industry as the four o'clock phenomenon, when efficiency tends to fall as the afternoon proceeds. Dr. Robinson (⁶) has shown that this applies in the laboratory also. He found that the coefficient of variation for serum potassium determinations rises alarmingly with time and the number of estimations completed.

Last but not least, the collection of the specimen. For blood letting many laboratories are now employing their own special service of phlebotomists who can be carefully trained, but most of us have to rely on ward staff for taking blood, and there are many other tasks which can only be done by the ward staff. Some of our tests need very close collaboration with the wards and for this aspect of the work the value of a comprehensive brochure, completely up-to-date and containing full instructions for all the tests offered by the laboratory, with notes on their interpretation, cannot be stressed too greatly. The brochure should also contain the latest figures for the laboratory error expressed as standard deviation or coefficient of variation. The use of the book by ward doctors and nursing staff, prevents many misunderstandings and innumerable telephone calls. Text books give many variations of the same test and one of them only should be selected by a laboratory, in consultation with the medical staff, for use on all patients served by the laboratory. Clinicians should be encouraged to complain about results and to be quality conscious in general. They should also make clear the purpose for which a test is intended; the clinical chemist would aim to provide a more precise result (for example by doing the test in duplicate or using a special technique) when he knows that some investigations may depend on observing a small change.

Once the sample has reached the laboratory there are many hazards of which we are all aware but do not always guard against, such as plasma standing for an excessively long time on the cells, samples for bilirubin standing in bright light, or specimens for the assay of labile substances standing on the laboratory bench instead of being frozen. All these points need to be guarded against continually.

In conclusion, it is worth moving away from the hospital for a few minutes to describe some of the invariably inexpensive and immensely valuable decisions that can and should be taken at higher levels. We will describe recent experiences in Britain to illustrate these points.

At the regional level, there is responsibility for many hospitals serving a population which may be as large as five million. This might well be a suitable level at which a senior person, active in hospital clinical biochemistry, should be appointed as part-time regional quality control officer. He might well start by obtaining information from hospitals about the details of their quality control programme asking for the names and functions of their local quality control officers, how much money is spent on control sera and what control techniques are used. He must make it clear that he will treat the information confidentially. He should invite laboratories to seek his advice on any problems they may have. And, finally, he should be willing to visit them if they request it.

At a higher, or national level, there is also much that can be done. For about a year, 200 large laboratories in Britain have participated in a survey organised from Birmingham by Professor Whitehead. Pooled sera are posted every two weeks to the laboratories which assay them for stated constituents and the results are sent to Birmingham where histograms, means, standard deviations and coefficients of variation are prepared by computer and posted back to the 200 laboratories. This trial is due to extend its operation to take in more hospitals and possibly also break down the analysis of results according to analytical methods used.

A British commercial company is interested in providing a similar and possibly more extensive service which, it is hoped, will start very soon on a free trial basis but later, if it is successful, laboratories would have to make some payment (possibly £ 100 a year) for the service.

We are also interested in providing some control of accuracy and are investigating complex techniques with a view to providing *referendary* standards. It should be possible to combine control of accuracy and precision control. Samples could be sent to a referendary centre for very accurate analysis.

At an even higher level, international work on quality control is also needed, and little has yet been done in this field. The International Union of Pure and Applied Chemistry and the International Federation of Clinical Chemists have important roles to play in defining units, establishing methods and practises regarding primary standards for accuracy control and in ensuring by example and publicity that what can be done is widely known and available throughout the world.

We have, we believe, covered most of the factors affecting the success of quality control, but it must be remembered that success in this particular field is never complete. Running a good quality control programme is one of the best exercises in humility anyone could wish for. If you are feeling pleased with yourself one day you can be absolutely certain that unpleasant surprises of any magnitude are only round the corner, but whether the quality is good or bad the comfort of knowing that it is under control is immense. We think the well known quotation from Bishop Westcott is most appropriate and a good note upon which to close:

« It is the vision of the ideal that guards the monotony of work from becoming the monotony of life».

It is a quotation which we think will stand a fair amount of thinking about.

Summary. — Accuracy is a more difficult aspect of quality control than precision and requires the use of special techniques which need not be applied recurrently. International co-operation is necessary. Factors which are stressed for the maintenance of a quality control programme are:

- 1) To use a relatively simple and manageable method.
- 2) To appoint a quality control officer.
- 3) Publicise the laboratory error.
- 4) Display quality control charts.
- 5) Take a pride in obtaining good results.

Producing good quality results depends on good management, satisfactory physical lay-out and organisation of the laboratory, care in the recruitement and training of staff and awareness that responsibility does not end at the door of the laboratory.

Much help can be obtained from regional quality control officers.

National, regional (possibly commercial) quality control schemes are of considerable value.

Riassunto (L'attuazione del controllo di qualità e i fattori che ne condizionano la riuscita). — L'accuratezza è un aspetto più difficile del controllo di qualità di quanto non sia la precisione, e richiede l'uso di tecniche speciali che non occorre vengano impiegate correntemente. È necessaria una collaborazione internazionale. I fattori da sottolineare per lo svolgimento di un programma di controllo di qualità sono:

- 1) usare un metodo relativamente semplice e agevole;
- 2) nominare un addetto al controllo di qualità;
- 3) rendere pubblici gli errori di laboratorio;
- 4) mettere in mostra la documentazione sul controllo di qualità;

۴

5) farsi un punto d'onore di ottenere buoni risultati.

Il produrre risultati di buona qualità dipende da una buona direzione, da una soddisfacente disposizione e organizzazione del laboratorio, da attenzione nell'assunzione e nell'addestramento del personale e consapevolezza che la responsabilità non finisce alla porta del laboratorio.

Molto aiuto può essere ottenuto da parte degli addetti al controllo di qualità regionale.

Schemi di controllo di qualità nazionali, regionali (possibilmente commerciali) hanno un valore considerevole.

REFERENCES

- (1) WHITBY, L. G., F. L. MITCHELL & D. W. Moss. Advan. Clin. Chem., 10, 65 (1967)
- (2) MITCHELL, F. L., J. PROFFITT & W. ANNAN. Proc. Assoc. Clin. Biochem., 4, 71 (1966)
- (8) CLARK, E. P. & J. B. COLLIP. J. Biol. Chem., 63, 461 (1925).
- (4) TRINDER, P. Analyst, 85, 889 (1960).
- (5) HALPIN, J. F. Zero Defects, Mc Graw-Hill, New York, 1967.
- (6) ROBINSON, R. Lancet, ii, 744 (1966).

274

The use of quality control data and interlaboratory test results in the standardization of clinical chemical methods

H. BÜTTNER

Institut für Klinische Chemie, Medizinische Hochschule, Hannover, Germany

Introduction

Clinical Chemistry is a kind of analytical work which may be more or less complex depending on the type of biological material to be analyzed. As a consequence of the difficulties arising in clinical chemical work, there is a number of different methods or modifications of methods for a given analysis. In most cases the Clinical Chemist selects the method to be used in his laboratory by subjective criteria: recommendation by colleagues or reagent dealers, habit etc. We can expect a real success in clinical chemical routine work also in the smallest laboratories only if we select the optimal methods from the multitude that are recommended. For this selection we need objective criteria to compare the different methods.

The value of a clinical chemical method is determined mainly by the reproducibility of the method. A method having a higher reliability is better than another one having a lower reliability. As a consequence we can use for comparison the parameters of analytical errors of a method. But these errors should not be evaluated in a pilot laboratory under ideal conditions. We have to use data obtained in routine laboratories. I will show that these data can be obtained from the results of statistical quality control within one laboratory and from the results of interlaboratory comparisons.

The following criteria should be used for the selection of the best analytical methods:

I	-	precision	within	ı ser	ies	(inform	nations	abtainab	la
		precision	from	day	to	day	finiori	liter	optamab	data
	1	accuracy					irom	quanty	control	uata

- long time performance
- in individual laboratories
- standard deviation between laboratories

Ann. Ist. Super. Sanità (1971) 7, 275-280

- II specificity
 - sensitivity
- III difficulty of the method
 - time requirement
 - instrumentation requirements
 - current costs

Definition and evaluation of the parameters of analytical errors

The modern theory of errors in analytical chemistry is based on the work of C. F. Gauss published at the beginning of the last century. According to his work we can distinguish three types of errors, which all contribute to the total error of a given result: gross errors, systematic errors and random errors.

As a measure of random errors we can use the dispersion of the results obtained by several determinations using the same sample. It is conventional to calculate the standard deviation and to call it « precision» (Fig. 1).



Fig. 1. - Parameters for the measurement of random and systematic analytical errors.

But the precision value of a given analysis will be different depending on the working conditions. We shall observe a small standard deviation or good precision if the same person carries out the analyses in the same batch or series (« precision within series»). And we shall find a greater standard deviation if the analyses are carried out by different persons on different days (« precision from day to day ») (Fig. 2). In statistical quality control systems, used in many clinical chemical laboratories today, precision data

276

٠

are an important part of the control system. This offers the opportunity to collect precision data from routine laboratories. Straumfjord and Copeland (¹) in 1965 published some precision data of this kind. But as a prerequisite for standardization we have to collect more data with distinction of different analytical methods.

As a measure of systematic errors (bias) we use the quantity « accuracy » which means deviation from the target value. This deviation can be deter-



Fig. 2. - Spread of analytical errors in clinical chemical laboratories.

mined by analyzing a standard solution with a known target value. As with precision data we should get accuracy data from routine quality control systems.

However, a big problem comes up at this point. If we wish to use the parameter « accuracy» as a criterion of the reliability of a method we have to determine an exact accuracy value. That means, we need a primary standard solution made by weighing the pure substance to be analyzed in a suitable solvent. In Clinical Chemistry it is often very difficult to prepare primary standards. In my opinion, it is urgent to select reference substances which can be used to prepare primary standards. The US-National Bureau of Standards has announced some clinical chemical standard substance. But I think it would be important for the IFCC to coordinate efforts in this direction.

In many cases there is no possibility today to prepare primary standards. So we have to use secondary standards, the concentration of which is

Ann. Ist. Super. Sanità (1971) 7, 275-280

analyzed by chemical methods. To obtain true target values of a secondary standard we have to analyze the solution using an independent reference method.

A reference method is that having the highest precision and accuracy available. This method should give absolute values. Generally a reference method is not suitable for use in the routine laboratory because it requires special and expensive instrumentation and much more time.

From the quality control system of a routine laboratory we can obtain important information on the long term performance of a given method. Hitherto this criterion is not used in the selection of a reliable method. As a measure of the long term performance we could use the frequency of the «out of control» event in a given period of time.

The dispersion of results obtained by several determinations using the same method and samples of the same source is greater among different laboratories than within the same laboratory.

This standard deviation can be determined by an interlaboratory comparison by analyzing samples of the same source in a group of different laboratories. In most interlaboratory comparisons the standard deviation among laboratories is two or three times the standard deviation within the same laboratory (Fig. 3). As an interpretation of this phenomenon we can



Fig. 3. — Dispersion of results obtained with the Autoanalyzer in different laboratories (CAP-Survey 1965).

make the hypothesis that in the different laboratories the systematic errors are of different size: in most cases statistical treatment of the results of interlaboratory tests confirms this hypothesis. That is, the standard deviation among laboratories is a measure of systematic and random errors. Interlaboratory comparisons are carried out in a « blind» manner, the indi-

279

vidual quality control system of most laboratories is functioning as a « known» control. Therefore the standard deviation between laboratories will be of additional value as a criterion in the selection of reliable methods.

Standardization of clinical chemical methods

From the results of interlaboratory comparisons we know that nearly all clinical chemical methods, used in routine laboratories, show systematic errors. As a consequence the results of different laboratories are not comparable. Some people expect that these difficulties could be overcome if all laboratories would use exactly the same method. This is the meaning of « standardized method». But we should realize that it is dangerous to standardize a method which is not perfect in its analytical basis. Such a method will show different systematic errors also after standardization. In other words, standardization will be successful only if we standardize an optimal method.

To select this optimal method we can use the parameters of errors which we discussed as optimal criteria.

The standardization of methods is closely connected with the standardization of instrumentation. With regard to analytical errors the reliability of an analytical method is limited by the reliability of the instruments used. In particular this is true for modern automated devices used in Clinical Chemistry.

After discussing some general problems of standardization philosophy in Clinical Chemistry let us now ask how to do standardization in practice. At the Geneva international meeting in 1969 I proposed a plan for the standardization of a clinical chemical method, based on objective criteria of reliability. As a prerequisite we need a group of reference laboratories. These laboratories are concerned with the development of absolute reference methods and with the preparation of pure primary standards. The method to be standardized should be checked against the reference method. The method then should be tested in a group of laboratories. In this period information should be collected on precision, accuracy and long time performance.

At this point we can tentatively define the minimum requirements for precision, accuracy, etc. In my opinion routine laboratories, in each country, should be free to choose this recommended method or another one. Using another method the laboratory should meet the minimum requirements.

This plan would secure the following:

1. Recommended methods should have a valid analytical basis. Analytical errors under conditions of routine work are known.

2. Any laboratory can choose another method with the same analytical error. Development and progress in clinical chemical methodology are not restricted.

Ann. Ist. Super. Sanità (1971) 7, 275-280

3. Recommended methods should be checked against absolute reference methods. As a consequence the comparability of the results of different laboratories will be improved.

The accomplishment of this plan would probably exceed the possibilities of one country. I think we need a teamwork across the frontiers to do this.

Summary. — The standardization of clinical chemical methods is connected with the urgent need to improve the present poor comparability of the results obtained by different laboratories.

It is necessary to use objective and rational criteria for the selection and evaluation of the methods, and to abandon the subjective criteria largely used so far. Among other things it should be kept in mind that a good method must provide reliable results not only in the pilot laboratory, but also in routine work.

The criteria which are believed to be valid are indicated and illustrated. An additional criterion to evaluate a method may be based on the standard deviation between laboratories, calculated on the data obtained in a group of laboratories utilizing the same material for quality control. It is proposed to indicate the minimum requirements of precision, accuracy, etc. of a given method, before starting its standardization.

Riassunto (L'impiego dei dati di controllo di qualità e dei risultati ottenuti in diversi laboratori di analisi nella standardizzazione dei metodi chimico-clinici). — La standardizzazione dei metodi chimico-clinici è connessa al problema serio ed urgente di migliorare l'attuale scarsa comparabilità dei risultati forniti dai diversi laboratori.

Si indica la necessità che a criteri soggettivi, diffusamente seguiti nella scelta dei metodi, vengano sostituiti criteri oggettivi e razionali di selezione e di valutazione che tengano tra l'altro conto del fatto che un buon metodo deve fornire risultati attendibili, non solo nel laboratorio pilota, ma anche nel lavoro di routine.

I criteri ritenuti utilizzabili vengono indicati ed illustrati. Un criterio addizionale per valutare l'idoneità di un metodo può essere basato sulla deviazione standard « tra laboratori» dei dati ottenuti in un gruppo di laboratori che utilizzino lo stesso materiale per il controllo di qualità. Si propone di fissare le minime richieste di precisione, accuratezza, ecc., di un metodo prima di dare effettivo inizio alla sua standardizzazione.

REFERENCES

(1) STRAUMFJORD, J. V. & B. E. COPELAND. Am. J. Clin. Pathol., 44, 452 (1965).

Ann. Ist. Super. Sanità (1971) 7, 275-208

The blind approach in quality control (*)

G. VANZETTI and D. PALAZZI

Laboratorio di Biochimica e Centro di Biochimica analitica, Ospedale Maggiore Ca' Granda, Milano, Italy.

As a rule, the quality control programs used in clinical laboratories are based on repeated analyses of the same control sera (2-4); this can give rise to the so-called observer's bias, since the analyst is usually aware of the expected results.

To obtain a truly objective assessment of the analytical performance, the analyst should have no prior knowledge of the control data, and become acquainted with them only after performing the analyses.

In other words, the analyst should adopt a « blind system», similar to that currently employed by clinicians for assessing the therapeutic activity of drugs. Only through a blind approach can the analyst be certain to avoid « observer's bias».

Starting from this viewpoint, in 1968, we planned a collective quality control program, based on the weekly monitoring of the routine analyses by the use of multiple control sera of a composition unknown to the analyst.

At our request, a well-known manufacturer of diagnostic products prepared and analyzed very carefully nine control sera of different composition (Table 1). The composition of the sera was checked in several reference laboratories, and the results were in good agreement. These liquid sera, prepared in large amounts, were labelled and distributed in 8-ml vials, and were then used for a collective quality control program, with the participation of 29 hospital laboratories located mainly in Northern Italy and especially in the Milan area.

Our program started in March 1969, and came to an end in February 1970, lasting 48 weeks. At the start of the program we gave a code number to each laboratory and we sent to each laboratory director a box of 12

の語いは

^(*) An interim report on this subject was given at the Symposium on Quality Control, 7th International Congress of Clinical Chemistry, Geneva 1969 and has been published in the Proceedings of the Congress (¹).

SUBSTANCE 2 3 6 7 1 4 5 8 9 UREA, g/1 0.23 0.59 0.29 0.51 0.40 0.27 0.35 0.46 0.55 GLUCOSE, g/l 0.92 2.70 1.28 2.34 1.81 1.10 1.54 2.07 2.52TOTAL PROT., g/100 ml 7.00 4.80 6.56 5.24 5.90 6.78 6.23 5.57 5.02 CHLORIDE, mEq/1 . . 101 85 98 88 93 100 96 91 87 9.9 CALCIUM, mg/100 ml . 7.3 9.4 7.8 8.6 9.6 9.0 8.2 7.5 SODIUM, mEq/1 . . . 144 111 137 118 127 141 132 122 114 6.30 5.60 5.25 6.55 POTASSIUM, mEq/1 . 4.40 6.80 4.90 4.65 5.95 INORG. PHOSPHORUS, 5.90 5.30 3.20 2.90 3.50 4.40 3.95 4.85 5.60 mg/100 ml CREATININE, mg/100 ml 0.90 2.00 1.10 1.80 1.50 1.00 1.30 1.65 1.90 IRON, µg/100 ml . . . 95 71 90 76 83 93 87 80 73

Composition of control sera

TABLE 1

vials containing the control sera, distributed at random and numbered from 1 to 12. Each vial was labelled with the scheduled date of analysis, at weekly intervals.

At the same time and for each laboratory we prepared a check-list of the 12 control sera as well as 12 scaled envelopes numbered 1 to 12 containing the control data for the corresponding vials.

During the first week, each laboratory analyzed the control serum of vial N. 1, and sent us the analytical results (Table 2).

TABLE 2

Specimen of form used for collection of weekly analytical results

Laboratory AZ Week 37

Control serum to be analyzed on 2-12-1969

UREA, g/1	. 0.25
GLUCOSE, g/1	. 1.15
TOTAL PROTEINS, g/100 ml	. 6.95
CHLORIDE, mEq/1	. 99
CALCIUM, mg/100 ml	. 10.2
Sodium, mEq/1	. 148
Ротаssium, mEq/1	. 4.30
INORG. PHOSPHORUS, mg/100 ml	. 2.70
CREATININE, mg/100 ml	. 1.10
IRON, µg/100 ml	. 110

Results should be sent within 2 days to the Center of Analytical Biochemistry, Ospedale Maggiore di Milano.

Ann. Ist. Super. Sanità (1971) 7, 251-292

After receiving the results, we sent each laboratory director the sealed envelope n. 1, with the control data of the corresponding serum (Table 3), to enable him to compare his results with the reference data and to evaluate

TABLE 3

Specimen of reference data sheet sent to each laboratory for control of results obtained

Ospedale Maggiore Ca' Granda di Milano Center of Analytical Biochemistry

Laboratory AZ Week 37

REFERENCE DATA

UREA, g/1			ŝ			•		•	•		•	•	•		0.23
GLUCOSE, g/l											×	•			0.92
TOTAL PROTEINS, g/10)0	ml						÷		•	2				7.00
CHLORIDE, mEq/1 .			÷		•	•				•	•	R	•		101
CALCIUM, mg/100 ml		•	•			•				•				•	9.9
Sodium, mEq/1		• •		×		÷				•				S	144
POTASSIUM, mEq/1	• •		•		•	•	÷		•	•	•	ł		•	4.40
INORG. PHOSPHORUS,	mg	/10	0	m	L			×					e.	- 285	2.90
CREATININE, mg/100	ml		÷		•						÷	÷			0.90
Iron, $\mu g/100~m1$.		•			•	•	•		ě	•	•	÷	X	•	95

the performance of his laboratory. This was repeated at weekly intervals for 12 weeks.

At the end of the first 12 weeks we supplied each laboratory with another 12-vial package for continuation of the program: we did this again after 24 and 36 weeks. After 48 weeks the program was concluded.

After each 12-week period, we prepared a summary sheet and an evaluation report for each laboratory. For the summary sheet, we transferred the weekly results to punch cards, and recorded with the help of a computer the individual data on a sheet, both as absolute values and as percentages of the expected values (Table 4).

For the evaluation report we carried out a statistical analysis on the percentages; for each substance we calculated the scatter and the confidence limits of the percentages, with an estimate of the average percentage and of its standard deviation (Table 5). We also divided percentages into classes, in order to show the distribution of the analytical data (Table 6).

For each laboratory and each substance we calculated the percentage of « acceptable» results, thas is, the percentage of the results that did not

TABLE 4

			States and Long							
WEEK	UREA	GLUCOSE	TOT. PROT.	сі	Са	Na	к	Р	CREAT.	IRON
25	24	96	720	102	99	144	460	290	85	92
	23	92	700	101	99	144	440	290	90	95
	104%	104%	103%	101%	100%	100%	104%	100%	94%	96%
26	28	122	680	97	96	139	460	400	110	88
	29	128	656	98	94	137	490	350	110	90
	96%	95%	103%	99 %	102%	101%	93%	114%	100%	97%
27	40	174	580	95	85	122	563	442	144	75
	40	181	590	93	86	127	560	440	150	83
	100%	96%	98%	102%	98%	96%	100%	100%	96%	90%
28	54	240	540	90	77	113	691	529	188	80
	51	234	524	88	78	118	630	530	180	76
	105%	102%	103%	102%	98%	95%	109%	99%	104%	105%
29	58	285	500	80	71	117	691	599	200	71
	59	270	480	85	73	111	680	590	200	71
	98%	105%	104%	94%	97%	105%	101%	101%	100%	100%

Summary sheet of analytical results from one laboratory (*) Laboratory 25

(*) In the table are reported the results of only five weeks; figures are reported without punctuation.

Evaluation report for one laboratory Laboratory 25 TABLE 5

SUBSTANCE	AVERAGE PERCEN- TAGE	ST. DEV.	m 2∗	\overline{m} + 2s	PRECISION SCORE	RELIAB. INDEX %
UREA	99.71	3.29	93	106	193	100
GLUCOSE	99.63	3.61	92	106	131	100
TOTAL PROTEINS	103.13	2.01	99	107	135	100
CHLORIDE	99.51	2.55	94	104	103	91
CALCIUM	100.73	2,65	95	106	150	83
SODIUM	100.28	3.40	93	107	84	83
POTASSIUM	102.51	4.08	94	110	89	100
INORG. PHOSPHORUS .	101.89	4.92	92	111	134	91
CREATININE	99.43	4.26	90	107	154	100
IRON	97.74	4.19	89	106	162	100

Average precision score, lab. 25 = 134; average precision score, all laboratories = 100; average reliability index, lab. 25 = 95; average reliability index, all laboratories = 75.

Ann. Ist. Super. Sanità (1971) 7 281-292

	< 85	85-95	95-105	105-115	> 115
UREA	0	0	11	1	0
GLUCOSE	0	0	11	1	0
TOTAL PROTEINS	0	0	11	1	0
CHLORIDE	0	1	11	0	0
CALCIUM	0	0	10	2	0
SODIUM	0	0	10	2	0
POTASSIUM	0	1	8	3	0
INORG. PHOSPHORUS	0	0	9	3	0
CREATININE	0	2	9	1	0
IRON	0	3	8	1	0

Summary sheet of percentage distribution for one laboratory Laboratory 25

deviate from expected values by more than $\pm 5 \%$ in the case of sodium, chloride and calcium, or $\pm 10 \%$ for the remaining routine analyses. We then calculated the overall percentage of acceptable results (so-called « reliability index») for a given laboratory.

We also assigned «precision scores».



 Fig. 1. — Estimation of the precision score S_{z,i} of a given laboratory (i) for a given substance (x) from the quotient <sup>S_{i, x}/_{S_x}, on the basis of formula [1]. (See text).
</sup>

For each substance, we calculated

an « average » standard deviation by computing the logarithmic average \bar{s}_x of the individual standard deviations. We then calculated a score $(S_{x, i})$ by comparing the individual standard deviation $s_{x,i}$ with this average, using the following equation:

$$S_{x,i} = 100 - \frac{50}{\log 2} \cdot \log \frac{s_{x,i}}{\bar{s}_x}$$
[1]

As shown in Fig. 1, a score of 100 means that the individual scatter coincides with the average; a score of 50 means that the individual scatter is twice the average; a score of 150 means that the individual scatter is half the average (precision better than average), and so on.

Ann. Ist. Super. Sanità (1971) 7, 281-292

Last, we calculated an overall precision score for each laboratory, by averaging the scores obtained for each substance. We included all these data — average percentages, confidence limits, distribution of percentages, reliability index, precision scores, etc — in the evaluation report shown in Table 5.



Fig. 2. — Average coefficient of variation obtained for glucose (a), total proteins (b), calcium (c), sodium (d), potassium (e), chloride (f), creatinine (g), inorganic phosphorus (h) and iron (i) in subsequent 12-week periods (all laboratories).

By sending the summary sheet and the evaluation report once every 12 weeks to each of the participants we enabled them to check the pertinent data and calculations and to assess the overall performance of their laboratory. We ensured secrecy to all participating laboratories; however, while ignoring the individual results of the other participants, each analyst was made aware of his own performance and score, compared to the average for the group.

286

After each 12-week period, we discussed the implications of our quality control program, as well as the analytical methods and inherent problems in several small group meetings, in order to promote mutual cooperation.

The response of all participants was good, sometimes even enthusiastic. Most of them were already using control sera routinely: however, they all agreed that the collective multilevel program was truly rewarding. This was confirmed by a questionnaire filled out by the laboratory directors before the end of the 48-week period: in answering the questionnaire, they all stressed the importance of a regular quality control program, and expressed approval for the « blind» approach adopted.

The average results are shown in Fig. 2. In the diagrams, we report the average values of the standard deviations (logarithmic averages) calculated for each substance and for each of the four 12-week periods under consideration.

We left out the analytical results concerning urea, because the titer of urea decreased during the experiment in several control sera, and therefore the reference data on urea became invalid.

For all the other substances there was an overall, gradual improvement in the performance of the laboratories, as shown by the significant decrease of the average standard deviation in time: the average decrease between the first and the last 12-week period was about 30 %.

The performance of many laboratories was still unsatisfactory, as shown by the high scatter of results, but at least all analysts became aware of their shortcomings; the overall improvement shows that in many laboratories steps were being taken in the right direction.

Our « multilevel» quality control program offers two main advantages over the current programs using one or two control sera of known composition. First, it gives objective results: the possibility of « observer's bias» is eliminated. Secondly, it is performed on a collective basis: this encourages a useful exchange of information, and fosters healthy competition as well as cooperation between the participants.

The multilevel program, however, has a few drawbacks also. Many different control sera are needed, and this may increase production costs; besides, the statistical calculations are more complex than in the usual programs.

For a more extensive application of our program, many lyophilized control sera of different composition are required. The same manufacturer who supplied us with the liquid control sera used in our experiment agreed to our request for multiple lyophilized sera: these are now being prepared in large amounts. These sera will enable us to broaden the range of our con-

Ann. Ist. Super. Sanità (1971) 7, 281-292

trols and to include several organic substances that we could not assay before.

We are now preparing a new 48-week quality control program with lyophilized sera. In our group, there will be room for a maximum of 50 participants; we hope that similar groups will be started elsewhere.

We also envisage a self-control program: in this case the laboratory director will be supplied with a package of 12 different control sera, labelled 1 to 12, and with 12 sealed envelopes, also labelled 1 to 12, containing the corresponding control data. The laboratory director will thus be able to entrust the analyses to his technicians, and then compare the analytical results with the control data without having to wait for the results from our Center.

Statistical calculations can be made in the laboratory, or entrusted to an agency having access to an electronic computer.

And now we wish to make a few critical considerations. The preliminary analyses on control sera, needed to establish the reference data for subsequent analyses, must be performed with the greatest care. The « reliability index» will be valid only if the control data themselves are reliable and precise.

Assigning a precision score independent of the reliability index is justified mainly because there are still uncertainties about the absolute concentrations of several serum components.

An appraisal of the standard deviation on the basis of 12 determinations only is subject to considerable error: the validity of our statistical calculations can therefore be challenged. We would obtain better results through analysis of two control sera a week: by the end of 12 weeks we would have 24 analytical data for each substance instead of 12, and this would lead to more reliable statistics. However, the cost of the program and the burden on the analyst would be greater.

A tolerance in the range of ± 10 % for the errors of most routine analyses may seem excessive; the same holds for the range of ± 5 %, adopted for sodium, chloride, and calcium. At least in our country stricter rules would not be realistic today, but it will perhaps be possible to adopt them in the future. Other criteria, for instance those recommended by Tonks (⁵), may be more rational, but they require more complex calculations.

The inclusion of confidence intervals in the evaluation sheet appears to be important, since these intervals show the analyst how much his data are off the mark.

Our program is very flexible and can even be adapted, if required, to suit the individual needs: it should be considered as an addition to, and not as a substitute for, the usual controls. Now a few words on another simple quality control system deserving a place in daily laboratory routine; this system, wich can be also performed under « blind » conditions, is the daily monitoring of routine analytical methods by randomly selected patient specimens (*).

For each method, a single patient specimen is selected at random as control sample for replicate analysis: an aliquot is analyzed as usual along with the daily routine, and another aliquot (so-called «carry-over» specimen) is stored and analyzed with the next day run. The analytical results and their differences are recorded on a difference control chart (Table 7).

TABLE 7

Control of analytical precision of total protein assay by means of "carry-over" specimens

Absorbance values of reference solution	a) 0.43 b) 0.43	0.44	0.43	0.42	0.43	0.43
Difference	0.00	0.01	0.00	0.00	0.01	0.01
Analytical results given by carry-over speci- mens	a) 7.20 b) 7.45	6.85 6.90	7.10 7.05	7.05 7.00	6.75 6.55	7.00 7.10
Difference	+ 0.25	+ 0.05	- 0.05	- 0.05	- 0.20	+ 0.10
Day of the month	1	2	3	4	6	7

Month of April 1970

We report here the data obtained during the first week of April 1970: in the real chart we report the data of the entire month, and we use the differences to construct an appropriate diagram and to estimate the «between-run» standard deviation. We include in another section of the chart the results given by, and the differences between duplicate analyses of a single randomly selected patient specimen; and we use these data to estimate the «within-run» standard deviation of the analyses.

This procedure is repeated daily for a month: the «between-runs» standard deviation and the corresponding confidence limits are estimated once a month from the collected data. One of the following formulas may be used for the calculation of the standard deviation:

$$s = \sqrt{\frac{\sum d_i^2}{2n}} \qquad [2 a] \qquad \qquad s = \frac{\sqrt{\pi}}{2} \cdot \frac{\sum d_i}{n} \qquad [2 b]$$

(*) A more detailed description of this system has been published elsewhere (*).

Ann. Ist. Super. Sanità (1971) 7, 281-292

where s it the standard deviation, d_i are the differences between duplicate analyses, and n is the total number of duplicate analyses.

We can also estimate the confidence limits of the results, expressed as $\pm F \%$, by the formula:

$$\pm F \% = \pm 100 \cdot \frac{2 s}{\overline{c}}$$

where \bar{c} is the average concentration of the given substance in the control specimen.

The control specimens are selected at random, either by the analyst himself before analysis, or by a supervisor: in this way the analyst has no prior knowledge of the expected results.

In order to obtain valid results, the « carry-over» specimens must be preserved with care. Unstable compounds must be stabilized by physical procedures like refrigeration or freezing, or by the addition of suitable blocking agents, such as fluoride in the case of glucose. In some cases, for instance bilirubin determinations, pathological (hyperbilirubinemic) sera should be preferred as control samples, since normal sera give low absorbance values.

This quality control system, already mentioned by Gray (7) and recommended by Whitby, Mitchell and Moss (3), has been used successfully in our laboratory for several years, for all routine chemical assays. Properly used, it allows an objective estimate of the standard deviation and of the confidence limits of analytical methods: it allows to monitor the precision, but not the accuracy, of analyses and must therefore be complemented by the analysis of control specimens of known composition to uncover systematic bias.

In conclusion, the «blind» system of quality control provides an objective evaluation of the performance of the clinical laboratory and of the individual analyst; therefore it seems to represent a real advance over the customary control systems, and to deserve further extensive testing in the laboratories.

Summary. — The results of a joint quality control experiment performed in 29 Italian hospital laboratories are reported.

The experiment lasted 48 weeks (from March 1969 to February 1970) and was performed using a « blind» system. Nine different liquid control sera, especially prepared and analyzed by a well known manufacturer were utilized, after being checked again.

Each week a different serum, the composition of which was unknown to the analyst, was analyzed in each laboratory. The analytical results were then compared to the expected results. í

17

The results collected after 12, 24, 36 and 48 weeks were analyzed by a computer.

After each 12-week period the mean percentage, the coefficients of variation and the confidence limits were calculated for each laboratory and for each substance analyzed. Moreover, a « precision score» and the percentage of « acceptable» results (values comprised between 95 % and 105 % of the expected results for sodium, chloride and calcium; between 90 % and 110 % for the other 7 substances measured) were calculated for each laboratory and for each substance.

The control system used allows the analyst to evaluate objectively the performance of his laboratory and was favorably accepted by all participants.

In the course of the experiment there was on the average a significant improvement in the quality of the analytical performance of the participating laboratories.

Riassunto (Sperimentazione di un sistema cieco per il controllo di qualità nei laboratori chimico-clinici). — Si descrivono i risultati di un esperimento collettivo di controllo di qualità effettuato in 29 laboratori ospitalieri italiani.

L'esperimento ha avuto la durata di 48 settimane (marzo 1969-febbraio 1970): esso è stato effettuato con un sistema « cieco», impiegando 9 diversi sieri di controllo allo stato liquido, appositamente preparati e analizzati da una nota Casa produttrice, e accuratamente ricontrollati.

In ogni laboratorio veniva analizzato ogni settimana un siero diverso, di composizione sconosciuta agli analisti. I risultati analitici venivano quindi confrontati con i risultati attesi.

I risultati raccolti dopo 12, 24, 36 e 48 settimane sono stati analizzati mediante un elaboratore elettronico.

Dopo ciascun periodo di 12 settimane sono stati calcolati per ogni laboratorio e per ogni sostanza la percentuale media, i coefficienti di variazione e i limiti fiduciari. Per ogni laboratorio e per ogni sostanza è stato calcolato inoltre un « punteggio di precisione», come anche la percentuale dei risultati « accettabili» (valori compresi tra il 95 e il 105 % dei valori attesi per sodio, cloro e calcio; fra il 90 e il 110 % per le altre 7 sostanze dosate).

Il sistema di controllo adottato permette agli analisti di valutare in modo obbiettivo le prestazioni del proprio laboratorio; esso è stato accolto con grande favore da tutti i partecipanti.

Nel corso dell'esperimento vi è stato mediamente un sensibile miglioramento nella qualità delle prestazioni analitiche dei laboratori partecipanti. We wish to express our thanks to the following people, to whom we are greatly indebted:

- The directors and analysts of the participating laboratories (*);
- Dr. F. Santini of Pharmascal, and Dr. G. Anido of Dade for their constant cooperation, and for the generous gift of control sera;
- Prof. W. De Ambrogio, Dr. E. Buro, and Dr. F. Taddei of Siemens Data for the development of the computer program;
- Mrs. B. Albertoni and Mrs. A. Nardeschi for secretarial work.

REFERENCES

- (1) VANZETTI, G. & D. PALAZZI. Proc. 7th Intern. Congr. Clin. Chem., Geneva/Evian 1969. Karger, Basel 1970. Vol. 1, p. 307.
- (2) HENRY, R. J. Clinical Chemistry: Principles and Technics. Harper and Row, New York, 1964, p. 122.
- (3) WHITBY, L. G., F. L. MITCHELL & D. W. Moss. Advan. Clin. Chem., 10, 65 (1967)
- (4) BÜTTNER, H. Z. Klin. Chem. Klin. Biochem., 5, 41 (1967).
- (5) TONKS, D. B. Clin. Chem., 9, 217 (1963).
- (*) VANZETTI, G. Proc. Second Intern. Symp. on Quality Control, Geneva 1968.
- (7) GRAY, G. H. Broadsheet No. 30 (N. S.), Assoc. Clin. Pathol., 1961.

- Laboratorio clinico, Ospedale Civile, Brescia (Dr. E. Bozzetti);
- Divisione di Biologia Sperimentale, Istituto Tumori, Milano (Dr. V. Carminati);
- Laboratorio clinico, Ospedale Civile, Cuggiono (Dr. F. Cefis);
- Laboratorio elinico, Ospedale Civile, Padova (Dr. G. Ceriotti);
- Laboratorio clinico, Ospedale di Circolo, Merate (Dr. I. Ferrario);
- Laboratorio clinico, Ospedale Civile, Saronno (Dr. S. Fiocca); Laboratorio clinico, Ospedale Civile, Vigevano (Dr. C. Franzini);
- Laboratorio clinico, Clinica S. Giuseppe, Milano (Dr. E. Garavoglia);
- Laboratorio clinico, Ospedale Civile, Vimercate (Dr. S. Garbin);
- Laboratorio clinico, Ospedale di Sesto S. Giovanni (Dr. L. Grassi);
- Laboratorio clinico, Ospedale Civile, Busto Arsizio (Dr. V. Malacrida);
- Laboratorio clinico, Ospedale Civile, Grosseto (Dr. G. Marini);
- Laboratorio clinico, Ospedale Civile, Gallarate (Dr. M. Martinazzi);
- Laboratorio clinico, Ospedale Civile, Monza (Dr. D. Palazzi);
- Laboratorio di Biochimica, Ospedale S. Carlo Borromeo, Milano (Dr. C. Petrini Arisi);
- Laboratorio clinico, Ospedale Bassini, Milano (Dr. R. Petrini);
- Laboratorio elinico, Ospedale Civile, Sondrio (Dr. F. Ports);
- Laboratorio clinico, Ospedale Civile, Lecco (Dr. L. Ragaini);
- Laboratorio clinico, Ospedale Maggiore, Novara (Dr. G. Rossi);
- Laboratorio clinico, Ospedale C. Borella, Giussano Brianza (Dr. A. Rovescalli);
- Laboratorio clinico, Ospedale Civile, Tradate (Dr. B. Sabotto);
- Laboratorio elinico, Ospedale Civile, Vicenza (Dr. L. Salviati);
- Laboratorio clinico, Ospedale di Circolo, Seregno (Dr. F. Sinisi);
- Laboratorio elinico, Ospedale Civile, Legnano (Dr. E. Strobino);
- Laboratorio elinico, Ospedale Civile, Rho (Dr. L. Tropeano);
- Laboratorio clinico, Ospedale Civile, Magenta (Dr. A. Tommasini Degna);
- Laboratorio di Biochimica, Ospedale Maggiore Ca' Granda, Milano (Dr. G. Vanzetti);
- Laboratorio Centrale, Ospedale Fatebenefratelli, Milano (Dr. E. Vidari).

^(*) List of participating laboratories :

Laboratorio Centrale, Ospedale Policlinico, Milano (Dr. L. Attili);