Pure & Appl. Chem., Vol. 45, pp. 59–62. Pergamon Press, 1976. Printed in Great Britain.

THE PRACTICAL APPLICATION OF REFERENCE METHOD TECHNOLOGY

FRED L. MITCHELL
Division of Clinical Chemistry, Medical Research Council, Clinical Research Centre, Harrow, Middlesex, UK

Abstract—The main value of reference technology will be in the selection of techniques for service use and in defining reference ranges for the concentration of substances in the body fluids of normal individuals. Initially it should find application in the present national and international quality control schemes to distinguish those techniques which, as currently operated, are sound, from those in which accuracy can be demonstrated to be faulty. Over the long-term it is to be expected that when new techniques for service use are submitted for publication, they should not be accepted unless the results which they have produced have been satisfactorily compared with results obtained by the appropriate reference technique.

In the application of the reference technique for calcium to the Wellcome Group Quality Control Program it has been shown that two widely used routine methods for calcium in serum tend to give lower results than the reference value at high calcium concentrations, and higher results at low concentrations. Attention can now be focused on this effect with a view to its elimination either by suggesting the modification or abandonment of the techniques in question. Herein lies the key to further improvement in interlaboratory quality control which has recently tended to plateau.

Reference technology is likely to be required for four main purposes:

1. The assay of commercial quality control material.
2. The assay, either in prospect or retrospect of material used in national quality control programs.
3. "One off" measurements required by individual laboratories for development work on instruments and methods.
4. For monitoring the effects of drugs or their metabolites, etc.

Definitive methods will almost certainly always be difficult and will be required only rarely, but even reference methods will not be easy to set up and are likely to prove uneconomic to keep running for individual laboratories to be able to do their own assays as required. Therefore centers need to be designated to carry out certain assays on a national and international basis.

1. INTRODUCTION

Barely a decade has passed since quality control was an unknown term in clinical chemistry and it was falsely assumed that all results obtained by recognized methods were sufficiently accurate and precise. Since then, complex systems of control have become common practice, so far they control precision only, and not accuracy. It has been said that this is all which is required, since the clinical value of a measurement lies solely in its ability to aid diagnosis or monitor therapy, and not in its ability to measure specifically a certain substance. This is not acceptable for several reasons:

1.1 Different techniques for measuring the same substance may produce widely different results with a high degree of precision.

1.2 Without a means of showing which results are inaccurate, it is difficult to persuade laboratory heads to accept one technique in preference to others. Thus interlaboratory variation is perpetuated.

1.3 The rationale behind the use of an assay cannot be fully elucidated in all circumstances if there is no certainty that a particular substance and that alone is being measured.

1.4 The purist attitude with which a responsible scientist approaches his work cannot be satisfied.

The suggestion has been made that the overall mean value obtained by a large number of laboratories on samples from the same specimen "state-of-the-art" or "consensus value" should be accepted as correct; then methods giving results widely different from this should be carefully investigated and possibly discarded. The basis for this belief is that an accurate result is likely to ensue when several different methods are used by the laboratories concerned. This argument has force only if the principles upon which the methods are based are widely different, with no common factor, and if the mean values obtained by the different methods are very close. Often these rules have not been applied and any assumption that the state-of-the-art value is accurate, is not valid. The problems which emerge because of an inability to control accuracy are exemplified by the situation with the measurement of uric acid; a variety of methods exist, each capable of high precision yet giving results which may differ one from another by as much as 25%. It is still not possible to determine which technique, if any, produces the right answers and unless a method can be incontrovertibly proven faulty, it will continue to be used.

Many more examples could be quoted, two of the worst being the current situations with the measurement of blood lead and of urinary oestriol in pregnancy. Interlaboratory variations in enzyme measurement are particularly worrying where not only does the measurement of an activity rather than a substance itself present problems, but general agreement is lacking upon the use of units of measurement.

J. P. Cali has explained previously that the only answer to the control of accuracy lies in the development and use of reference technology.

2. NOMENCLATURE

Before any new principles can be applied in practice, the terminology used must be clearly defined. The present limited nomenclature in current use in the area of reference technology has evolved by trial and error and by cooperation between research workers and field testers.

J. P. Cali has explained the development of *definitive*

and *reference* methods. Other terms which have been suggested, such as absolute, referee and referendary have been discarded. Final definitions are presently subject to debate, but the following might be said to be currently acceptable.

2.1 *Definitive method.* A technique for measuring the concentration of a given substance in samples of a particular biological material which (irrespective of the total cost involved) gives results which are accepted as the nearest attainable to the true values, and the accuracy of which can be clearly proved on theoretical grounds.

2.2 *Reference method.* A technique which can be carried out with generally available equipment and which gives results which are statistically indistinguishable, or which are biased in a clearly defined way, from those obtained by the correspondingly recommended definitive method.

2.3 *Special reference method.* (It is necessary to specify a third type of method for measuring those substances for which definitive and reference methods cannot apply, e.g. enzymes and most proteins)—a procedure specifically selected to produce the best possible result and specified in considerable detail.

2.4 *Selected method.* A technique designed for routine use and, in competition with others, selected as being satisfactory on grounds of accuracy, precision, economy of labor and materials, ease of operation, etc. Eventually the choice of these methods will be made by comparison with results obtained by a reference method, but if this is not available, selection will have to be achieved by any other criteria available.

2.5 *Ascribed value.* The value found for the concentration of a substance in a certain specimen to be used for quality control purposes using a selected method. (If a reference method is used, the term would be "reference method value".)

2.6 *Standard reference material (SRM).* (Sometimes termed "certified reference material" in Europe, a term which will probably replace "standard reference material")—well characterized materials which are certified for specific chemical and/or physical properties, and are for use in calibrating instruments and analytical methods, and in establishing working standards.

The importance of correctly defining terms is shown by the confusion which has already arisen over the use of the word reference. The IFCC Expert Panel on Theory of Reference Values has chosen the word "reference" to replace "normal" in the ambiguous term "normal value" when used in connection with the range of assay values found in "normal" individuals. Any other usage of "reference" must therefore be qualified as in the terms "reference material," "reference method," "reference method value," etc.

When values in the past have had to be ascribed to certain specimens for quality control purposes (i.e. as legally required in the West German Quality Control Program), the technique used has been termed "reference" and considerable confusion has arisen.

3. AREAS WHERE REFERENCE TECHNOLOGY IS REQUIRED

Reference methods are likely to be required for the following five purposes:

3.1 The assay of commercial quality control material when this is to be marketed with reference method values quoted for certain of its constituents.

3.2 The assay, either in prospect or retrospect of material used in national quality control programs. The measurements then being required either for the information of the participating laboratories, or for monitoring the performance of the participating laboratories or the methods being used.

3.3 Together with definitive methods for "one off" measurements required by individual laboratories for development work on instruments and methods.

3.4 Together with definitive methods for monitoring the effects of drugs or their metabolites, etc. upon results obtained using routine methods.

3.5 Definitive methods will be required from time to time to check the performance of new or modified reference methods.

4. THE PRACTICAL APPLICATION OF THE CALCIUM REFERENCE TECHNIQUE

The National Bureau of Standards "reference method" for the determination of calcium in serum[1] has been evaluated at the Clinical Research Centre in London and as is to be expected when reference techniques are tried out world wide, certain problems emerged and solutions were suggested.[2] It was systematically applied by Brown and his colleagues to samples distributed in the Wellcome Group Quality Control Program (Wellcome Reagents Ltd., Beckenham, Kent, England).[3] This is an international control program now covering some 700 laboratories but at the time some 300 laboratories were assaying calcium on specimens distributed.

During 1973, 24 samples were analyzed by the reference method. One object was to estimate the coefficient of variation of the technique when used in the laboratories of the Clinical Research Centre, London. This was calculated from 40 pairs of samples and at 2% agreed with the findings in the test laboratories used by the National Bureau of Standards. (The coefficient of variation of the definitive method at the National Bureau of Standards was approx ±0·2%.) Duplicate samples from batches of sera from the program were assayed by the reference technique at the Clinical Research Centre, and by the definitive method at the National Bureau of Standards. The concordance of results was good, but the reference method showed a slight positive overall bias. This was not unexpected but remains to be explained. The technique is time-consuming and one worker can only carry out approximately six complete assays per week.

The other major point of the exercise was to check the overall accuracy of the results returned by laboratories to the Wellcome Centre. Figure 1 shows the relationship between the reference method results and the mean values obtained by all 300 laboratories. It will be seen that samples of low calcium concentration tended to give high results in the program, whereas samples of high concentration gave low results. This indicated the presence of a concentration-related bias.

In an effort to identify the possible source of this bias, the types of assay used by the different laboratories were investigated. They were found to fall into four main groups: AutoAnalyzer I methods, AutoAnalyzer II (including SMA) methods, atomic absorption methods, and (manual) EDTA methods. Approximately 50 laboratories reported results in each of the four classes, so valid statistical analysis was possible. No bias was found for the AutoAnalyzer I methods. For the AutoAnalyzer II, the results tended to be consistently slightly lower than the corresponding values obtained by the reference methods. The overall concentration-related bias was more pronounced for the two remaining manual methods. The

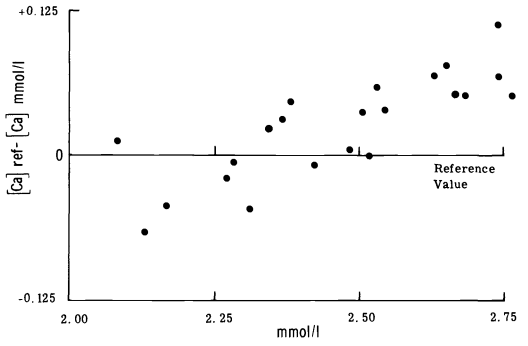

Fig. 1. Deviations of the mean values for calcium found by all laboratories in specimens in the Wellcome Group Quality Control Program, from the reference value.

EDTA technique was found to be particularly bad, and showed larged discrepancies at both high and low calcium concentrations.

Therefore, for the first time, two widely used techniques, as currently carried out, have been shown to be unsatisfactory. This could not have been done without the use of reference technology, and with this knowledge it should now be possible to effect improvement in the international quality control results for the assay of calcium.

5. THE CONTROL OF ERRORS PRODUCED BY INTERFERING SUBSTANCES

A definitive technique for a given substance should ideally measure that substance alone, regardless of any abnormal material which may be present in the matrix. The possible presence of any such material must be borne in mind in choosing and defining the technique, but normally the principles involved should render such caution almost unnecessary. This does not hold however for a reference technique, or a special reference method, the use of which may be perfectly valid for serum specimens used for quality control purposes, but gross interference may occur from drugs or their metabolites or other abnormal substances which may be encountered in specimens being assayed in any hospital laboratory. The problem may be even more serious for a selected method or any technique used routinely. The use of reference technology alone therefore, will not completely control accuracy. It cannot be assumed that specimens of serum or plasma containing abnormal material, or even normal material in abnormal amount, can be correctly assayed by reference methodology, and therefore any specification of a reference method should be accompanied by information on likely sources of interference.

6. REFERENCE TECHNOLOGY FOR SUBSTANCES FOR WHICH NO SRM IS AVAILABLE

The accuracy of definitive and reference methods depends upon the purity of the SRM used for calibration, and when a satisfactory SRM is not available, as for many proteins, or is virtually impossible to obtain, as for most enzymes, definitive and reference techniques as defined, cannot be developed and alternatives must be found. If the yard stick is not available, the means must be specified whereby its equivalent can be reproduced in all circumstances. This means that a technique of measurement must be selected and then defined in considerable detail so that the variation of results found in different laboratories is minimal. As with definitive and reference methods, it may be necessary to have two grades of technique, one so rigorously defined that it is only carried out in special centres (special definitive method) and another more suitable for widespread use (special reference method).

For enzyme measurements there is no alternative to this approach and the position is further complicated by the fact that in many cases a mixture of isoenzymes and not a single enzyme is being measured. This situation causes much of the present interlaboratory variation in enzyme measurement, since different substrates and conditions of temperature, pH, etc. give different overall activities for groups of similar enzymes. In these circumstances the reference technique must have identical conditions to those used in the routine technique, and any difference between the two must be limited to detail of specification, ease of operation, use of automation, etc.

In any technique which is based upon these principles, the importance of specifying universally acceptable units of measurements must be stressed. The usual SI unitage of mmol/l cannot be applied and the alternative for enzymes, is some agreed measure of activity, and for proteins the best solution might be to express their concentration in terms of a protein for which it is possible to prepare a SRM.

7. WORLD AVAILABILITY

So far the major initiative in the development of reference technology has come from the National Bureau of Standards, to whom a debt of gratitude is owed by the clinical chemistry profession. However, they cannot, and should not be asked to, carry the world burden of this work indefinitely. Other centres are working on the development of other techniques but there is a need for international organizations such as the World Health Organization and the International Federation of Clinical Chemistry to take over the distribution of development, and to ensure the availability of materials and methods worldwide. For development work in this field, the cost of equipment and materials and the organization of field testing is high, and progress is likely to be more rapid if the work is distributed between a number of countries, with duplication prevented by the allocation of different assays to different centres by agreement.

Once developed, the techniques need to be kept running, with definitive methods occasionally available and reference methods continuously available in accessible centres. This too is costly but can be financed by those requiring the work to be done. As with the certification of grade "A" glassware, etc. this work is probably best done by government centres, but reliable commercial centres could well have a place. The production of SRM's has been, and probably will remain, the province of government laboratories.

8. CONCLUSIONS

For further progress to be made in the control of quality in clinical chemistry, it is clearly necessary to proceed with an international program for the further development of SRM's and definitive and reference methods, covering hopefully and eventually the complete range of assays required in clinical chemistry. The overall aim for the future must be the achievement for all measurements, of an ability to trace back performance via a reference method to a definitive method. The pioneer work done on calcium points the way to solutions in other areas but it has by no means produced a blue-print. Many problems need to be solved, but hopefully none are unsurmountible.

For the assay of enzymes and other proteins and for radio-immunoassay, the difficulties are formidable indeed, yet in these areas the need for control of accuracy is most urgent.

The importance of the ultimate reference point—a definitive technique—cannot be overstressed. Without it, any reference technology can only be second best.

For the assay of many substances, it may be necessary for many years, to accept selected methods for reference purposes, but it is important that any such temporary measure should be accepted as such, and effort should be continually directed toward seeking a true reference point.

REFERENCES

[1]J. P. Cali, G. N. Bowers, Jr. and D. S. Young, *Clin. Chem.* **19**, 1208 (1973).

[2]J. F. Pickup, M. J. Jackson, E. M. Price and S. S. Brown, *Clin. Chem.* **20**, 1324 (1974).

[3]J. F. Pickup, M. J. Jackson, E. M. Price, M. J. R. Healy and S. S. Brown, *Clin. Chem.* **21**, 1416 (1975).