

Analytical Performance Characteristics Should Be Judged against Objective Quality Specifications

The discipline of clinical chemistry is dynamic. Even a superficial glance at contents pages of the Journal over recent years demonstrates a marked evolution of the field. Appropriately, the Information for authors has also changed over the years to reflect new concepts of what constitutes an acceptable publication.

Readers of the Journal may have noticed an addition to the 1999 Information for authors (1), under the heading *Description of Analytical Methods and Results*, namely:

"Analytical quality. Results obtained for the performance characteristics should be compared objectively to well-documented quality specifications, e.g., published data on the state of the art, performance required by regulatory bodies such as CLIA '88, or recommendations documented by expert professional groups". In addition, quality specifications can be derived from analysis of performance on clinical decision-making.

Many manuscripts deal with the development of new analytical methods or evaluation of commercially available analytical systems. Experimental designs and statistical techniques used to derive data on imprecision and bias are generally more than satisfactory. This is hardly surprising in view of the very many published protocols for the evaluation of methods (2). In contrast, objective analysis of whether the imprecision and bias are satisfactory is often less well done.

This is not a new phenomenon. Although the idea of utilizing quality specifications in assessing the acceptability of method performance was firmly stated in 1974 (3), it was pointed out more than a decade ago that few evaluators actually did compare the performance achieved with preset quality specifications (4, 5). This seems rather difficult to understand because quality specifications based on the state of the art (6), the views of an expert individual (7), and biological variation (8) had been available for many years. Indeed, even a superficial reading of the more recent literature would demonstrate that many papers, reviews, conference proceedings, and book chapters deal with the generation and application of quality specifications (9).

It may be that there are too many contradictory published recommendations, and it might not be easy for authors to select the most appropriate. In addition, new strategies to set quality specifications continue to appear, which might suggest that there is no ubiquitous professional consensus. Moreover, industry does not appear to use professionally set quality specifications as major considerations in either development or marketing.

In spite of these difficulties, there are many quality specifications against which experimental data can be compared, at least for tests reported on ratio (and difference) scales, and we suggest the following approach, which like the types of evidence and grading of recommendations used in clinical practice guidelines (10) can be

placed in a hierarchy of objectivity, the best being first and the worst being last.

ASSESSMENT OF THE EFFECT OF ANALYTICAL PERFORMANCE ON CLINICAL DECISION-MAKING

Quality specifications in specific clinical situations. Ideally, quality specifications should be derived objectively from an analysis of medical needs. Thus, the effect on clinical decision-making of the analytical quality found in any evaluation should be the subject of objective assessment. Unfortunately, this is very difficult. Scandinavian approaches (11) have shown how the calculations can be done for a variety of analytes in a number of different clinical settings. Others, notably Klee (12), have performed very detailed and useful studies using similar approaches. When such clear strategies can be identified, this is probably the best possible approach; however, a major disadvantage is that only a few tests are used in single well-defined clinical situations with standard well-accepted medical strategies directly related to the test result. Another significant drawback is that the quality specifications calculated depend very much on the assumptions made regarding how test results are used by clinicians (13); therefore, they must be related to well-characterized strategies. Quality specification have been derived from questionnaires using vignettes submitted to clinicians on the use of individual tests in specific clinical situations: these studies have very serious flaws, which have been discussed previously by us (9), and we do not recommend their use.

General quality specifications based on medical needs. Consideration of the two major clinical settings in which test results are used, namely monitoring individual patients and diagnosis using reference intervals, shows that generally applicable quality specifications might best be based on the components of biological variation, namely, within-subject (CV_I) and between-subject (CV_G) variation (14).

A very widely held view is that imprecision should be $<0.50 CV_I$ and bias should be $<0.25(CV_I^2 + CV_G^2)^{1/2}$. This strategy has advantages in that data on components of biological variation are easily available on >180 quantities (15). This concept has been expanded recently (16).

For imprecision, although desirable performance is defined as $CV_A < 0.50CV_I$, users of quality specifications based on biology might consider that optimum performance could be defined by $CV_A < 0.25CV_I$, and that the more stringent quality specifications generated using this formula should be used for those quantities for which the desirable performance standards were easily achieved with current technology and methodology, and that minimum performance could be defined by $CV_A < 0.75CV_I$, and that the less stringent quality specifications generated using this formula should be used for those quantities for

which the desirable performance standards were not attainable with current technology and methodology.

An additional basic concept is that, ideally, laboratories throughout a homogeneous population area should use exactly the same reference intervals. This was first proposed by Gowans et al. (17), who showed that for this to be achieved, the bias (B_A) should be $<0.250(CV_I^2 + CV_G^2)^{1/2}$. Analogously to imprecision, this quality specification for bias can be termed desirable performance. However, users of quality specifications based on biology might consider that optimum performance could be defined by $B_A < 0.125 (CV_I^2 + CV_G^2)^{1/2}$, and that the more stringent quality specifications generated using this formula should be used for those quantities for which the desirable performance standards are easily achieved with current technology and methodology, and that minimum performance could be defined by $B_A < 0.375(CV_I^2 + CV_G^2)^{1/2}$, and that the less stringent quality specifications generated using this formula should be used for those quantities for which the desirable performance standards are not attainable with current technology and methodology.

PROFESSIONAL RECOMMENDATIONS

Guidelines from national or international expert groups. A small number of national or international professional groups have proposed detailed quality specifications for imprecision and bias. The recommendations made by the National Cholesterol Education Panel have been used extensively as criteria of acceptability of new methods (18). One European Working Group has proposed quality specifications for use in the evaluation of analytical systems based upon biological variation data as detailed above and the state of the art attained by the best 20% of laboratories (19); these have been widely used in Europe for judging the acceptability of analytical systems for the past few years. Another European Working Group has suggested quality specifications for reference methods when used for validation of routine methods and for assigning values to materials used in external quality assessment schemes (EQAS) or proficiency testing (PT) programs (20). Such guidelines have the major advantage of being based on very extensive laboratory and clinical experience of a number of experts from different backgrounds and with a variety of professional experiences. Moreover, they are usually based on extensive discussion of existing or new scientific theories or experimental data before publication. In addition, the method by which the recommendations were generated is published in the peer-reviewed literature, which allows users to evaluate the objectivity of the process used to reach the conclusions.

Guidelines from expert individuals or institutional groups. Quality specifications have been proposed in a number of sets of published guidelines on what has become termed best practice or good laboratory practice. These are often developed or presented at a single consensus conference without significant discussion. Examples include the recommendations made on analytes used in the assessment

of thyrometabolic status (21) and on therapeutic drug monitoring (22). These guidelines have an advantage in that they usually are based on the very extensive laboratory and clinical experience of an expert or an expert group from a single institution. Although a disadvantage is that these guidelines may be somewhat subjective and not based on scientific theory or experimental data, at least the procedure used to generate the recommendations is published, which again allows users to evaluate the objectivity of the conclusions.

QUALITY SPECIFICATIONS LAID DOWN BY REGULATION OR BY EQAS ORGANIZERS

Quality specifications laid down by regulation. A number of countries have detailed the exact levels of performance required for an analytical technique to be judged as acceptable. The US CLIA '88 legislation (23) documents acceptable total error for a number of commonly assayed analytes, and it is easy to calculate [for example, by using the approaches of Westgard et al. (24)] whether the performance characteristics achieved for a new method meet the criteria to pass. Similar legislation exists in Germany (25), but the quality specifications are very different. The advantage of this approach is that these standards, particularly the CLIA '88 approaches, are well known and understood. However, in contrast to the specifications used in Germany, which are clearly based on the theoretical grounds developed by Stamm (26), a disadvantage of the CLIA '88 quality requirements is that they appear to be based upon the state of the art and therefore reflect what is achievable rather than what is desirable.

Quality specifications laid down by EQAS organizers. EQAS organizers use a variety of measures of location and allowable dispersion. In Europe (27), some countries use statistical analysis of the data returned from the participant laboratories; however, fixed limits are increasingly used. Fixed limits of acceptability are also widely used in other parts of the world. Again, it would be easy to calculate, using the approaches of Westgard et al. (24), whether the analytical performance achieved in an evaluation of a method would allow the laboratory to perform inside or outside the fixed limits of acceptability. The major disadvantage of these quality specifications is that, although often based on expert opinion, they tend to be empirical and are clearly influenced by what is actually achievable at the time.

PUBLISHED DATA ON THE STATE OF THE ART

Published data from external quality assessment and proficiency testing schemes. Comparison of analytical quality could be accomplished through reference to the performance achieved by groups of laboratories participating in EQAS and PT. This has an advantage in that many data are often available. However, the documented analytical performance may not truly reflect the state of the art because the materials used in the challenges may not behave exactly as specimens from patients because of

matrix effects, and participants may adopt special analytical techniques to ensure good performance. In addition, a prerequisite for evaluation of bias is that the materials are as genuine as possible (for example, fresh-frozen human serum) with traceable concentration (or other) values assigned by a reference method or transferred by a reliable method from a reference material (certified when this is available). Moreover, the state of the art inferred from such schemes will change with time (and not always for the better). Furthermore, performance achieved analytically may bear no relationship to actual medical needs.

Published methodology. Comparison may be done by reference to performance documented in original works on similar or other methods for measurement of the analyte(s) under investigation. This also has an advantage in that many data are often available, but has a disadvantage in that the method evaluations are carried out under optimal conditions and the performance documented in the laboratory of the originator or the original evaluator may be the best possible rather than that achieved in practice. Again, performance achieved analytically may bear no relationship to actual medical needs.

Now that the requirement to compare analytical quality to some type of quality specifications has been published in the Information for authors, those preparing their work for submission for publication should follow it. Authors should recognize that it might be difficult to assess the scientific or clinical validity of some of the published quality specifications: it is hoped that in the future those who prepare recommendations, particularly regulators, make their methodology completely transparent. For now, authors should compare their results to quality specifications based on explicitly stated criteria.

References

- Information for authors. *Clin Chem* 1999;45:1–5.
- Haeckel R. Evaluation methods in laboratory medicine. Weinheim, Germany: VCH, 1993:314pp.
- Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974;20:825–33.
- Fraser CG, Singer R. Better laboratory evaluations of instruments and kits are required. *Clin Chem* 1985;31:667–70.
- Hackney JR, Cembrowski GS. Need for improved instrument and kit evaluations. *Am J Clin Pathol* 1986;86:391–3.
- Tonks DB. A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. *Clin Chem* 1963;9:217–33.
- Barnett RN. Medical significance of laboratory results. *Am J Clin Pathol* 1968;50:671–6.
- Cotlove E, Harris EK, Williams G. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. III. Physiological and medical implications. *Clin Chem* 1970;16:1028–32.
- Fraser CG, Hyltoft Petersen P. Desirable standards for laboratory tests if they are to fulfil medical needs. *Clin Chem* 1993;39:1447–55.
- US Department of Health and Human Services. Agency for Health Care Policy and Research. Acute pain management: operative or medical procedures and trauma. Clinical Practice Guideline No. 1. AHCPR Publication No. 92-0023. Rockville, MD: The Agency, 1993:107pp.
- Hyltoft Petersen P, Horder M. Influence of analytical quality on test results. *Scand J Clin Lab Investig* 1992;58(Suppl 208):65–87.
- Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem* 1993;39:1514–8.
- Lytken Larsen M, Hyltoft Petersen P, Fraser CG. A comparison of analytical goals for haemoglobin A1c assays derived using different strategies. *Ann Clin Biochem* 1990;28:272–8.
- Stockl D, Baadenhuisjen H, Fraser CG, Libeer JC, Petersen PH, Ricos C. Desirable routine analytical goals for quantities assayed in serum. Discussion paper from the members of the external quality assessment (EQA) Working Group A on analytical goals in laboratory medicine. *Eur J Clin Chem Clin Biochem* 1995;33:157–69.
- Sebastian-Gambaro MA, Liron-Hernandez FJ, Fuentes-Arderiu X. Intra- and inter-individual biological variability data bank. *Eur J Clin Chem Clin Biochem* 1997;35:845–52. [Also available at www.westgard.com].
- Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8–12.
- Gowans EMS, Hyltoft Petersen P, Blaabjerg O, Horder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Investig* 1988;48:757–64.
- National Cholesterol Education Program Laboratory Standardization Panel. Current status of blood cholesterol measurement in clinical laboratories in the United States. *Clin Chem* 1988;34:193–201.
- Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems in clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992;30:311–7.
- Thienpont L, Franzini C, Kratochvila J, Middle J, Ricos C, Siekmann L, Stockl D. Analytical quality specifications for reference methods and operating specifications for networks of reference laboratories. *Eur J Clin Chem Clin Biochem* 1995;33:949–57.
- 1995 National Academy of Clinical Biochemistry Standards of Laboratory Practice Symposium on Thyroid Testing. *Clin Chem* 1996;42:119–92.
- 1997 National Academy of Clinical Biochemistry Standards of Laboratory Practice Symposium on Therapeutic Drug Monitoring. *Clin Chem* 1998;44:1072–140.
- US Department of Health and Human Services. Medicare, Medicaid and CLIA programs: regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. *Fed Regis* 1992;57:7002–186.
- Westgard JO, Seehafer JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. *Clin Chem* 1994;40:1909–14.
- Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien. *Dtsch Arztebl* 1988;85:A699–712.
- Stamm DA. A new concept for quality control of clinical laboratory investigations in the light of clinical requirements and based on reference methods. *J Clin Chem Clin Biochem* 1982;20:817–22.
- Ricos C, Baadenhuisjen H, Libeer JC, Hyltoft Petersen P, Stockl D, Thienpont L, Fraser CC. External quality assessment: currently used criteria for evaluating performance in European countries, and criteria for future harmonization.

Callum G. Fraser^{1*}
Per Hyltoft Petersen²

¹ Directorate of Biochemical Medicine
Ninewells Hospital and Medical School
Dundee, Scotland DD1 9SY

² Department of Clinical Biochemistry
Odense University Hospital
DK-6000 Odense C, Denmark

*Author for correspondence. Fax 44 (0)1382 645333; e-mail callumf@dh.scot.nhs.uk.