

Prerequisites for Use of Common Reference Intervals

Ferruccio Ceriotti

Diagnostica e Ricerca San Raffaele S.p.A., 20132 Milano, Italy

For correspondence: Dr Ferruccio Ceriotti e-mail: ceriotti.ferruccio@hsr.it

Abstract

The theory of reference values was developed more than 30 years ago, but its application in most clinical laboratories is still incomplete today. This is for several reasons, the most relevant ones being the lack of standardisation of the analytical methods, resulting in method-dependent values, and the difficulty in recruiting the proper number of reference subjects for establishment of reference intervals. With the recent progress in method standardisation the first problem is reducing while the second can be addressed optimally via multicentre collaborative studies that aim to establish common reference intervals. To be effective this approach requires the following prerequisites: 1) the existence of a reference measurement system for the analyte; 2) field methods producing results traceable to the reference system; and 3) a carefully planned multicentre reference interval study. Such a procedure will produce results traceable to the reference measurement system for a large number of reference subjects, under controlled pre-analytical conditions. It will also enable a better understanding of the various sources of population variability, if there is the need for partitioning of a reference interval or if there are any limitations to adopting the established reference intervals on a national or global scale. Once reference intervals are determined, clinical laboratories can adopt a common reference interval provided: 1) the population that the laboratory services is similar to the one studied; 2) methods producing traceable results are used; and 3) analytical quality is within defined targets of precision and bias. Moreover, some validation of the interval using a small sample of reference individuals from the laboratory's population is advisable.

Introduction

The Theory

In Schneider's 1960s paper entitled "Some thoughts on normal, or standard, values in clinical medicine", he states: "... practical medicine is basically founded on comparison. If medicine is to be scientific, we must not only understand the structural, functional and chemical relations operating in individuals, but we must also understand the basis of our comparisons".¹

Today, almost 50 years later, in the age of evidence-based medicine, and in contrast with the enormous developments in the field of medicine, a sound basis for these comparisons is often lacking in the clinical laboratory. Nevertheless, according to Horn and Pesce: "the reference interval is the most widely used medical decision-making tool", even if its practical usefulness is lower than its theoretical power.² This is due to the fact that obtaining a "good" reference interval is a very demanding activity, in terms of time, money and knowledge.

But what is meant by a "good" reference interval? It is an interval that, when applied to the population serviced by

the laboratory, correctly includes most of the subjects with characteristics similar to the reference group and excludes the others. Usually we consider "health-related" reference intervals to mean that the subjects with values within the interval have a lower probability of being affected by a specific disease, while those outside the interval have a higher statistical probability of having the disease or, at least, that the observed value is not normal for a healthy person. The percentage of unhealthy people included in the reference interval or, vice versa, the percentage of healthy subjects outside the interval, defines the "goodness" of the interval.

The factors responsible for this misclassification were already recognised by Schneider who identified the three contributing causes namely, intraindividual, interindividual and analytical variability.¹ Intra- and interindividual variability are inextricably bound. Nevertheless, the relative sizes of these two sources of variation can substantially affect the usefulness of the reference interval as a guide to the status of an individual. Eugene Harris in 1974 demonstrated that only when intraindividual variability (CV_I) is greater than interindividual variability (CV_G), i.e. $CV_I/CV_G > 1.4$, does the distribution of values from any individual cover much

of the reference interval.³ But this is an uncommon situation. In contrast, when CV_I/CV_G is <0.6 , which occurs quite commonly, the dispersion of values for an individual will span only a part of the population-based reference interval. In this case the reference interval will not be sensitive to changes for that individual and, on average, for any individual.⁴

One way to improve the usefulness of reference intervals is to reduce the interindividual variability by partitioning the intervals as much as possible. Stratification by age and gender is the minimum pre-requisite, but other ways include by race, ethnic group, body mass index or nutritional habits (e.g. vegetarians). Herein is the problem of the selection of an appropriate number of reference subjects, properly screened to exclude relevant pathologies, and subdivided by gender, age, race, ethnicity and lifestyle. Diagnostic manufacturers routinely perform reference interval studies for hundreds of analytes using their different assays and platforms and produce method-specific results that may not be comparable across space and time, or equivalent between methods depending upon calibration traceability. Thus, there is difficulty in providing evidence for the existence of authentic and clinically significant differences among races or ethnic groups, even for the most common analytes.

As a shortcut to circumvent these problems several authors have proposed the so called “indirect methods” to define reference intervals.⁵⁻¹⁶ These methods are based on statistical manipulation of existing data to select the “healthy” group from the entire population. There are two main reasons against this approach. Firstly, it is not in keeping with the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) requirements for reference intervals which stipulate that the characteristics of the reference subjects be clearly defined. With the data mining approach very little is known about the subjects and the method relies upon statistical methods to exclude the unhealthy group. Only if you can combine relevant medical history information and analytical data can you fulfil the requirements,^{10,11} but this is an “a posteriori” approach, not an indirect one. Secondly, there is insufficient control of the pre-analytical and analytical conditions. Although the indirect approach can be very useful for local situations or “difficult” groups of subjects like neonates and children, or as a means to confirm the “goodness” of the selected reference interval, it cannot be used to set common reference intervals.

The Practice

Given the enormity of obtaining sufficient numbers of reference subjects, the requirement that each clinical laboratory produces its own reference intervals is a practical impossibility. The effort needed for one single analyte let alone multiplying this hundreds of times for all analytes is

beyond the routine laboratory’s capacity, and to repeat this process continuously to keep the intervals updated with assay and platform changes appears completely unrealistic. The difficulties with the current approach to reference intervals are clearly described in the commentary by Jones et al.¹⁷

To address these difficulties a concerted approach is required by the interested parties who include:

- Diagnostic manufacturers, who must fulfil the legislative requirements and the customers’ needs in a globalised market. The European Directive 98/79/EC on In Vitro Diagnostic medical devices has in the chapter regarding the information supplied by manufacturers (Annex 1), the following requisite: “Where appropriate, the instructions for use must contain ... the reference intervals for the quantities being determined, including a description of the appropriate reference population”;¹⁸
- Clinical laboratory professionals, who want to provide a high quality report that complies with the requests of quality standards like ISO 15189:2003, e.g. paragraph 5.5.5 specifically asks for a periodical review of biological reference intervals;¹⁹
- Clinicians, who use the reference intervals for the management of the patient and would appreciate common intervals among laboratories; and
- Patients themselves, who can be puzzled by reports with different reference intervals from different laboratories in which the same analytical results may be judged normal by one site and abnormal by the other.

Defining common physiological (health-related) reference intervals through a multicentre collaborative experiment can be an efficient way to fulfil the expectations of all stakeholders. This approach overcomes the problem of recruiting large numbers of reference individuals and, at the same time, enables investigation of the influences of race, environmental conditions and life style habits on reference intervals. Sharing the work among multiple sites reduces the cost at each site and the time needed to complete the task.

The main obstacle preventing the adoption of common reference intervals, besides the real existence of local population differences, was (and still is) the lack of comparable values between routine methods due to insufficient standardisation. This is true in particular for the enzymes but also for many other common analytes such as creatinine. To improve the analytical quality it is necessary to establish a sound accuracy base and, to obtain this, a reference measurement system is required.²⁰ Up to now, complete reference measurement systems have been established for only a minority of analytes, but much work is in progress and further standardisation efforts will allow the spread of common reference intervals.²¹

Table. Prerequisites for use of common reference intervals.

Category of Prerequisite	Prerequisite	Responsibility
Analytical	Existence of a Reference Measurement System (accuracy base)	IFCC, JCTLM, National and International Metrology Institutes
	Existence of traceable field methods	Manufacturers
	Correct implementation of the methods in the clinical laboratories	Clinical laboratories
	Control of the performances of the field methods to keep them within stated limits for imprecision and bias	Clinical laboratories, EQAS organisers
Clinical	Accurate definition of reference intervals providing information on the influence of biological and environmental factors	Joint effort – IFCC, manufacturers and clinical laboratories
	Compatible pre-analytical phase	Clinical laboratories
	Validation of the applicability of the common reference intervals to the laboratory's own population	Clinical laboratories
	Adoption of the common reference intervals	Clinical laboratories

IFCC, International Federation of Clinical Chemistry and Laboratory Medicine

JCTLM, Joint Committee for Traceability in Laboratory Medicine

EQAS, External Quality Assessment Scheme

Common Reference Intervals

The Theory

The concept of adopting common reference intervals is simple. If the analytical method is the same or yields identical results because it is correctly standardised, and the population has the same characteristics or it is known that a specific analyte is not significantly influenced by ethnicity or the environment, then common reference intervals can be used.

The Practice

Unfortunately the practical application of this simple concept is not as easy as it would appear. A number of Prerequisites must be in place before adopting common reference intervals (Table).

Establishing Reference Intervals

Assuming that a reference measurement system exists, the most demanding task is the definition of an adequate set of reference values. This should include subjects from different races and ethnic groups and from various environments in order to document if clinically significant differences do really

exist. If they do, then it may be worth while to do partitioning, or else not recommend the application of a common reference interval unless it is wider, which may reduce its utility. The best way to gain this information is to develop a multicentre study involving several clinical laboratories. This approach has been done by a Spanish group and further developed by the Nordic countries.²²⁻³¹

Multicentre Reference Interval Study

A multicentre study for the establishment of common reference intervals must be carefully organised to produce results that can be adopted by any clinical laboratory that operates under similar pre-analytical and analytical conditions. It requires:

- A priori selection of reference subjects according to well-defined criteria, e.g. according to the IFCC and Clinical and Laboratory Standards Institute (CLSI) specifications - inclusion and exclusion criteria for reference subjects must be stated.^{32,33} The number of participating centres and enrolled individuals should be determined according to the number of subjects required for partitioning by

age, gender, race, life style, etc. To obtain sufficiently narrow confidence intervals for the reference limits the optimal number of individuals within each group should be around 500 and the minimal sample size to allow for non-parametric calculation of the confidence limits is 120.^{29,34} The criteria for partitioning are according to Lahti et al.³⁵⁻³⁷

- A clear definition of the pre-analytical phases (according to CLSI C28-A2 or to IFCC specifications) - preparation of the individuals (e.g. fasting, exclusion of physical exercise), the type of sample to collect (serum or plasma), sample tube centrifugation speed and time, time of blood collection, use of a tourniquet, time from collection to analysis or storage and temperature of storage.^{33,38} Ideally, to reproduce the normal analytical flow of the clinical laboratories, the analyses should be performed on fresh samples. However, to reduce the analytical variability it can be useful to freeze the samples and analyse them as one batch or in fewer analytical runs. This procedure can be acceptable only if it has been demonstrated that freezing does not introduce any bias. The storage of one or more aliquots for further use is highly recommended.
- Use of methods providing results traceable to the reference measurement system and high interlaboratory standardisation.
 - a) Standardisation of the analytical procedures - if a reference measurement system exists for the analyte of interest, there is no need for all participating laboratories to use the same analytical system, but an essential pre-requisite is that they all use methods certified by the manufacturers as producing results traceable to the analyte reference measurement system. Method specificity is paramount (e.g. the presence of pyridoxal phosphate for aspartate aminotransferase or alanine aminotransferase measurement, or the type of substrate for α -amylase). Traceability to the reference measurement system must be verified through the use of two or more fully commutable reference materials (frozen pools) with target values assigned by the reference measurement procedure, possibly by a network of reference laboratories.²⁰
 - b) The interlaboratory comparability must be checked via the use of common quality control materials. An internal quality control program must be implemented with clear criteria, defined a priori, for acceptance or rejection of each analytical series.³⁹
- Proper data analysis for the calculation of the reference limits. The data from the different centres must

be compared to identify the presence of analytical bias (from the quality control data) or by an atypical distribution of the reference values. These differences can be tested against the common distribution by use of the non-parametric Mann-Whitney test or Kruskal-Wallis test.^{26,28} Before computing reference limits, outliers should be detected and eliminated using Dixon's test or more sophisticated approaches.^{40,41} Eventually the central 95% of the distribution can be calculated with the simple non-parametric method, applicable to distributions of any form, or with more complex methods requiring data transformation followed by parametric or non-parametric calculations.^{34,42}

Adopting Common Reference Intervals

To be able to apply common reference intervals a clinical laboratory has to verify similarity of the pre-analytical conditions, the analytical method used and its performance, and the characteristics of the population to be serviced.

Pre-analytical Conditions

The reference intervals can be used only if the same pre-analytical conditions are applied (e.g. use of serum, fasting subjects), or if it is possible to demonstrate that the modification has introduced no effects, e.g. analyte levels are not modified by meals, lithium heparin and serum equivalence.

Analytical Aspects

The method in use must produce results traceable to the reference measurement system for the specified analyte. For European countries, if the reagent is 'CE' marked and it is used strictly according to manufacturer's specifications this should be the rule. The European IVD Directive puts traceability as an essential requisite but as demonstrated by the measurement of some enzyme activities, traceability does not necessarily ensure result comparability.^{6,43} The analytical specificity of the method is a key point, especially for the measurement of enzymes. If the specificity of the measurand is not the same as for the reference method, traceability cannot be obtained, e.g. measurement of transaminases in the absence of pyridoxal phosphate or the use of different substrates for α -amylase.

The analytical quality of the method in use should be controlled in order to keep the imprecision and the bias within a stated limit. Targets for maximal imprecision can be derived from the criteria related to biological variability.⁴⁴ According to Gowans et al., the maximum acceptable analytical bias, for use of the same reference interval, is defined as $<0.25 (CV_I^2 + CV_G^2)^{1/2}$ (population-based variability) which is equivalent to the confidence interval of the reference limit for the sample size of 120 individuals.⁴⁵ Gradation of the specifications to 0.125, 0.25 and 0.375 times the population-based variability allows

for optimum, desirable and minimum quality, respectively.⁴⁶ A list of estimated biological within and between-subject variations and analytical quality specifications can be found at Westgard's website.⁴⁷ The presence of a bias and its magnitude can be verified by method comparison on fresh patient's samples or from External Quality Assurance Scheme (EQAS) results or from interlaboratory internal quality control programs.

Characteristics of the Population Served by the Laboratory

These can vary according to the analyte. If race or life habits are known not to significantly influence the reference intervals, it is sufficient to verify the pre-analytical and analytical aspects (e.g. electrolytes).

If race, ethnic groups or life style are known to influence reference intervals or if no information is available on these, it is advisable that the clinical laboratory validates the intervals on a small sample group of its own population. This validation can be done according to the CLSI Document C28-A2, paragraph 8.2.³³ The Document suggests examining 20 reference individuals from a laboratory's own subject population. They should represent a healthy population and satisfy the selection criteria. After discarding any apparent outlier, if no more than 2 of the 20 tested values fall outside the interval, this can be adopted. If three or more fall outside these limits, the experiment should be repeated with another 20 subjects. If this second time no more than 2 of the 20 tested values fall outside the interval, adopt the interval; if again 3 or more fall outside it means that probably the population differs and a specific reference interval is needed (provided that all the pre-analytical and analytical aspects are correct). This binomial test works well when the reference values have a Gaussian-like distribution, but is very insensitive when the distribution is highly skewed (J. Boyd, personal communication). More powerful statistical tests than the binomial test described above can be carried out, e.g. the Kolmogorov-Smirnov test, which compares the full dataset from tested reference individuals with the 20 reference specimens for a given laboratory. Finally a further approach is to calculate the reference intervals from the 20 subjects using a robust algorithm like the one proposed by Horn et al. to check if the obtained limits are within the confidence limits of the common reference interval.⁴²

Discussion and Conclusions

The processes described are not easy, fast or straightforward. The reality is that development of the reference measurement systems and compliance by manufacturers to calibration traceability happen slowly. The time, effort and money required to establish reference intervals are large and clinical laboratories are disinclined to modify reference intervals as

this is a demanding task also requiring education of clinicians and patients. Large multicentre studies are needed for the definition of common reference intervals and are probably the only way to make real progress in this field and bridge the large gap now existing between a very nice theory (IFCC and CLSI documents) and a very poor practice.

The difficulties are related to the need to verify traceability by the distribution of frozen sera and to co-ordinate several centres to perform thousands of tests and enrol hundreds or thousands of individuals, all which have considerable costs. One objection to this is whether such an effort is needed for something that is outdated and will be surpassed by decision limits. The answer is if a common reference interval is established following stringent scientific criteria (both biological and analytical), in theory this is done forever and will not need to be repeated over and over by an infinite number of laboratories around the world. To define a decision limit properly requires that a reference limit is correctly calculated. In fact, excluding the peculiar situation of total cholesterol and lipoproteins, the definition of a decision limit implies the comparison between two populations, namely, the healthy reference population and the unhealthy or ill one (or more depending on the test and on the type of decision limit to be set). Thus reference intervals are only the beginning and once developed work has to be done with patients in various pathological conditions to be able to define cut-offs and decision limits according to the diagnostic sensitivity and the specificity of the test.

The challenge for the future is the development of intraindividual reference values, as already foreseen by Harris.³ This is a concrete possibility today with the progress in information technology and the improvements in analytical standardisation.^{48,49} The computation of individualised reference limits should eventually fulfil Schneider's plea for more reliable means of comparisons with which to judge the health of a patient.

Competing Interests: None declared.

References

1. Schneider AJ. Some thoughts on normal, or standard, values in clinical medicine. *Pediatrics* 1960;26:973-84.
2. Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta* 2003;334:5-23.
3. Harris EK. Effects of intra- and interindividual variation on the appropriate use of normal ranges. *Clin Chem* 1974;20:1535-42.
4. Fraser CG. Inherent biological variation and reference values. *Clin Chem Lab Med* 2004;42:758-64.
5. Martin HF, Hologgita JV, Driscoll J, Fanger H,

- Gudzinowicz BJ. Reference values based on populations accessible to hospitals. In: Gräsbeck R, Alström T, editors. *Reference Values in Laboratory Medicine*. Chichester: Wiley; 1981. p. 233-62.
6. Hoffmann HG. Statistics in the practice of medicine. *JAMA* 1963;185:864-73.
 7. Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. *Biometrics* 1967;23:115-35.
 8. Naus AJ, Borst A, Kuppens PS. The use of patient data for the calculation of reference values for some haematological parameters. *J Clin Chem Clin Biochem* 1980;18:621-5.
 9. Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. *J Clin Chem Clin Biochem* 1985;23:829-39.
 10. Kairisto V, Hänninen KP, Leino A, Pulkki K, Peltola O, Nanto V, et al. Generation of reference values for cardiac enzymes from hospital admission laboratory data. *Eur J Clin Chem Clin Biochem* 1994;32:789-96.
 11. Kouri T, Kairisto V, Virtanen A, Uusipaikka E, Rajamaki A, Finneman H, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem* 1994;40:2209-15.
 12. Krøll J, Saxtrup O. On the use of patient data for the definition of reference intervals in clinical chemistry. *Scand J Clin Lab Invest* 1998;58:469-73.
 13. Bock BJ, Dolan CT, Miller GC, Fitter WF, Hartsell BD, Crowson AN, et al. The data warehouse as a foundation for population-based reference intervals. *Am J Clin Pathol* 2003;120:662-70.
 14. Grossi E, Colombo R, Cavuto S, Franzini C. The REALAB project: a new method for the formulation of reference intervals based on current data. *Clin Chem* 2005;51:1232-40.
 15. Solberg HE. Using a hospitalized population to establish reference intervals: pros and cons. *Clin Chem* 1994;40:2205-6.
 16. Ferré-Masferrer M, Fuentes-Arderiu X, Puchal-Añe R. Indirect reference limits estimated from patients' results by three mathematical procedures. *Clin Chim Acta* 1999;279:97-105.
 17. Jones GR, Barker A, Tate J, Lim CF, Robertson K. The case for common reference intervals. *Clin Biochem Rev* 2004;25:99-104.
 18. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. *Official Journal of the European Communities* 07.12.1998, L331/1-L331/37.
 19. International Organization for Standardization. *Medical laboratories — Particular requirements for quality and competence ISO 15189*. Geneva: ISO; 2003.
 20. Müller MM. Implementation of reference systems in laboratory medicine. *Clin Chem* 2000;46:1907-9.
 21. Bureau International des poids et mesures, JCTLM database: Laboratory medicine and in vitro diagnostics. <http://www.bipm.org/jctlm/> (Accessed 21 April 2007).
 22. Fuentes-Arderiu X, Ferré-Masferrer M, Gonzales-Alba JM, Escola-Aliberas J, Balsells-Rosello D, Blanco-Cristobal C, et al. Multicentric reference values for some quantities measured with Tina-Quant reagents systems and RD/Hitachi analysers. *Scand J Clin Lab Invest* 2001;61:273-6.
 23. Ferré-Masferrer M, Fuentes-Arderiu X, Alvares Funes V, Güell-Miró R, Castiñeiras-Lacambra MJ. Multicentric reference values: shared reference limits. *Eur J Clin Chem Clin Biochem* 1997;35:715-8.
 24. Ferré-Masferrer M, Fuentes-Arderiu X, Gomà-Llongueras M, Aluma-Trullas A, Aramendi-Ramos M, Castano-Vidriales JL, et al. Regional reference values for some quantities measured with the ADVIA Centaur analyser. A model of co-operation between the in vitro diagnostic industry and clinical laboratories. *Clin Chem Lab Med* 2001;39:166-9.
 25. Fuentes-Arderiu X, Ferré-Masferrer M, González-Alba JM, Villarino-Gonzalez MI, Arrimadas-Esteban E, Cabrero-Olive D, et al. Multicentric reference values for some quantities measured with the Elecsys 2010 analyser. *Clin Chim Acta* 2001;304:143-6.
 26. Fuentes-Arderiu X, Mas-Serra R, Alumá-Trullás A, Martí-Marcet MI, Dot-Bach D. Guideline for the production of multicentre physiological reference values using the same measurement system. A proposal of the Catalan Association for Clinical Laboratory Sciences. *Clin Chem Lab Med* 2004;42:778-82.
 27. Bäck SE, Nilsson JE, Fex G, Jeppsson JO, Rosén U, Tryding N, et al. Towards common reference intervals in clinical chemistry. An attempt at harmonization between three hospital laboratories in Skåne, Sweden. *Clin Chem Lab Med* 1999;37:573-92.
 28. Rustad P, Felding P, Lahti A. Proposal for guidelines to establish common biological reference intervals in large geographical areas for biochemical quantities measured frequently in serum and plasma. *Clin Chem Lab Med* 2004;42:783-91.
 29. Hyltoft Petersen P, Rustad P. Prerequisites for establishing common reference intervals. *Scand J Clin Lab Invest* 2004;64:285-92.
 30. Rustad P, Felding P, Franzson I, Kairisto V, Lahti A, Martensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. *Scand J Clin Lab Invest* 2004;64:271-84.

31. Rustad P, Felding P, Lahti A, Hyltoft Petersen P. Descriptive analytical data and consequences for calculation of common reference intervals in the Nordic Reference Interval Project 2000. *Scand J Clin Lab Invest* 2004;64:343-70.
32. PetitClerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987;25:639-44.
33. CLSI. How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline—Second Edition. CLSI document C28-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2000.
34. Solberg HE. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values (EPTRV) and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem* 1987;25:645-56; *Clin Chim Acta* 1987;170:S13-S32.
35. Lahti A, Hyltoft Petersen P, Boyd JC, Fraser CG, Jørgensen N. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. *Clin Chem* 2002;48:338-52.
36. Lahti A, Hyltoft Petersen P, Boyd JC. Impact of subgroup prevalence on partitioning of Gaussian distributed reference values. *Clin Chem* 2002; 48:1987-99.
37. Lahti A, Hyltoft Petersen P, Boyd JC, Rustad P, Laake P, Solberg HE. Partitioning of nongaussian-distributed biochemical reference data into subgroups. *Clin Chem* 2004;50:891-900.
38. Solberg HE, PetitClerc C. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values (EPTRV). Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem* 1988;26:593-8; *Clin Chim Acta* 1988;177:S1-S12.
39. Westgard JO. Design of internal quality control for reference value studies. *Clin Chem Lab Med* 2004;42:863-7.
40. Dixon WJ. Processing data for outliers. *Biometrics* 1953;9:74-89.
41. Horn PS, Pesce AJ. Reference Intervals a user's guide. Washington DC: AACC Press; 2005. p. 29-36.
42. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem* 1998;44:622-31.
43. Jansen R, Schumann G, Baadenhuijsen H, Franck P, Franzini C, Kruse R, et al. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum: an international study in the EC4 framework of the Calibration 2000 project. *Clin Chim Acta* 2006;368:160-7.
44. Cotlove E, Harris EK, Williams GZ. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. 3. Physiological and Medical Implications. *Clin Chem* 1970;16: 1028-32.
45. Gowans EM, Hyltoft Petersen P, Blaabjerg O, Hørdler M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988;48:757-64.
46. Fraser CG, Hyltoft Petersen P, Libeer J-C, Ricós C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34: 8-12.
47. Westgard QC: Desirable specifications for total error, imprecision, and bias, derived from biological variation. <http://www.westgard.com/biodatabase1.htm> (Accessed 22 April 2007).
48. Harris EK. Some theory of reference values. II. Comparison of some statistical models of intraindividual variation in blood constituents. *Clin Chem* 1976;22: 1343-50.
49. Queralto JM. Intraindividual reference values. *Clin Chem Lab Med* 2004;42:765-77.