

# Prerequisites for establishing common reference intervals

P. HYLTOFT PETERSEN\*† & P. RUSTAD‡

\*Department of Clinical Biochemistry, Odense University Hospital, Odense, Denmark;  
†NOKLUS, Norwegian Centre for External Quality Assurance of Primary Care Laboratories,  
Division for General Practice, University of Bergen, Norway; ‡Først Medical Laboratory, Oslo,  
Norway

Hyltoft Petersen P, Rustad P. Prerequisites for establishing common reference intervals. Scand J Clin Lab Invest 2004; 64: 285–292.

Establishment of common reference intervals for homogeneous populations within regions is based on the same basic principles as the IFCC recommendations for individual laboratories, but a few additional prerequisites are needed. Thus, the need for common standardization and traceability during production of the reference values and with the application of the common reference intervals in the laboratories becomes crucial. Furthermore, the external control system must be geared to the purpose, using matrix-correct control materials with concentration values traceable to the same reference methods, and validation of results according to analytical quality specifications designed for the use of common reference intervals. Here, the standards may have a restrictive influence on the establishing of common reference intervals, with their demands for the use of the producers' traceability, instead of a relevant high-quality reference preparation shared by all the participants. Two main strategies for measurements are analysis immediately after the sampling, and storage of samples until analysis in one or a few analytical runs. The former strategy needs constant standardization and stability of the performance in many laboratories and in several analytical runs, resulting in between-run variation, whereas the latter precludes this between-run variation, but makes demands on the stability of the components under storage. When a considerable number of laboratories decide to establish common reference intervals, it is possible to obtain large sample sizes of reference values, which reduces the confidence intervals around the reference limits. It also makes it possible to collect samples from many subgroups, such as racial groups and groups related to different environmental conditions, as well as the traditional groupings according to age and gender, pregnancy and use of oestrogens. If all these subgroups are large, e.g.  $n > 500$ , the confidence limits will be small and criteria for partitioning can be applied. Choosing reference individuals is not easy, as definitions of health, as well as rule-in and rule-out criteria vary from one investigation to the other. Therefore, the strategy and the criteria must be thoroughly described. Arguments for establishing common reference intervals are not needed. On the contrary, lack of such common reference intervals should be explained.

*Key words:* Common control; common standardization; homogeneous populations; partitioning according to subgroups; statistical evaluations

*Per Hyltoft Petersen, Department of Clinical Biochemistry, Odense University Hospital, DK-5000 Odense C, Denmark. Tel. +45 65 412 836, fax. +45 65 411 911, e-mail. per.hyltoft.petersen@ouh.fyns-amt.dk*

## INTRODUCTION

In a series of articles on reference values and reference intervals, the IFCC outlined recommendations for establishment of reference intervals in the individual clinical chemical laboratory [1–6]. In this series, for example, reference values and reference intervals are defined, and the question of health is discussed [1], together with further details on the selection of reference individuals and exclusion criteria as well as partitioning according to subgroups related to age and gender [2]. Biological factors (preparation of the individuals, etc.) and methodological factors (specimen collection and handling) are also described [3]. The most comprehensive recommendation of the series consists of a detailed description of the procedures for calculation of reference limits, dealing with both parametric and non-parametric estimation of these limits [5], whereas the last paper deals with the presentation of reference values and classification of observed values [6]. The fourth part of the series [4] describes “Control of analytical variation in the production and application of reference values”, concentrating on the internal quality control (IQC) and external quality assessment (EQA), while the section on transferability of reference values between laboratories is very short, and the main impression is that the IFCC recommends that each laboratory should produce its own reference intervals.

Production of laboratory-specific reference intervals, however, is a laborious and costly process if all recommendations are to be followed [7], so laboratories often borrow reference intervals from other laboratories or produce reference intervals from small sample sizes of less well-described reference individuals, if they do not use medical students.

Since the publication of these IFCC recommendations, there has been increasing interest in production of common reference intervals for homogeneous populations to be shared by laboratories within a region, because clinicians and other physicians as well as the patients do

not understand that reference intervals tend to vary among laboratories, and that measurement results from different laboratories can be different for the same blood specimens. Thus, from Spain, two publications on “multicentric reference values” illustrate that it is possible to establish common reference intervals for a whole country for laboratories using the same equipments and reagents [7, 8]. A project on common reference intervals for specific plasma proteins based on the Certified Reference Material (CRM 470) covered the Nordic countries [9], and a Swedish project on common reference intervals was designed for a county [10], whereas a Japanese project covered a whole prefecture [11], and now there is “the Nordic project on common reference intervals for 25 common analytes” [12–14]. Furthermore, some of the prerequisites for determining racial and environmental similarities and differences for plasma proteins have been published [15].

The new concept of sharing common reference intervals for a homogeneous population within a region has added new aspects and new prerequisites to the IFCC recommendations and, consequently, this contribution deals with these new prerequisites in relation to the IFCC recommendations.

## SAMPLE SIZE IN RELATION TO COMMON REFERENCE INTERVALS AND PARTITIONING

When establishing common reference intervals, the sample size can be expanded considerably compared to production of local reference intervals in each individual laboratory. When many laboratories can share the common reference intervals, the investment for each laboratory is limited or the whole work can be concentrated in one or a few institutions, which may be funded. Consequently, large sample sizes of 500, or preferably more reference individuals can be obtained [9, 11–15]. A large sample makes it relevant to perform a detailed investigation of subgroups where it now is possible to obtain

reliable estimates of the subgroup reference intervals, with reasonably small confidence intervals for the reference limits, as each subgroup will at least fulfil the minimum size of 120 recommended by the IFCC [5]. The 90% confidence interval (CI) for a sample size of  $N=120$  is, however,  $\pm 0.24*s$  (or more generally  $0.25*s$ ) where  $s$  is the population standard deviation.

The criteria for partitioning are according to Lahti *et al.* [16]:

- A. If one or both differences between the lower reference limits and difference between upper reference limits of two subgroups are  $>0.75*s_{\text{lowest}}$ , where  $s_{\text{lowest}}$  is the smallest  $s$  of the subgroup standard deviations, partitioning is recommended.
- B. If both these differences between lower reference limits and higher reference limits of two subgroups are  $\leq 0.25*s_{\text{lowest}}$ , partitioning is not recommended.
- C. For differences between the two extremes ( $0.25*s_{\text{lowest}} < \text{difference} < 0.75*s_{\text{lowest}}$ ), other than statistical reasoning is to be considered. This could be due to genetic differences, which are not investigated routinely, whether it is easy to distinguish between, for example, racial subgroups, based on information from the literature, clinical importance, etc.

Criterion A. can also be based on the fraction of reference individuals from each subgroup and must not exceed the interval of 0.9 to 4.1% of the population when the common reference limit is applied. When a prevalence of a subgroup, different from 0.5, is taken into consideration, the recommendations for partitioning become more complicated [17].

In order to keep the confidence intervals of reference limits small compared to the partitioning criteria, each subgroup should be of a size of at least 500, whereby the parametric criterion for partitioning becomes reasonably large compared to the 90% CI.

## CHOICE OF REFERENCE POPULATION

The question of how to choose reference individuals cannot be answered in general terms, and

it often depends on the purpose of the investigation and the possibilities of finding these reference individuals. The most common way is to establish reference values from a healthy population, but even here, the definition of the state of health is problematic. A number of strategies can be reported:

- a. Selection of homogeneous reference groups according to racial, ethnic, geographical and environmental conditions, aiming at being representative of the population to which it is to be applied
- b. Stratification according to age and gender, menstrual cycle, pregnancy and oestrogen supplement
- c. Definition of state of health, i.e. rule-in and rule-out criteria according to stated requirements; e.g. allowable medication, drug abuse, measured concentrations of other quantities indicating chronic or acute diseases, sometimes with a medical examination

Examples are Harris *et al.* [18], who applied criteria for partitioning to a population of Hispanics, Asians, Whites and Blacks for serum creatine kinase, with the interesting result of three reference intervals. A high interval for Black men, an intermediate for Hispanic, Asian and White men together with Black women, and a low interval for Asian, Hispanic and White women. Jørgensen *et al.* [19] invited 2100 people according to the "local personal identification register" for measurements of fasting plasma glucose, and 755 responded. Of these, 29 were diseased and of the remaining 726 participants a large number were at risk for diabetes, leaving only 424 as a "low-risk" population. In a very painstaking investigation involving 255 individuals participating in a longitudinal reference interval study, starting in 1963, all reference individuals underwent medical examinations by a specialist in internal medicine [10]. In a comparable investigation on plasma proteins based on 999 reference individuals, who were first selected through a medical examination, but afterwards 25 usual clinical chemical quantities were measured and based on a 99% criterion of these reference intervals, only individuals without values outside these limits were accepted for estimation of reference intervals for the plasma proteins [20]. In another investigation on plasma proteins [9], the reference

individuals were selected among the laboratory personal and their relatives, and with this group only 37 out of the 553 participants had to be ruled out. This is very much like the Nordic project on reference intervals for 25 of the most frequently used properties in clinical chemistry [12–14]. In this project variation between countries is also under consideration.

These examples illustrate different strategies for inviting reference individuals, and at the same time indicate that invitation according to a random selection from, for example, social security numbers leads to the exclusion of a large proportion of the participants, whereas selection of hospital personal and their relatives make the exclusion more moderate. There are no recommendations on which selection method is the more appropriate, and this may depend on the purpose of the investigation and the possibilities for recruiting reference individuals. It is important, however, to describe the strategy and the individuals and to apply clear criteria for rule in and rule out.

## PRE-ANALYTICAL CONDITIONS

The pre-analytical conditions can be divided into preparation of the reference individual before sampling, the conditions during sampling, the tubes and handling of samples and the storage. Guder *et al.* have written an excellent book on the subject [21] and a booklet on quality of samples [22].

### *Preparation of the reference individuals*

This preparation is mainly in the form of restrictions such as fasting/usual intake of food, no exhausting exercise and no alcohol and smoking for some time before the sampling, etc.

### *Sampling procedure*

Usually the sampling procedure should be performed in accordance with the routine of the laboratory. For blood sampling, the general condition is to sit in a relaxed position for 15 min before the sampling is taken in an arm vein, using vacuum tubes, but sampling of capillary blood from the finger may be relevant, and for measurements of coagulation analyses, the first glass must be discharged. In this context

the use of anticoagulation and separator gels may be crucial for the measurement results.

### *Handling of samples*

Blood for production of serum must be allowed to coagulate before centrifugation, but tubes with anticoagulants can be handled immediately, and for quantities like plasma glucose, instant cooling is necessary.

### *Storage of samples*

In contrast to the routine measurements, which are performed shortly after sampling in the laboratory, the production of reference values is usually performed after storage of the samples until measurements, which are often done in one single or a few analytical runs. This storage must not have any influence on the measurement results, if the reference values are to represent the population and are to be useful for calculation of reference intervals.

## ANALYTICAL QUALITY

The standardization of the analytical procedures and the traceability of the measurement values are crucial for establishing common reference intervals, as any analytical bias will result in disclosure of the estimated reference interval and, thereby, disagreement between the reference interval and the population to which it is applied. Increasing bias will result in increasing deviations in the assumed 2½% of the reference population outside each reference limit. The standardization needed for establishing common reference intervals must therefore be based on reference methods with tight traceabilities to the calibrators in use via matrix-correct materials in the chain or by use of split samples, i.e. patient samples measured by both the reference method and the field method. If not possible, the standardization should be based on a matrix-correct reference material, which could be a non-processed liquid frozen serum pool for serum analytes.

Two strategies for measurements in a project on common reference intervals are (i) to measure as soon as possible, which requires strict standardization of the methods in use, or (ii) measurement in a single/few analytical run(s)

after storage of all samples until the measurement in a single laboratory. The latter strategy may introduce a systematic error owing to the actual calibration but at the same time reduces the between-run analytical imprecision. Here the stability during storage is vital.

## STATISTICS

The statistical estimations of reference intervals for common reference intervals are based on the same statistical principles for parametric as well as for non-parametric evaluations [5] and now also by using the non-parametric estimation according to the bootstrap-based procedures [23], whereas the possibilities for collection of large samples make the subgrouping relevant with partitioning according to well-defined criteria [16, 17].

## ANALYTICAL QUALITY SPECIFICATIONS

When the common reference intervals are established according to the best analytical standardization, it is important to maintain the quality in all laboratories using the common reference intervals. Analytical quality specifications for this purpose are outlined for gaussian distributions [24] as well as for log-gaussian distributions [25], based on the idea of large sample sizes to reduce the confidence intervals around the reference limits, and if this, instead, is used to allow for analytical deviations in the user laboratory, each laboratory could obtain the same quality of reference interval by establishing its own according to the IFCC [5], and then all laboratories would have the same reference intervals. The first analytical quality specifications [24] were that a combination of bias and imprecision (in relation to the produced reference intervals) should not result in the percentage of reference values outside each reference limit exceeding the interval 1.4 to 4.4%. This corresponds to 0.25 times the population-based standard deviation, which is equivalent to the confidence interval for the sample size of 120, which is the minimum recommended by the IFCC [5]. The specifications were later graduated for the bias term to 0.125, 0.25 and 0.375 times the population-based

standard deviation for optimum, desirable and minimum quality, respectively [26]. To obtain the combined specifications for bias and imprecision, the calculations are more complicated, but the different combinations of analytical quality specifications can be illustrated graphically [27]. A list of estimated biological within- and between-subject variations and analytical quality specifications can be found at Westgard's web side [28, 29].

## EXTERNAL CONTROL

For regions with common reference intervals, the external control should be related to this purpose and based on the same standardization as that used during the establishment of the common reference intervals and with acceptance limits according to the analytical quality specifications for sharing the common reference intervals. External control related to the establishment of common reference intervals, and using the analytical quality specifications for possible recommendation of use of the common reference intervals has been performed [9].

## STANDARDS

The reference intervals are also mentioned in the EU Directive 98/79/EC (Annex I 8.7) and the EN/ISO 15189 (5.5.5), but where the Directive writes that the producers of kits and reagents are responsible for the reference intervals, the ISO-standard holds the individual laboratories responsible [30]. This discordance is not readily resolved, as the Directive demands the companies to direct their reference intervals to all areas and races and at the same time makes it difficult for the laboratories to obtain reliable reference intervals, as the standardization is also the producer's responsibility. This makes it extremely difficult to apply the common reference intervals to laboratories, which are not able to establish the same standardization as that under which the common reference intervals are produced, owing to differences in methods and calibrators from the various producers. The directive does not make it easy to establish the common standardization needed for production and use of common reference intervals.

In reality, the standards are very demanding, without giving relevant advice for the reference intervals, and they do not reflect the concept of common reference intervals.

## GENERAL DISCUSSION

It should not be necessary to argue for establishing common reference intervals for homogeneous populations and within regions like the Nordic countries. On the contrary, arguments for not having common reference intervals should be demanded.

The prerequisites for establishing common reference intervals are basically the same as those for production of reference intervals in the individual laboratory [1–6], but a number of prerequisites need to be added. These are related to the standardization needed for harmonizing the analytical measurements in the region to use the common reference intervals and to the quality of the external control needed to monitor the current analytical quality with reliable control materials and with traceable concentration values where possible. Furthermore, this control should be performed according to the analytical quality specifications for common reference intervals.

With cooperation between all laboratories in a region, it should be possible to obtain large sample sizes, in order to gain confidence about the estimated limits and in order to investigate variations among defined subgroups according to age and gender as well as race and environmental conditions, etc.

It would be relevant to produce a new standard based on the IFCC recommendations and with the extra requirements for establishing common reference intervals, which could overrule the conflicting and insufficiently available standards.

## CONCLUSIONS

Special prerequisites for the establishing of common reference intervals are:

1. Large sample sizes. Each subgroup should have a sample size of > 500.
2. Partitioning criteria based on the acceptable

fractions of reference individuals outside the reference limits.

3. Analytical quality specifications for establishing common reference intervals.
4. Common standardization and common external control according to the analytical quality specifications.
5. Standards that not render the establishment of common reference intervals difficult.

## REFERENCES

- 1 Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem* 1987; 25: 337–42.
- 2 PetitClerc C, Solberg HE. 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.
- 3 Solberg HE, PetitClerc C. 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.
- 4 Solberg HE, Stamm D. Approved recommendation on the theory on reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. *Eur J Clin Chem Clin Biochem* 1991; 29: 531–5.
- 5 Solberg HE. 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.
- 6 Dybkær R, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. *J Clin Chem Clin Biochem* 1987; 25: 657–62.
- 7 Fuentes-Arderiu X, Ferré-Masferrer M, Gonzales-Alba JM, Escola-Aliberas J, Balsells-Rosello D, Blanco-Cristobal C. 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.
- 8 Ferré-Masferrer M, Fuentes-Arderiu X, Alvares Funes V, Güel-Miró R, CastiñeirasLacambra MJ. Multicentric reference values: shared reference limits. *Eur J Clin Chem Clin Biochem* 1997; 35: 715–8.
- 9 Hyltoft Petersen P, Blaabjerg O, Irjala K, editors. Assessing quality in measurements of plasma proteins. The Nordic Protein Project and related projects. *Upsala J Med Sci* 1994; 99: 195–389.
- 10 Bäck S-E, Nilsson J-E, Fex G, Jeppsson J-O, Rosén U, Tryding N. 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.
- 11 Kinoshita S, Toyofuku M, Iida H, Wakiyama M, Kurihara M, Nakahara M. Standardization of laboratory data and establishment of reference intervals in the Fukuoka prefecture: a Japanese perspective. *Clin Chem Lab Med* 2001; 39: 256–62.
  - 12 Rustad P. Reference intervals for 25 of the most frequently used properties in clinical chemistry. Proposal by Nordic reference interval project (NORIP). *Klinisk Biokemi I Norden* 2003; 2: 10–7.
  - 13 <http://www.furst.no/norip/>
  - 14 Mårtensson A. Reference intervals for haematology analytes. Preliminary results from the Nordic reference interval project (NORIP). *Klinisk Biokemi I Norden* 2003; 2: 20–1.
  - 15 Hyltoft Petersen P, Whitcher JT, Myron Johnson A, Itoh Y, Carlström A. on behalf of the International Federation of Clinical Chemistry and Laboratory Medicine Committee on Plasma Proteins. Strategy for determining racial and environmental similarities and differences for plasma proteins. *Clin Chem Lab Med* 2001; 39: 1146–53.
  - 16 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.
  - 17 Lahti A, Hyltoft Petersen P, Boyd J. Impact of subgroup prevalence on partitioning of Gaussian-distributed reference values. *Clin Chem* 2002; 48: 1987–99.
  - 18 Harris EK, Wong ET, Shaw ST. Statistical criteria for separate reference intervals: race and gender groups in creatine kinase. *Clin Chem* 1991; 37: 1580–2.
  - 19 Jørgensen LGM, Stahl M, Brandslund I, Hyltoft Petersen P, Borch-Johnsen K, de Fine Olivarius N. Plasma glucose reference interval in a low-risk population. 2. Impact of the new WHO and ADA recommendations on diagnosis of diabetes mellitus. *Scand J Clin Lab Invest* 2001; 61: 181–90.
  - 20 Ichihara K, Kawai T. Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM 470) and NCCLS proposed guideline (C28-P, 1992): trial to select reference individuals by results of screening tests and application of maximal likelihood method. *J Clin Lab Analysis* 1996; 10: 110–7.
  - 21 Guder WG, Narayanan S, Wisser H, Zawta B. Samples: from the patient to the laboratory. The impact of preanalytical variables on the quality of laboratory results. 2nd ed. Germany: Git Verlag GMBH; 2001.
  - 22 Guder WG, da Fonseca-Wollheim F, Heil W, Töpfer G, Wisser H, Zawta B. Die Qualität diagnostischer Proben. Deutschen Gesellschaft für Klinische Chemie und Deutschen Gesellschaft für Laboratoriumsmedizin. Dritte Auflage 2002.
  - 23 Linnet K. Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clin Chem* 2000; 46: 867–9.
  - 24 Gowans EMS, Hyltoft Petersen P, Blaabjerg O, Hørder 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.
  - 25 Hyltoft Petersen P, Gowans EMS, Blaabjerg O, Hørder M. Analytical goals for the estimation of non-Gaussian reference intervals. *Scand J Clin Lab Invest* 1989; 49: 727–37.
  - 26 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.
  - 27 Hyltoft Petersen P, Fraser CG, Jørgensen L, Brandslund I, Stahl M, Gowans EMS, Libeer J-C, Ricós C. Combination of analytical quality specifications based on biological within- and between-subject variation. *Ann Clin Biochem* 2002; 39: 543–50.
  - 28 <http://www.westgard.com/intra-inter.htm>
  - 29 <http://www.westgard.com/biodatabase2.htm>
  - 30 Hyltoft Petersen P. Report from the second European symposium on clinical laboratory and in vitro diagnostic industry. “Physiological reference values: A shared business?”. *Clin Chem Lab Med* 2003; 41: 825–8.

Received: 11 November 2003

Accepted: 10 March 2004