Editorial

Joseph Henny

Multicenter reference intervals studies: a promising perspective for the future?

Up to now the establishment of reference limits has been based on the recommendations of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC-LM) which give greater place to determining reference intervals (RIs) in each laboratory [1]. In fact, the situation is complex: regulations (e.g., European Directive 98/79/EC [2] and the new revised version pending approval) require that manufacturers mention RIs in the package inserts (for all countries). Clinical laboratories are responsible for RIs printed on laboratory reports (ISO 15189) [3]. Laboratories should also verify them if they use RIs from literature or manufacturers. Also, producing RIs is actually too expensive and a heavy task for most laboratories. Alternatives have been proposed, such as the determination of RIs from databases from each laboratory (data mining) [4], but relevant clinical information, needed to select healthy individuals, is rarely available. That is why the IFCC/CLSI remains reserved on this method. The production of common RIs for homogeneous populations shared by all laboratories within a country or a region has attracted a lot of interest. In Spain, the Catalan Association for Clinical Laboratory Science has demonstrated that it is possible to establish common RIs for an entire country for laboratories using the same equipment and reagents, with ensured metrological traceability [5]. Almost simultaneously, a Nordic European Group proposed a new concept of shared reference values based on the recommendations of the IFCC-LM, but specifying prerequisites [6]. This protocol is based on the need for a common standardization and traceability throughout the production phase of the reference values. A process of external quality control using commutable materials (no matrix effect) and traceable to the same reference method is implemented. The analytical specifications will be defined beforehand. These authors point out that there are no recommendations for the selection of reference individuals, these being dependent on the study objectives. However, if the subgroups are large enough (>500) the partitioning criteria can be applied and the confidence limits for RIs will be small. This protocol applied to a group of laboratories in several Scandinavian countries has shown its effectiveness. On behalf of the IFCC-LM, Ceriotti [7, 8] completed these prerequisites to determine common RIs: 1) the population that the laboratory services is similar to the one studied; 2) careful study design; 3) use of traceable analytical methods; 4) predefined analytical goals; 5) adequate statistical treatment.

In this issue, Ichihara et al. present two very ambitious articles [9, 10] attempting to provide answers to many questions: 1) observe possible differences between the populations studied in seven regions of East and South-East Asia and in seven regions of Japan; 2) assess the possible need for partitioning RIs; 3) derive common RIs for standardized analytes (traceable to a reference method); 4) propose a method of transference of RIs for not standardized analytes (by comparison of methods: cross-check testing); 5) derive RIs for not standardized analytes.

In the first study [9] the authors determined RIs for 72 analytes traceable to a reference system. As in their previous study [11] they limited the component of inter-laboratory variability by centralizing analyzes of deep-frozen sera of volunteers from 63 participating laboratories (in Japan and South-East Asia) in a single laboratory in Japan. Significant regional differences were identified for 11 analytes.

In the second study [10] volunteers were recruited from 48 laboratories in 14 cities in Japan and 15 laboratories in nine cities in South-East Asia. Analyses of deep-frozen sera of patients were centralized in a single laboratory in Japan, following the same protocol as in the first study. Simultaneously a comparison of methods study (comparing methods used in the central laboratory versus methods used in each laboratory) was conducted on a limited number of serum samples. As a result, the analyses of these samples were carried out simultaneously in each laboratory and in the central Japanese laboratory. Thanks to the cross-check testing method, the authors claimed that 74% of RIs determined from the central laboratory database may be transferred to each participating laboratory. In addition biological and related regional changes were observed for several analytes.

The study design was similar in both studies; it is certainly one of the highlights of the study: careful
management of the study, standardization of all steps as complete as possible (to minimize the inter-laboratory specimen-handling effects), centralization of measures (to avoid inter-laboratory variability), external assessment scheme using commutable control sera to avoid the matrix effect (use of fresh-frozen pooled sera from healthy individuals).

The analytical quality of the data represents the second major strength of the first study [9]. The verification of the traceability to a reference system ensures that RIs may be used by another laboratory using a similar traceable method. In this case, the determination of common RIs makes sense. It must also be remembered that traceability does nothing for poorly specific routine methods. Development by manufacturers and use by clinical laboratories of traceable methods should be encouraged. However, we must be realistic. For the moment, the number of methods traceable to a reference system is limited (about 70), but it takes into account most of the current clinical chemistry tests.

For some laboratory tests, for which there is no traceable methods to a reference system, alternative such as transference of RIs between analytical systems should be implemented. To compare analytical systems, the investigators propose in the second study [10] a cross-comparison of test results method. The principle is similar to the method recommended by the IFCC/CLSI (C28 A3) [1] for the transference of RIs, but the detailed interpretation of the results are original. Ichihara et al. used the standard error of the slope “b”, expressed in coefficient of variation of the slope to assess the convertibility of RIs derived from a central laboratory to participating laboratories (each of them using their own analytical system). For most laboratory tests and analytical systems studied, this method gives good results (rate of transferability close to 80%) for a cut-off value of CV (b)=10%. This 10% threshold was arbitrarily determined by consensus among all participating laboratories, after considering all the scattergrams. This method appears promising to transfer RIs produced in a central laboratory to other laboratories using different analytical methods for most non-standardized laboratory tests. However, several limitations should be considered, some of which have been previously pointed out by the IFCC/CLSI [1]: 1) the use of appropriate distribution of values is critical. If not, we may over- or underestimate the quality of the correlation. When the physiological range is narrow, linear regression is not the most appropriate to assess the transferability of RIs, as observed by the authors for the electrolytes and albumin; 2) consider the magnitude of the intercept in comparison of the range of the data is useful. It must be much smaller than the RI, although in the case of this article, the authors claimed that this condition is fulfilled; 3) this method of comparison of analytical systems is purely based on statistical methods. It assumes that populations of each laboratory are homogeneous and comparable. The issue of transference of RIs is not simply a matter of comparing analytical systems, but also comparing population.

The process of selecting healthy volunteers a priori is highly relevant. The exclusion criteria are clearly defined as recommended by the IFCC-LM guidelines [1], the Nordic group [6] and Ceriotti [8]. However, the healthy reference population studied was limited to hospital workers. We understand that conducting such a huge study involving a large number of laboratories should meet feasibility criteria, one of the most crucial is the ability of recruitment of healthy volunteers. However, the use of common RIs by clinical laboratories assumes that populations are comparable (population of the reference laboratory having determined the RIs vs. population of the user laboratory). Presumably the possible differences between hospital workers and the general population are minimal but, to date, there is no evidence that a population of hospital workers is representative of the healthy general population. Usually, in all population studies, one verifies the study population is representative of the general population.

Investigating possible regional differences is one of the objectives of Ichihara and coworkers. Undoubtedly, this is a very ambitious goal, but useful and essential. Few validated information is available on this subject in the scientific literature. However, regional differences studied are limited to those observed in populations of hospital workers from participating Asian laboratories: they may not be representative of the diverse population’s subgroups of the various states where are located the laboratories participating in the study. For practical application, additional studies are needed to ensure that the observed results are applicable to the entire population (and subgroups) of the states considered.

As regards the statistical aspects of these studies Ichihara et al. used for calculating the RIs a careful approach for the treatment of their data. The RIs were derived parametrically by the use of a modified Box-Cox transformation [12]. Previously, the reference values were carefully selected a priori by a health questionnaire and then, in a second step, by exclusion of individuals with clinical or laboratory abnormalities. To refine the set of reference values the authors use a third step of exclusion, an iterative approach called “Abnormal Latent Value Exclusion” (LAVE) [12].
By combining data from several laboratories it is possible to obtain reference values samples of a large size which reduce the confidence intervals around the reference limits. It is also possible to collect samples of multiple subgroups (e.g., age, gender, ethnic, under different environmental conditions). If these subgroups are large enough (at least >500) the criteria for partitioning can be applied. It must be recognized that in both articles the number of individuals per region is relatively low (e.g., between 59 and 232 in the first study). It is sufficient to calculate global RIs but not enough to consider multiple criteria for partition. It is not responded to the question whether the number of reference individuals is sufficient to derive RIs per subclass (for analytes for which significant differences due to several factors of variation were observed).

To determine whether the differences due to factors of biological variation were large enough to justify a partition, Ichihara and coworkers using a method derived from nested ANOVA [12]. The authors compare the standard deviation (SD) of the factor of variation considered (e.g., age, sex, between cities, etc.) corresponding to the SD of the component of the inter-individual variation [standard deviation ratio (SDR)]. A threshold ratio >0.30 for the SDR was used to request the partition of the reference values by a factor of variation. Initially, in a previous study [11], Ichihara proposed, empirically, a threshold ratio >0.40. This cut-off value of the SDR was adjusted to 0.30 by comparing the results obtained with the authors’ own method to that proposed by Harris and Boyd (C28 A3) [1]. Using this new threshold, results obtained with both statistical approaches are consistent to judge the partitioning for gender. In contrast, in the articles published in this issue of Clinical Chemistry and Laboratory Medicine [9, 10] another threshold value of the SDR (>0.25) was chosen specifically to assess the need to partition RIs on a geographical basis. To justify their position the authors claimed SDR >0.30 is sometimes insensitive if only one or two regions show a difference compared to all other regions. Further studies are definitely needed to clarify the validity of this approach, including the determination of threshold value to be used to judge the appropriateness of partitioning reference values.

Undoubtedly, these two studies represent a tremendous effort from all the scientists involved. Coordination work and study design is equally remarkable. It is a field of valuable experience for all the researchers involved in reference values studies, opening new promising perspectives for the future. However, before the many facets of this protocol can be generalized, some additional studies are needed. The field of reference values is one of the most complex of the clinical laboratory sciences, at the interface between laboratory medicine, clinical medicine and epidemiology. Thus we can dream that future works involve experts from these three disciplines of the medical sciences.

Conflict of interest statement

Author’s conflict of interest disclosure: The author stated that there are no conflicts of interest regarding the publication of this article.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

References


Joseph Henny,
UFR des Sciences de la Santé,
Versailles Saint-Quentin University,
UMRS 1018 INSERM, Villejuif, France,
E-mail: joseph.henny@cmp.u-nancy.fr