The IFCC recommendations for determining reference intervals: strengths and limitations

Abstract

Based on the original recommendation of the Expert Panel on the Theory of Reference Values of the International Federation of Clinical Chemistry (IFCC) dating back to the 1980s, updated guidelines were recently published under the auspices of the IFCC and the Clinical Laboratory Standard Institute (CLSI). Whereas the theory and the fundamentals of the reference value concept (definition, selection of reference individuals, preanalytical and analytical requirements, analysis of reference values) have not been modified, the updated guidelines add valuable improvements (transference, validation and verifying reference intervals). However, certain limitations remain which will be discussed in the present article.

Keywords: decision limits; reference intervals; reference values.

Zusammenfassung


Schlüsselwörter: Entscheidungsgrenzen; Referenzbereiche; Referenzwerte.

Assistance with interpretation of laboratory tests is one of the major concerns of laboratory staff. Accompanying results with reference intervals for quantitative analytes help the clinicians (and the laboratorians) to interpret laboratory results.

This concept of reference values was instigated in the 1970s under the guidance of Ralph Gräsbeck associated with a Scandinavian working group [1]. Some years later, the International Federation of Clinical Chemistry (IFCC) and a French speaking group within the Société Française de Biologie Clinique, comprising representatives of Spain, Belgium, Switzerland and Canada, published a series of articles introducing not only the concept of reference values but also recommendations which were internationally accepted [2, 16]. In the middle 1990s, the National Committee for Clinical Laboratory Standards (NCCLS) published guidelines for defining and determining reference intervals in the clinical laboratory [2–17].

At the same time, the directive 98/79/CE of the European Community imposed on the manufacturers of the diagnostic in vitro industry to provide reference limits (or intervals) on the package insert of laboratory reagent kits [18]. Many international organisations (EC4, ISO, with the ISO 15189 standard) recommend providing such information [19].

One should recognise that the guidelines proposed by the IFCC and then by NCCLS were less and less adapted to the needs of the professionals, manufacturers and laboratorians. In 2005, a joint meeting gathered the Committee on Reference Limits and Decision Limits of the IFCC-LM and a Working Group of the Clinical Laboratory Standards Institute (CLSI, formerly NCCLS) during the XIX International Congress of the IFCC-LM Orlando (FL, USA). During the course of this meeting, it was decided to revise the original document on the determination of
reference values (C28.A) published in 1995 by the NCCCLS [17] and this was updated in 2000 [20].

Finally, the working group was set up with members designated by the CLSI, advisor members of the Committee on Reference Limits and Decision Limits of the IFCC (C-RIDL) and additional experts.

The goals of the joint committee were to revise the former documents of the IFCC and of the CLSI in order:

- to provide clinical laboratories and diagnostic test manufacturers with updated guidelines for determining reference intervals for quantitative analytes,
- to provide recommendations regarding procedures that can be used to verify reliable reference intervals for use in laboratory medicine, and
- to publish a document common to IFCC and CLSI.

The updated document was published in November 2008 under the guarantee of the IFCC and CLSI and was entitled “Defining, Establishing and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline – Third Edition – C28-A3” [21].

In the present article, the general rules proposed by the IFCC Expert Panel on the Theory of Reference Values (EPTRV) will be briefly recalled and described. The update of the joint document of the IFCC and CLSI will be mentioned each time that it will be necessary.

One shall clearly recognise that the major guidelines of the original recommendation of the IFCC are preserved in the updated document as shown in Table 1.

Terminology and protocol for determining reference intervals

The terminology of the Expert Panel on Theory of Reference Values of the IFCC and International Council for Standardisation in Haematology (ICSH) was endorsed by the World Health Organisation. The definitions concern an observed value, a reference distribution, a reference individual, a reference interval, a reference limit and a reference value. The complete definitions of the three most important terms are given in Table 2. All these definitions have not been changed in the updated versions as well as the protocol for determining reference intervals. This protocol implies to carefully follow a sequence of operations summarised in Table 3.

Selection of reference individuals

There are two common methods for selecting reference individuals. The “direct sampling technique”, which is strongly recommended, requires defining in advance the health status needed for the reference individuals. Defining “good health” is extremely complicated and implies to collect a variety of preclinical and clinical information either by a short medical examination and/or by questionnaires for each candidate reference individual. During the first step, one excludes the non-healthy individuals from the reference sample (on the basis of exclusion criteria), and in the second step one divides the reference sample into significant subclasses (e.g., the most common being age/gender). All these criteria (exclusion and partition) must be documented.

When the process of selection takes place before blood sampling, it is referred to as a priori. When the same criteria as before are applied following the sample collection, it is referred to as a posteriori.

The second method for selecting reference individuals is known as “indirect sampling technique”. This is not recommended as a primary approach. With this technique, a set of laboratory values from a database are

Table 1 IFCC/CLSI newly revised recommendations in comparison to the former guidelines by the IFCC.

<table>
<thead>
<tr>
<th>That has not been modified</th>
<th>That was updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>• IFCC/ICSH definition</td>
<td>• Analysis of reference values</td>
</tr>
<tr>
<td>• Selection of reference individuals</td>
<td>• Robust method</td>
</tr>
<tr>
<td>• Direct sampling techniques</td>
<td>• Transference</td>
</tr>
<tr>
<td>• Indirect sampling techniques</td>
<td>• Comparability of the analytical system</td>
</tr>
<tr>
<td>• Preanalytical and analytical requirements</td>
<td>• Comparability of the test subject population</td>
</tr>
<tr>
<td>• Analysis of reference values</td>
<td>• Verifying reference intervals</td>
</tr>
<tr>
<td>• Parametric and non-parametric methods</td>
<td>• Multicentre reference intervals studies</td>
</tr>
<tr>
<td>• Minimum number of reference values</td>
<td></td>
</tr>
<tr>
<td>• Outliers detection</td>
<td></td>
</tr>
<tr>
<td>• Presentation of laboratory reports</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Definition of a reference individual, a reference value and a reference interval [2, 20].

<table>
<thead>
<tr>
<th>Reference individual</th>
<th>A person selected on the basis of well-defined criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference value</td>
<td>The value obtained by the observation or measurement of a particular type of quantity on a reference individual</td>
</tr>
<tr>
<td>Reference interval</td>
<td>The interval between and including two reference limits (e.g., the central 95% apparently healthy men between 18 and 65 years)</td>
</tr>
</tbody>
</table>
used as reference values. It is used when it is too difficult to collect samples from healthy subjects (e.g., paediatrics, cerebrospinal fluid (CSF), but special precaution must be taken not to include a large number of values from unhealthy individuals. This is why it could be appropriate to use data from individuals who are supposedly healthy. The reference intervals determined by indirect techniques should be considered as an approximate estimate.

**Technical recommendation: preanalytical and analytical phase**

The methodology followed for determining reference values is of utmost importance. Standardisation of the different steps of the preanalytical and analytical phases is strongly recommended by the IFCC [4, 5].

The subject preparation should take into account clinically meaningful preanalytical factors to minimise their effects on reference values. The most important include: fasting vs. non-fasting, drug, biological rhythms, physical activity, tourniquet time, etc.

A detailed list of these factors is cited in a document by the IFCC EPTRV [4]. The specimen type, the sample collection and processing (handling and storage, temperature) must be precisely defined.

The analytical variability of the method used must also be carefully controlled and be described in detail. The main characteristics of the method used should be reported [between run imprecision, limit of detection, linearity, recovery, interference, trueness (traceability if applicable), etc].

**Determination of reference limits**

Once the reference values are obtained, then the reference limits and reference intervals are determined. In the documents by the IFCC EPTRV [7], two main statistical methods were described:

- The non-parametric method (which is recommended) makes no specific assumptions about the mathematical form of the probability distribution of the observed reference values.
- The parametric method assumes that the observed values follow a Gaussian probability curve. This implies in practice that for many analytes not following the Gaussian form, the data must be transformed (“normalisation”).

In the last issue of the IFCC/CLSI guidelines, a new method, “the robust method”, proposed by Horn and Pesce was introduced [22]. The robust method is a compromise between the parametric and non-parametric methods. It requires neither a large number of observations nor a Gaussian distribution. Instead, the mean and standard deviation, robust method measures location and spread. It could be used when the available size is < 120, and when it is impossible to obtain more reference samples.

**Partition of reference values set into subclasses**

The partitioning of reference values in separate subclasses is justified only if this is clinically useful and/or well founded physiologically. This point should be considered before the analysis of data. Two approaches are proposed. The Harris and Boyd method offers the advantage of its simplicity [23]. This method proposes to calculate the statistical significance of the difference between subclass means by the standard deviation test. On the other hand, it imposes a Gaussian distribution and an equal prevalence of each subclass.

An alternative was proposed in the revised document in 2008. The Lahti method is based upon direct estimation of the proportions of two subclasses outside the reference limit at each end of the combined distribution [24].

However, neither the Harris and Boyd nor the Lahti method can solve the partitioning problem in the presence of several subclasses.

**Minimum number of reference values**

The minimum number of reference values remains a challenge. For the non-parametric method it is desirable that as many reference values as possible be collected for obtaining the reference set. A minimum of 120 reference individuals is recognised by the IFCC as a good compromise. One has to consider that the major factor which influences the width of confidence intervals of the reference limits is the number of available reference values.
With the robust method, no specific minimum number is required. The same remarks as above concerning the uncertainty of reference limits obtained with this method will apply.

Detection of outliers

The determination of reference intervals implies that the tested population represents a homogeneous collection of reference individuals. Usually, this cannot be identified by the technician performing the analysis. Generally, outliers are values which are outside the global set of reference values. These outliers can be detected and "eliminated" in three steps: the first step consists of a visual inspection of the frequency distribution. Second, several statistical tests are available for detection of outliers. The Dixon test is well known and is recommended by the IFCC Committee [25]. Another usual test, proposed by Tukey could equally be used [26]. Finally, when one or several outliers are excluded, it is proposed to check the remaining data for an additional outlier.

Common reference intervals

In the face of such difficulties for determining reference intervals (mainly to recruit reference individuals), it is proposed to calculate "common reference intervals" from multicentre reference intervals studies. However, a number of conditions must be met. The comparability of analytical systems from the different participating laboratories should be achieved. In addition, the usual prerequisites should be satisfied: a priori selection of reference individuals according to the main protocol, definition of the preanalytical phase in each laboratory, interlaboratory standardisation, common quality control program and traceability of the results.

Transference of reference limits

One of the major innovations of the new document C28.A3 is to consider that the determination of reliable reference intervals is a costly and heavy task and that it is unrealistic to expect each laboratory to develop its own reference intervals for each new test and method introduced in the laboratory. Considering that several sources of reference intervals are regularly published in the literature, in package inserts by the manufacturers, and sometimes determined by the clinical laboratories themselves the transfer of reference intervals by some process could be useful.

Consequently, cooperation between different laboratories and/or diagnostic manufacturers for transferring appropriate and adequate reference values could be an acceptable solution. Several scenarios can arise.

In the first case, the laboratory has determined reference intervals following the original process described in the document by the IFCC. This laboratory decides to change one of its analytical systems. In this case, a simple comparison of the two analytical systems (the old one and the new one) could be sufficient. The major advantage of this procedure is to avoid collecting reference samples from reference individuals. If the assays are completely comparable (excellent correlation coefficient, small slope bias, small intercept, range of values comparable), then the values from the two methods are comparable. The reference intervals of the previous method can be used for the newly introduced analytical system. If one of the assays yields results biased higher or lower, but the two assays are highly correlated then the reference limits in the new analytical system could be recalculated using the regression equation.

This protocol is very simple and convenient for common use. However, some precaution must be taken, including an appropriate range of values represented. For a narrow range of values, mean bias evaluation may allow reference intervals recalculation of the new method.

In the second case, the clinical laboratory would like to transfer a reference interval for the same or comparable analytical system from another laboratory or from diagnostic test manufacturer package insert data. In this case, the question of transference becomes the question of the comparability of the two reference populations.

In the document C28.A3, three methods of validation of published reference intervals are proposed:

a. Subjective assessment: the reference interval may be transferred without validation study, if all pertinent factors of the original study are consistent with those of the laboratory (reference population demographic/geographic, preanalytical and analytical procedures, careful description of the method for determining reference intervals).

b. Validation using small numbers of reference individuals. This method consists of the selection of reference individuals (no > 20) representative of the laboratory’s healthy population, then of the determination of the reference values on this set of population. If no > 2 of the 20 tested subjects’ values fall outside those original reported limits, then these limits are considered valid. If not, one has to resample a new set of values from different reference individuals or to determine its own references limits according to the original method.

c. An alternative approach uses larger numbers of reference individuals (e.g., 60 individuals or more). The methodology is similar to that described above. The only advantage is to increase the power for observing statistical differences between the two sets of reference specimens.

Presentation of reference limits

To conclude, some guidelines are briefly presented below for the presentation of reference limits for the benefit of
the manufacturers and for the presentation of laboratory results for the information of the laboratory.

It is recommended to the manufacturers to provide clear and complete information on reference intervals in package inserts [reference interval for each subclass, with the indication of partitioning factors examined, complete data about the reference population (number, geographical origin, ethnicity, percentiles used), data about analytical procedures (traceability, description of the analytical system) and the statistical methodology].

The recommendations for the presentation of laboratory results are neither new nor innovative. Each laboratory results should be accompanied by an appropriate reference interval adapted to the age and gender of the patient and to the analytical method used. Reports should highlight a laboratory result outside of the reference interval. However, the revised document takes into consideration that for some analytes, reference intervals are replaced by decision limits (e.g., for total cholesterol, glycated haemoglobin, etc.). In this case, it is highly recommended that reports should cite only the decision limits but not the reference intervals.

Discussion

The revised recommendations for determining reference intervals undoubtedly have some strength. The first and major advantage is the pragmatic approach to the difficult question of reference values. Furthermore, this updated version shows some improvements: a clearer guideline for the manufacturers in the aim to fulfil the European directive, a protocol for establishing reference intervals with small numbers of reference samples, and finally introducing the concept of verifying the applicability of reference intervals proposed by the manufacturers or from the literature to its own laboratory instead of the determination of its own reference limits, which is heavy, costly and often impossible.

However, the situation is far from being perfect, if this guideline constitutes real progress in comparison to the first articles by the IFCC and the first version of the NCCLS document, some limitations remain which should be known.

Selection of reference population according to the original method is by no means easy, it is challenging. Petitcher pointed out that it is virtually impossible to select a small number of healthy individuals representative of the biological diversity (metabolic and physiological processes that might be different amongst individuals, different ethnic and genetic background, lifestyle, etc.) [27].

In fact, without a careful selection, it is utopia to obtain unbiased data. A good example was shown by Kallner et al. who calculated reference intervals for glucose in two different sets of populations: a “non-disease” group and a “non-healthy” group [28]. In the two groups, the upper limit (97.5 percentiles) varied very much and reached approximately 15 mmol/L in the “non-disease” group and 25 mmol/L in the “non-healthy” group. This perfectly demonstrates that the original population was largely biased and a careful selection of reference individuals on clinical criteria is useful. It is often assumed that healthy individuals’ subsets are representative of a defined population and are homogeneous. However, probability sampling of individuals is rarely carried out and mathematical transformation will never make the data unbiased.

Finally, practical considerations limit the use of the original method to laboratories with large resources (large number of samples, clinical data availability, cooperation with medical doctors for obtaining informed consent, if needed by local regulation, etc.).

Analytical variability is often observed between different analytical systems for the measurement of the same analyte. Standardisation of the methods is one method for progress. The revised protocol recommends promoting the use of analytical methods traceable to a reference system. In practice, despite efforts of the manufacturers, harmonisation of analytical techniques has not yet been achieved. Secondly, very few analytes are measured with traceable analytical systems.

The recommended statistical methodology is the non-parametric method which does not require special competence in statistics, but which requires a minimum of 120 reference samples.

The other proposed methods (bootstrap, parametric methods) require trained personnel who can interpret these complex procedures. The robust method is simple and neither requires that the underlying distribution is normal nor a minimum number of samples. However, an effect of sample size on confidence intervals of reference limits exists. This is why a higher number of reference subjects would be recommended (particularly with a highly skewed distribution).

The transference procedures are perhaps the major contribution of the revised guidelines. The subjective validation study supposes that the required information concerning the reference population (demographics, geographics, statistical methodology, etc.) are carefully described. This is rarely the case in the manufacturer's package insert. The manufacturers should make efforts to be more informative.

On the other hand, the transference validation study using small numbers of reference individuals proposed for testing the comparability of the laboratory population with data from another population is very simple and within all the laboratories capability. However, selecting healthy individuals who are representative of the laboratory population is not easy. In many cases, clinical data are not always available.

The question of the use of reference intervals has been addressed neither by the IFCC/CLSI recommendations nor by the former expert panel of the IFCC. The recent trend aiming to replace reference limits by decision limits
for some analytes suggests clarifying the respective use of these two types of limits.

The definition of reference limits is clear as they are descriptive of a specific population and are determined from the reference distribution (see above). The decision limits are threshold above or below which a specific medical action is recommended [29]. In practice, the presentation of results should be adapted to these new trends depending on (national or international) recommendations and/or on clinical goals. For example, a quantitative result will be accompanied by:

- decision limits: if national (or international) consensus exists (e.g., total cholesterol, HbA1C)
- decision limits: for screening (e.g., for defining impaired fasting glucose: 100 mg/dL, 5.5 mmol/L)
- decision and reference limits: two-step diagnosis strategy (e.g., B-type Natriuretic Peptide (BNP) and N-terminal fragment 1–76 of its precursor (NT-proBNP):
  - above the upper reference limit in a presumably healthy population, the risk of cardiovascular disease is low
  - above the decision limit: an increase in cardiovascular mortality is identified. These patients should be offered diagnostic work-up and treatment
  - between upper reference limit and high risk decision limit: “grey zone”
- reference limits: in all other cases.

In summary, the concept and the theory of reference values proposed by the IFCC > 25 years ago remain in place. However, the recent update proposed jointly by the IFCC-LM and the CLSI add a valuable improvement for responding to the requests:

- of the manufacturers in order to fulfil the requirements of the European regulation, and
- of the laboratorians in order to have at their disposal some practical way for fulfilling the same regulation: the notion of determining reference values can be replaced by the notion of verifying reference intervals every time it is possible.

New proposals on the methodological level for the determination of reference intervals (selection of reference individuals and/or statistical analysis) are welcome provided that they contain practical and more efficient additions for clinical laboratories.

In conclusion, one should rethink the reference intervals within the framework of their use. For medical decision-making, the reference intervals (or reference limits) are only a guide for the interpretation of laboratory results. They should be used together with decision limits adapted to the clinical situation of each patient.

References

17. National Committee for Clinical Laboratory Standards. How to define and determine reference intervals in the clinical laboratory; approved guideline. NCCLS document C28-A.
21. Clinical and Laboratory Standards Institute. Defining, establishing and verifying reference intervals in the clinical labo-