EFLM Recommendation

Joseph Henny, Anne Vassault, Guilaine Boursier, Ines Vukasovic, Pika Mesko Brguljan, Maria Lohmander, Irina Ghita, Francisco A. Bernabeu Andreu, Christos Kroupis, Ludek Sprongl, Marc H.M. Thelen, Florent J.L.A. Vanstapel, Tatjana Vodnik, Willem Huisman and Michel Vaubourdolle*, on behalf of the Working Group Accreditation and ISO/CEN standards (WG-A/ISO) of the EFLM

Recommendation for the review of biological reference intervals in medical laboratories

DOI 10.1515/cclm-2016-0793
Received September 6, 2016; accepted September 15, 2016; previously published online October 14, 2016

Abstract: This document is based on the original recommendation of the Expert Panel on the Theory of Reference Values of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), updated guidelines were recently published under the auspices of the IFCC and the Clinical and Laboratory Standards Institute (CLSI). This document summarizes proposals for recommendations on: (i) The terminology, which is often confusing, noticeably concerning the terms of reference limits and decision limits. (ii) The method for the determination of reference limits according to the original procedure and the conditions, which should be used. (iii) A simple procedure allowing the medical laboratories to fulfill the requirements of the regulation and standards. The updated document proposes to verify that published reference limits are applicable to the laboratory involved. Finally, the strengths and limits of the revised recommendations (especially the selection of the reference population, the maintenance of the analytical quality, the choice of the statistical method used…) will be briefly discussed.

Keywords: decision limits; reference values; reference interval; transferability.

Introduction

Support for the interpretation of laboratory tests is a major concern for medical laboratories particularly in the interpretation of results. Properly validated reference intervals for each quantitative result as given in the report is one of the main criteria for medical decision taken using biological examination.

The concept of reference values was designed in the 1970s by a Scandinavian group, then, it was developed and completed by numerous works of national societies (French and Spanish) as well as at the international level, particularly within the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and National

***Corresponding author: Michel Vaubourdolle,** Service de Biochimie, Hôpital Saint-Antoine, 184, rue du Faubourg Saint-Antoine, 75012 Paris, France, E-mail: michel.vaubourdolle@aphp.fr; and Department of Medical Biology, Hospital Saint-Antoine, University Hospital Paris East, AP-HP, Paris, France
Joseph Henny: UMS 011 Inserm UVSQ, Hôpital Paul Brousse, Université de Versailles, Saint-Quentin, Villejuif, France
Anne Vassault: Executive manager, Asqualab, Paris, France
Guilaine Boursier: Laboratoire de Génétique des Maladies Rares et Auto-Inflammatoires, Hôpital A. De Villeneuve – CHU de Montpellier Université de Montpellier, Montpellier, France
Ines Vukasovic: University Department of Chemistry, Sestre Milosrdnice University Hospital Center, Zagreb Croatia
Pika Mesko Brguljan: Department of Clinical Chemistry, University Clinic for Respiratory and allergic Diseases Golnik, Golnik, Slovenia
Maria Lohmander: Laboratoriemedicin NU-sjukvarden, Trollhättan, Sweden
Irina Ghita: Synevo Central Lab Clinical trials, Iflov County, Romania
Francisco A. Bernabeu Andreu: Servicio de Analisis Clinicos, H.U. Principe de Asturias, Alcalá de Henares, Spain
Christos Kroupis: Department of Clinical Biochemistry, Attikon University General Hospital, University of Athens Medical School, Greece. http://orcid.org/0000-0002-5876-2599
Ludek Sprongl: Central Laboratory, Sumperk Hospital a.s., Sumperk, Czech Republic
Marc H.M. Thelen: Laboratory for clinical chemistry and haematology, Amphia ziekenhuis, Breda, The Netherlands
Florent J.L.A. Vanstapel: Laboratory Medicine, University Hospital Leuven, Belgium, Department of Public Health, Biomedical Sciences Group, KU Leuven, Belgium
Tatjana Vodnik: Center for Medical Biochemistry, Clinical Center of Serbia, Belgrade, Serbia
Willem Huisman: Consultant European Specialist in Clinical Chemistry and Laboratory Medicine, The Hague, The Netherlands

Committee for Clinical Laboratory Standards (NCCLS) (now CLSI – USA) in the 1980s. Documents from public or standardization bodies have institutionalized the recommendations of the national scientific societies.

Thus, EN ISO 15189 standard [1] for Medical Laboratories and Directive 98/79/EC [2] for in vitro diagnostic devices (IVD) recommend the report of results to include, as applicable, reference interval limits provided by Medical Laboratories and the package inserts of IVD as well.

Numerous articles published in recent decades are partly outdated and does not always correspond to the needs expressed by professionals [3–9]. A joint working group from the CLSI and the IFCC has revised documents published previously. A joint paper was published under the title "Defining, Establishing and Verifying Reference Intervals in the Clinical laboratory: Approved Guideline-Third Edition – EP28-A3c" [10]. The French Society (Société Française de Biologie Clinique – SFBC) provided recommendations for determination and review of reference intervals [11].

Terminology

The following definitions have been approved by the IFCC, the International Council for Standardization in Hematology (ICSH) and the World Health Organization (WHO), then by the CLSI. They have been included in full in the latest document published jointly by the IFCC and CLSI [10].

- *Observed value:* value of an analyte obtained by observation or measurement of a test subject, which should be compared with reference values, a reference distribution, reference limit or reference interval
- *Reference distribution*: the distribution of reference values
- *Reference individual*: a person selected on the basis of well-defined criteria
- *Reference population*: a group consisting of all reference individuals
- *Reference interval*: the interval between two reference limits (these included) e.g.: 95% of apparently healthy men from 18 to 65 years
- *Reference limits*: a value derived from the reference distribution and used for descriptive purposes
- *Reference values*: the value obtained by observation or measurement of a defined quantity on a reference individual

Figure 1 shows the relationship between the various terms defined above.

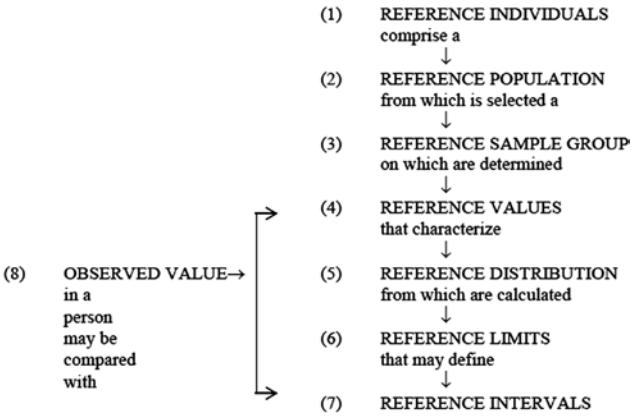

Figure 1: Relationship between the terms used for the "Reference Values Concept" [4, 10].

The *reference interval* is the interval specified in the distribution of values obtained from populations of healthy subjects. This is generally defined as an interval corresponding to 95% of the population, centered on the median. It can vary depending on the type of primary population sample and the analytical method. In some cases, only one reference limit may be used, usually an upper limit. The determination of the reference interval is based on statistical calculations. It is purely descriptive of a given population.

The *reference limits* [12] (defining a reference range) are associated with a well-defined reference population, generally consisting of healthy individuals. They are used to compare an observed value (a result from the patient) to reference data obtained from this group of well-defined subjects. They are one of the keys for medical decision making which should take into account the specificities of each patient. Reference limits are descriptive of a given health state and can sometimes be used, in well-defined cases, as decision limits.

The *medical decision limits* [12] are used by the clinician as a threshold below or above which a medical action is recommended. While reference limits are generally two (upper and lower limits), the number of decision limits is variable according to concerned laboratory test and clinical setting. They are based on a clinical assessment and are set either by statistical methods (e.g. Bayesian approach) or from epidemiological studies.

In some cases, for some analytes, reference ranges are replaced by decision limits set by national or international consensus (e.g. total cholesterol, glycated hemoglobin…). For these analytes it is unnecessary to determine reference limits or to validate data from the literature.

Finally, it should be borne in mind that the comparison of a laboratory result to a reference or decision limits is not the only way of interpretation of laboratory tests: as recalled again Dalton about the measurement of serum creatinine? This also applies to many other analytes: low variation of an observed value (even within the reference range) may have a pathophysiological significance (for the diagnosis, monitoring of treatment or prevention) [13].

Protocol for determining reference limits

For a new examination or a new method, if there is no reliable data in the literature, the laboratory will use the basic protocol to determine the reference interval limits. When information is available, it may be preferable to validate published reference intervals.

The basic protocol includes a series of successive steps fully described in documents formerly published by the IFCC [3–9] and recent recommendations of the IFCC and CLSI [12]. Here is a simplified summary:

1. List the factors of biological and analytical variations (from literature data)
2. Determine the exclusion and partition criteria on the basis of adapted questionnaire
3. Make a written consent form and have it signed by the selected individuals
4. Categorize potential reference individuals based on questionnaire data and other assessment of health status modes
5. Exclude individuals from the reference sample based on predetermined criteria
6. Set the appropriate number of individuals reference
7. Prepare selected individuals for sample collection according to the procedures normally used for patients in the laboratory
8. Collect and process samples
9. Collect reference values: analyze specimens following well-defined methods
10. Check the reference values. Establish a histogram to evaluate the distribution of data
11. Identify possible errors and/or outliers
12. Analyze the reference values: select a statistical method and calculate the limits of reference and the reference interval
13. Document all steps and procedures followed.

The following describes some points.

Selection of reference individuals

The definition of status of "good health" is particularly difficult to establish and assume that a multitude of conditions are met. In a first step, individuals "sick" or with "risk factors" will be excluded from the sample. In a second step, the reference sample will be divided into representative sub-classes, in practice most frequently limited to gender and age.

Preanalytical factors

The objective is to control and manage the significant preanalytical factors to minimize their effects. It relates to the preparation of the subject before collection and processing of the sample (handling and storage) [6].

Exclusion criteria are designed to select groups of healthy individuals by removing diseased, conditions as pregnancy, intense exercice, drug use or at risk individuals.

Partition criteria are designed to classify reference individuals into different subclasses. The two most common are age and sex. In some cases an exclusion criteria can be considered as a partition factor.

Exhaustive lists of various variability factors (preanalytical, exclusion and partition) are published in the documents of IFCC and CLSI [5, 10].

Analytical factors

Reference intervals are related to the method of measurement used which should be carefully described. The factors of variation over time (including changes from batch to batch) must be controlled and mastered. The issues to be considered are well described in an article published by Klein [14].

The traceability to a system of reference concerns very few laboratory tests so far: if it exists, it will be described.

Statistical data analysis

Three different statistical methods were described in the official IFCC/CLSI documents [10]. The methods presented below have been proven and are internationally recognized.

– The *parametric method* is applicable to populations whose distribution is normal ("Gaussian"). If the distribution is not Gaussian, a statistical transformation

to normalize it, is to apply. Then it is to verify that the new distribution follows a normal distribution. This method is not widely used in laboratory medicine, because the observed distributions are usually skewed.

- The *non-parametric method* requires nothing of the laws of probability, so it is still applicable. However, it requires careful selection of reference individuals and a sufficient number of individuals (≥120). This is the method currently recommended by IFCC.
- The *robust method* was recently introduced in the last document of the IFCC/CLSI [12]. This is of interest when the number of subjects is limited. It does not require that the distribution is Gaussian. Statistically it is a method similar to the parametric method except that it measures the position and dispersion instead of the mean and standard deviation.

Numerous other statistical methods have been described in the literature (traditional parametric methods, bootstrap techniques, etc. …) but they require the assistance of experienced statisticians.

In practice, the first two techniques, parametric and nonparametric, are fully described in the original documents of the IFCC [8], robust technology in a book published in 2005 by Horn and Pesce [15].

Minimum number of reference values

Conventional statistical methods requires a minimum number of at least 120 values by class or subclass. Indeed, the number of values directly affects the accuracy of the calculation of the reference limits. Reaching this number is sometimes difficult (e.g. expensive tests, pediatrics, difficult sampling…), and then it is recommended to use only nonparametric methods (or robust method as an alternative).

The calculation of the confidence interval for each reference limit allows validating the number of individuals selected. It is generally accepted that the confidence interval for each reference limit should be < 0.2 times the width of the reference interval concerned.

Transferability

The IFCC protocol is considered the "gold standard", but it is unsuitable for routine practice of clinical laboratories because it is too heavy and too complex to implement. Also it seems unrealistic that each laboratory determines its own reference ranges for each new test method or analytical system introduced in the laboratory. There is no simple, universal method. The main difficulty is the selection of the population and the definition of a "healthy" individual. If recruitment is not properly done, the calculated reference limits may be skewed regardless of the statistical method used. In practice, the most common factors of exclusion are: the presence of an acute or chronic illness, long-term medication, overweight and consumption of tobacco and/or alcohol. Inclusion and exclusion criteria are to be documented according to the test and disease associated.

Review of reference intervals

In an attempt to overcome this difficulty it is proposed that only a "review or verification" of the published reference limits shall be made. The transfer of data produced by other laboratories or IVD manufacturers, combined with a simple validation process, could be a great help.

However, certain conditions must be fulfilled so that the transfer process is acceptable, including the selection process of the population and that measurement methodologies (preanalytical and analytical) are similar. For this purpose, the revised recommendations IFCC/CLSI [10] propose several solutions based on different scenarios:

Case 1: Comparison of analytical systems

Reference limits were determined from the laboratory population for a given analytical system: If the laboratory decides to change a method (or an analyzer), transferability within the same laboratory to another analytical platform turn into a "comparison of analytical systems". It is not necessary in this case to select reference individuals. The operation comes down to a comparison of methods following a recognized protocol (SFBC accreditation [16] Valtec [17, 18], CLSI EP9-A3 [19]). We calculate the equation of the regression (slope, intercept, uncertainty) after a verification of the homogeneity of the data. Fresh patient samples will be used (while the range of measurement of the relevant method will be respected). The accuracy of each method and the calibrators should be similar.

Two scenarios are possible:

1. **There is no systematic difference between the two methods**
   - The slope of the regression line is close to 1.0 (x ± %) depending of the test and defined acceptable limits. It is necessary to carefully check the distribution and homogeneity of the data.

- The intercept (positive or negative) is weak, below the predefined criteria (depending of the test and defined acceptable limits [17, 18].
- The range of measurement is similar accordingly, the results obtained with one or the other methods are compatible. The reference interval of the previous method can be used for the new method.

2. **There is a proportional difference between the two methods**
   - The data of the comparison are consistent.
   - The slope of the regression line deviates from 1 and more than x ± % depending of the test and defined acceptable limits.
   - The intercept (positive or negative) is weak, below the predefined criteria depending of the test and defined acceptable limits [17, 18].
   - The range of measurement is similar accordingly; the reference ranges for the new method can be recalculated using the equation of the regression line after previous residual analysis.

Comparison of analytical systems assumes that:

1. The protocol for method comparison is followed strictly, including the uniform distribution of values throughout the measuring range, if not the risk of leading to erroneous statistical calculations is high.
2. If the intercept is too high, the direct transfer is not recommended.
3. Linear regression is not always the best method to compare two sets of values. For example, if the range of values is too small, the evaluation of the bias between the averages of the two methods is better suited to recalculate the reference limits of the new method.

Case 2: Comparison of populations

When a laboratory wishes to transfer the reference intervals established by another laboratory (article, laboratory with the same analytical system, users group…) or by a manufacturer, the transfer of the reference interval exceeds the strict comparison of analytical systems framework, it becomes a matter of comparing reference populations.

Several approaches are proposed in the latest recommendations of the IFCC and CLSI [10].

1. **Subjective method**
   To verify that the essential elements of the original study are consistent with the working conditions and the population of the laboratory.

   Key elements to consider are:
   - the geographic and demographic criteria
   - the preanalytical procedures
   - the analytical performance
   - the description of the reference population and the protocol used
   - the statistical method for determining the reference range

   If this is the case, the reference range of the original study can be transferred without verification. Providing all the necessary information is still the main limitation.

2. **Verification of the reference interval from a sample of apparently healthy subjects (Figure 2)**
   If the subjective method is not applicable, the laboratory verifies the reference interval published (IVD manufacturer's application sheets, another laboratory data, scientific articles, etc. …).
   Protocol:
   - Selection of 20 apparently healthy individuals, taking into accounts the required exclusion and partition criteria (gender, age, absence of disease, medication…).
   - Determination of reference intervals with the method to be tested. The homogeneity of the set of data will be checked to ensure that the whole range is covered by data. If some data are discrepant as compared to the whole, they could be eliminated using a so-called method for rejection of outliers. Preanalytical and analytical conditions of the tested method and the original one will be consistent.

   Interpretation:
   - reference limits to be checked are accepted if the number of results outside the limits is ≤ 2.
   - New selection of 20 biological samples is analyzed if the number of results outside of the proposed limits is equal to 3: the same procedure as above is applied. Under these conditions, the reference limits to be checked are accepted if the number of results of the new selection out of range is ≤ 2.
   - If four or more results are outside of the limits, it is advisable to review the analytical procedure, to consider the possible presence of biological and/or demographic differences and to determine the reference limits of the new method following the original protocol. It would ne helpful to examine possibility if given reference interval data could be matched with some other refrenec interval source prior to start detailed RI examination with a large number of samples.

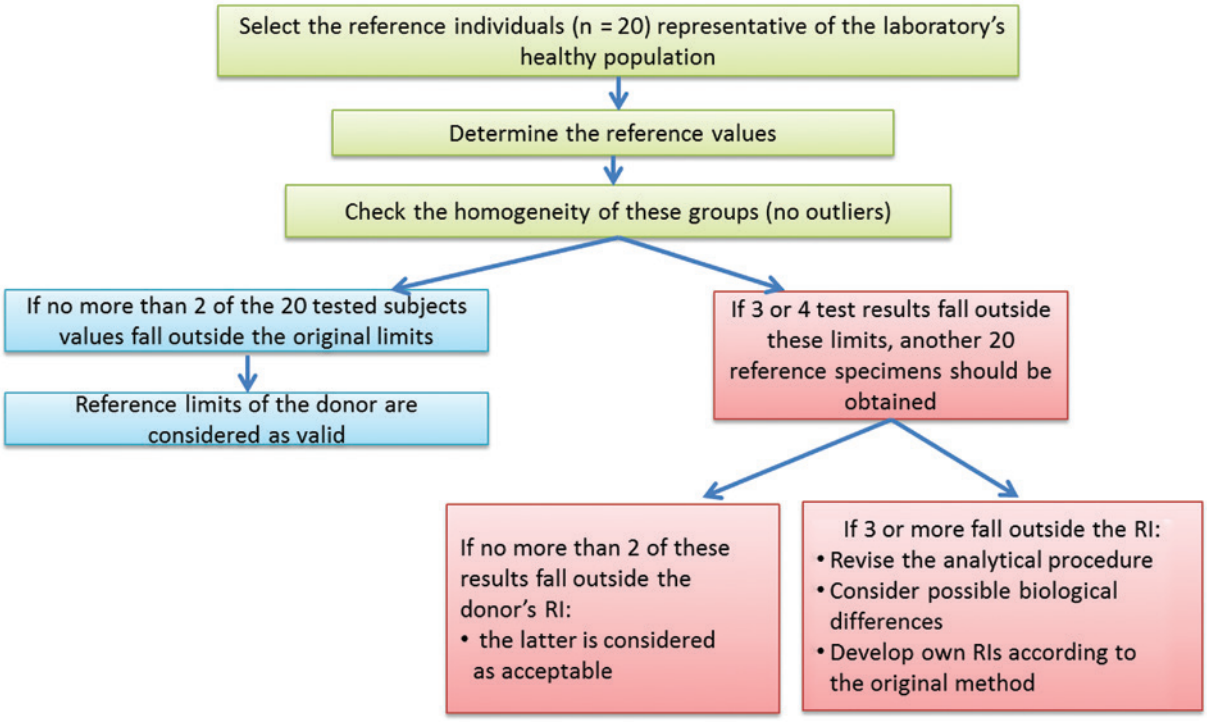

Figure 2: Validation of reference interval [10].

Multicenter studies to determine reference intervals

Given the difficulty in selecting reference individuals, pooling of data produced by different laboratories can simplify some tasks to achieve the desired goal [10]. Indeed, among the many factors of variation, two are particularly important:

- The *reference population*: the documentation of ethnic and racial differences is relatively scarce, but it cannot be overlooked.
- The *influence of the analytical method* is true. However, standardization efforts undertaken in recent years allow reducing the effects from laboratory to laboratory. However, the inter methods differences are often very important, especially in immunochemistry. It is therefore possible to produce common reference intervals through multicentre studies, subject to comparable analytical systems.

To do this the following prerequisites must be met:

- The selection of reference individuals will be consistent with the basic protocol
- The implementation of the pre analytical phase will be the same on each site and well described
- Traceability of results and the inter-laboratory standardization will be effective
- A common program of quality control will be implemented
- Finally it is recommended that each laboratory validates established reference ranges in its own environment.

Traceability

Traceability of all actions will be provided. All steps for determining Reference Intervals will be documented:

- Preanalytical conditions
- Analytical method
- Selection of the reference samples
- Statistical method

Other procedures for transferring and validating reference intervals

Tate et al. [20] published recently a critical review. Some examples are given, including the Canadian study from

Adeli et al. and those from Koerbin et al. in Australia. Adeli et al. [21–23] conducted the study of the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) and the Canadian Health Measures Survey (CHMS) which was carried on well-defined populations according to a pre-analytical and analytical protocol tightly controlled. This multicenter study compared the references intervals between the key analytical systems. Data for children, adults and elderly Canadians are presented.

Koerbin et al. [24] determined the bias between eight major analytical platforms using a similar method of comparison.

Tate [20] also offers as an alternative approach to mining data from its own laboratory. This approach is acceptable only if the laboratorian is able to identify healthy individuals not affected by a disease. This method is not recommended by the CLSI-IFCC.

Ozarda et al. [25] aims to derive reliable country specific Reference intervals through multicenter studies at a global level and a protocol was developed. It could be applied to markers traceable to a reference system and non-standardized biochemical markers if there are no regional and/or ethnic groups' differences.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. EN ISO 15189:2012. Medical laboratories – Particular requirements for quality and competence, Geneva: ISO, 2012.
2. Directive 98/79/CE. Brussels: European Community, 1998.
3. Siest G, Vernet-Nyssen M. Le concept de valeurs de référence en biologie clinique. Document A, 2ème version. Ann Biol Clin 1981;39:381–4.
4. Solberg HE. Approved recommendation on the theory of reference values (1986). Part 1. The concept of reference values. J Clin Chem Clin Biochem 1987;25:336–42.
5. Petitclerc C. Approved recommendations on the theory of reference values (1987). Part 2. Selection of individuals for the production of reference values. J Clin Chem Clin Biochem 1987;25:639–44.
6. Solberg HE. Approved recommendation on the theory of reference values (1988). Part 3. Preparation of individuals and collection of specimens for the production of reference values. J Clin Chem Clin Biochem 1988;26:593–8.
7. Solberg HE, Stamm D. Approved recommendation on the theory of 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.
8. Solberg HE. Approved recommendation on the theory of reference values (1987). Part 5. Statistical treatment of collected reference values Determination of reference limits. J Clin Chem Clin Biochem 1987;25:645–56.
9. Dybkaer R. Approved recommendations on the theory of reference values (1987). Part 6. Presentation of observed values related to reference values. J Clin Chem Clin Biochem 1987;25:657–62.
10. CLSI Document C28-A3c:2008. Defining, establishing and verifying reference intervals in the clinical laboratory; approved guideline, 3rd ed. Wayne, PA, 2008;28:61.
11. Henny J, Arnaud J, Giroud C, Vassault A; membres du sous-groupe 2 «analytique» de la SFBC. [Determination and verification of reference intervals.] Recommendations. Ann Biol Clin (Paris) 2010;68:305–13. French.
12. Ceriotti F, Henny J. "Are my laboratory results normal?" Considerations to be made concerning reference intervals and decision limits eJIFCC, 2008;19. Available at: http://www.ifcc.org.
13. Dalton RN. Serum creatinine and glomerular filtration rate: perception ad reality. Clin Chem 2010;56:687–9.
14. Klein G, Junge W. Creation of the necessary analytical quality for generating and using reference intervals. Clin Chem Lab Med 2004;42:851–7.
15. Horn PS, Pesce A-J. Reference intervals. A user's guide – Washington, DC: AACC Press, 2005.
16. Vassault A, Hulin A, Chapuzet E, Arnaud J, Giroud C; membres du sous-groupe 2 «analytique» de la SFBC. Verification/validation of the performances of analytical method. Ann Biol Clin (Paris) 2010;68:247–94. French.
17. Vassault A, Grafmeyer D, de Graeve J, Cohen R, Beaudonnet A, Bienvenu J. Quality specifications and allowable standards for validation of methods used in clinical biochemistry. Ann Biol Clin (Paris) 1999;57:685–95. French. Erratum in: Ann Biol Clin (Paris) 2000;58:110.
18. Vassault and the members of the French Society of Clinical Biology (SFBC). Protocole for the validation of methods (VALTEC protocol). Ann Biol Clin 1986;44:686–745.
19. CLSI Document EP9-A3: 2013: method comparison and bias estimation using patient samples; approved guideline – 3rd ed. Wayne, PA, 2013;22:55.
20. Tate JR, Yen T, Jones GR. Transference and validation of reference intervals. Clin Chem 2015;61:1012–5.
21. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, et al. Biochemical marker reference values across pediatric, adult and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian health measures survey. Clin Chem 2015;61:1049–62.
22. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, et al. Complex reference values for endocrine and special chemistry biomarkers across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals

on the basis of the Canadian health measures survey. Clin Chem 2015;61:1063–74.
23. Adeli K, Raizman JE, Chen Y, Higgins V, Nieuwesteeg M, Abdelhaleem M, et al. Complex biological profile of hematologic markers across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian health measures survey. Clin Chem 2015;61:1075–86.
24. Koerbin G, Sikaris KA, Jones GR, Ryan J, Reed M, Tate J, et al. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014;15:99–107.
25. Ozarda Y, Ichihara K, Barth JH, Klee G; Committee on Reference Intervals and Decision Limits (C-RIDL), International Federation for Clinical Chemistry and Laboratory Medicine. Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values. Clin Chem lab Med 2013;51:e223–4.