Review

Reference intervals: current status, recent developments and future considerations

Yesim Ozarda

Department of Medical Biochemistry, Uludag University School of Medicine, Bursa, Turkey

Corresponding author: yesim@uludag.edu.tr

Abstract

Reliable and accurate reference intervals (RIs) for laboratory analyses are an integral part of the process of correct interpretation of clinical laboratory test results. RIs given in laboratory reports have an important role in aiding the clinician in interpreting test results in reference to values for healthy populations. Since the 1980s, the International Federation of Clinical Chemistry (IFCC) has been proactive in establishing recommendations to clarify the true significance of the term 'RIs, to select the appropriate reference population and statistically analyse the data. The C28-A3 guideline published by the Clinical and Laboratory Standards Institute (CLSI) and IFCC is still the most widely-used source of reference in this area. In recent years, protocols additional to the Guideline have been published by the IFCC, Committee on Reference Intervals and Decision Limits (C-RIDL), including all details of multicenter studies on RIs to meet the requirements in this area. Multicentric RIs studies are the most important development in the area of RIs. Recently, the C-RIDL has performed many multicentric studies to obtain common RIs. Confusion of RIs and clinical decision limits (CDLs) remains an issue and pediatric and geriatric age groups are a significant problem. For future studies of RIs, the genetic effect would seem to be the most challenging area.

The aim of the review is to present the current theory and practice of RIs, with special emphasis given to multicenter RIs studies, RIs studies for pediatric and geriatric age groups, clinical decision limits and partitioning by genetic effects on RIs.

Key words: multicenter study; common reference intervals; reference values; clinical decision limits; genetic information

Introduction

In the mid 20th century, Grasbeck and Fellman published a paper entitled 'Normal Values and Statistics' as an initial study in the field of reference intervals (RIs) (1). This was followed by a presentation by Grasbeck and Sais on 'Establishment and Use of Normal Values' (2). In subsequent years it was realized that the terminology of 'normal values' was not adequate and even partially incorrect, so the term 'reference values' came into use. From 1987 to 1991, the International Federation of Clinical Chemistry (IFCC) published a series of 6 papers, in which it was recommended that each laboratory follow defined procedures to produce its own reference values (3-8). Although there were very important developments and implementa-

tions between the 1990s and 2008 (9-12), the C28-A3 guideline, published in 2008 by CLSI and IFCC constituted the most significant step in the development of RIs and is still in current use (13). This guideline entitled 'Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory' provides the necessary steps mainly for the selection of reference individuals, pre-analytical and analytical considerations, and analysis of reference values for a RI establishment study. In the C28-A3 guideline, in order to perform a multicenter RI study, criteria need to be satisfied described with the topics (i.e. a priori selection of reference subjects, clear definition of the pre-analytical phases, demonstration of traceability of results and stan-

dardization, and well defined quality control program with clear criteria) (13). In recent years, knowledge additional to the Guideline has come from the multicenter RI studies, especially those conducted by IFCC.

Interest has been renewed in the topic as a result of the following regulatory initiatives in the last two decades (14): according to the European Directive 98/79 on *in vitro* diagnostic (IVD) medical devices, diagnostic kit manufacturers are obliged to supply their clients with appropriate reference RIs for use with their assay platforms and reagents (15), and the International Organization for Standardization (ISO) 15189 standard for clinical laboratory accreditation states that each laboratory should periodically re-evaluate its own RIs (16). In the present-day era of evidence-based medicine, there is still a big gap between theory and practice with respect to the application of RIs as decision-making tools, despite the mandatory requirements (14). Through the continuing and increasing studies initiated by the IFCC, Committee on Reference Intervals and Decision Limits (C-RIDL) in recent years on multicenter RI studies, it has been possible to derive 'common' or 'harmonized' RIs on a national level from multicenter studies that follow a common protocol (17). The C-RIDL recently published two papers including a protocol and comprehensive standard operating procedures (SOPs) for multicenter RI studies (18), with indication of the utility of a panel of sera for the alignment of test results among laboratories in multicenter studies (19).

For pediatric and geriatric RIs, the challenges are even greater since samples from reference individuals are difficult to obtain (20). This problem can be overcome by gathering large populations of reference individuals (21). Another point of discussion is the confusion which arises from RIs and clinical decision limits (CDLs). Reference values are calculated specific to health whereas CDLs indicate sensitivity to disease (22).

The aim of the review is to present the current theory and practice of RIs together with a detailed evaluation of the most recent multicenter studies, an assessment of the RIs of the pediatric and geri-

atric age groups, which is still regarded as a problem in this area, a clarification of the confusion which arises from the use of CDLs and future possibilities based on partitioning by genetic information to generate RIs.

Reference intervals; the theory and the practice

RIs are derived from reference distribution, usually of 95% interval, and describe a specific population. The classical cascade is defined from reference individuals, a reference sample group, reference values, reference distribution, reference limits and RIs. The reference individuals form the reference sample group for measurement of the values from the reference population. Through statistical analysis of the distribution of the obtained values, the reference limits are calculated. These limits then define the RI (3).

The selection of reference individuals using a sample questionnaire is explained in detail in the CLSI/IFCC document, C28-A3 (13). Health is a relative condition lacking a universal definition. The designation of good health and determination of normality for a candidate reference individual may involve a variety of examinations, such as a history and physical and/or certain clinical laboratory tests. The exclusion and partitioning criteria can be implemented appropriately through a well-designed questionnaire. Exclusion criteria are features which prevent the individual from being included in the reference sample. Although some criteria, such as alcohol, tobacco and some environmental factors, may be potential exclusion criteria, amounts of consumption of alcohol and tobacco can be recorded in detail on the sample questionnaire and the effects are evaluated statistically, primarily using multiple regression analysis (MRA) (18,19). Written informed consent from participants is needed from each reference individual who agrees to participate in the study. The consent form should state clearly that laboratory personnel are allowed to obtain specimens, and to use the associated laboratory values and questionnaire information for the determination of RIs (13).

In the *a priori* sampling approach, exclusion criteria are applied before sampling collection and it is the more appropriate approach when the biology of an analyte is known. In *a posteriori* sampling, the exclusion criteria are applied after the sampling. Both of these methods are known as direct sampling, which is the primary recommendation of the IFCC. Ideally, RIs are determined on the basis of a healthy population using direct methods (4). However, indirect methods, which are also known as data mining, based on previous laboratory data can also be useful (23). Various methods may be used for the selection of a group of healthy individuals from a general hospital population and reference values are calculated from hospital data using statistical methods, such as Bhattacharya analysis (24) and some modifications of the method (25,26). There is opposition to this approach from some, as there is insufficient knowledge of the subjects and reliance on statistical methods to exclude the unhealthy subjects as explained in C28-A3. It has also been emphasised that as there is little control of the pre-analytical and analytical conditions, the indirect approach could be used for local situations or difficult groups of subjects such as neonates, children or the elderly, or as a means to confirm the goodness of the selected RI (27). Other researchers favour the indirect method as the results are clinically relevant and much simpler for an individual laboratory to implement than the time-consuming direct a priori method, which requires considerable data and professional input (28,29).

Pre-analytical and analytical aspects must be taken into consideration in the implementation of a RI study. Generally, the pre-analytical considerations involve biological (i.e. sampling time in relation to biological rhythms, fasting or non-fasting and physical activity) and methodological factors (i.e. sample collection techniques, type of additives, with or without tourniquet and sampling equipment, specimen handling, transportation, time and speed of centrifugation, and storage conditions). For reproducibility and standardization, it is essential that the pre-analytical aspects are accurately defined and described as the preanalytical phase is known to have the highest errors

in the total test process (30). Because of the importance of harmonizing pre-analytical phase of the total testing process, an effort has been made by the European Federation for Clinical Chemistry and Laboratory Medicine (EFLM) Working Group for Preanalytical Phase (WG-PRE) to support the worldwide harmonization of color coding for blood collection tube closures (31,32). EFLM, WG-PRE believes that such harmonization would reduce pre-analytical errors and substantially improve patient safety (32).

Analytical aspects include the analytical variability of the method used for the measurement, equipment/instrumentation, reagents, calibration standards, and calculation methods. Different commercial methods may be used in a trueness-based approach to the reference measurement system providing results traceable to the system and thus, comparable results can be produced in clinical laboratories. When performing a RI study, the reference measurement systems and standard reference materials are of great importance to ensure traceability of the test results in comparisons (33).

Calculation of RIs includes parametric and non-parametric calculation methods, detection of outliers, partitioning, and confidence intervals. The lower reference limits are estimated as the 2.5th percentile and the upper limits as the 97.5th percentile of the distribution of test results for the reference population. 5% of all results from healthy people will fall outside of the reported RI and as such will be flagged as being 'abnormal'. In the parametric calculation method, there is an assumption that the observed values, or some mathematical transformation of those values, follow the Gaussian or "normal" probability distribution. The reference values of many analytes do not display Gaussian distribution, so the parametric method can be applied after data transformation. The most suitable transformation method must be selected (e.g. logarithmic, power or some other function) and testing is then applied to establish whether the transformed reference values conform to Gaussian distribution. The nonparametric method of estimation does not assume the probability distribution of the observed reference values (7). Although the C28-A3 recommends the non-

parametric calculation method, the RIs calculated by the parametric and nonparametric methods were compared in the recent IFCC, C-RIDL study which concluded that the results of the two methods are very close and parametric methods can also be used as a first choice (13).

Whichever method is used in the calculation of the RIs, detection and exclusion of the outliers are very important to obtain reliable RIs. A simple but effective method for the detection of outliers is visual inspection of the data. The most common method is the D/R method proposed by Dixon (D: the absolute value of the difference between the suspected outlier and the next or proceeding value, R: the entire range of the observations) (34). If the D/R ratio is more than 1/3, the outlier is discarded. However, this method is not very sensitive when there is more than one outlier. The Horn using Tukey method is a more sophisticated method, which includes Box-Cox transformation of the data to obtain Gaussian distribution followed by identification of the outliers in interquartile ranges (IQR: Q3-Q1; Q1: lower quartile, Q3: upper quartile). At levels of < Q1 - 1.5 IQR and / or > Q3 + 1.5 IQR, the outliers are discarded (35,36). The latent abnormal value exclusion (LAVE) method proposed by Ichihara *et al.* (37) is a secondary exclusion method to exclude possibly abnormal results hidden within the reference values. This method is an iterative approach for the derivation of multiple reference RIs simultaneously, when no exclusion of values has been made in the initial computation of the RIs. The algorithm then uses those initial values of RIs to judge the abnormality of each individual's record by counting the number of abnormal results in tests other than the one for which the RI is being determined. Several statistical methodologies have been proposed to be able to make the extremely important decision of whether or not to separate different groups.

The most widely-used partitioning method is that of Harris and Boyd, in which the means and standard deviations of the subgroups are considered as a separate different standard deviation that may produce different limits (38). However, this method is only appropriate for analytes with a Gaussian distribution with subclasses, where the values are of similar size and standard deviation. A similar method was proposed by Lahti *et al.* allowing the estimation specifically of the percentage of subjects in a subclass outside the RIs of the entire population in any situation (39,40). More recently, Ichihara and Boyd recommended a partioning method on the basis of the magnitude of the standard deviations of test results named standard deviation ratio (SDR) (37). An SDR greater than 0.3 can be regarded as a guide for the consideration of partitioning reference values. This method is based on two or three level nested analysis of variance (ANOVA). Sensitivity of the population based-RIs can be increased and thereby, the usefulness of RIs is improved by stratification of age, gender, race, ethnicity and lifestyle. Stratification by age and gender is the minimum pre-requisite and other means include race, ethnicity, body mass index or nutritional habits (41).

In the IFCC publication in 1987 (7), it was recommended that reference limits should always be presented together with their 90% confidence intervals (CIs). The CI is a range of values including the true percentile (e.g. the 2.5th percentile of the population) with a specified probability, usually of 90% or 95%, as the 'confidence level' of the interval. In the C28-A3 guideline, non-parametric CIs are given from the observed values corresponding to certain rank numbers from Reed *et al.* (42). Although one can theoretically determine 95% RIs with a lower number (as few as 39 samples), it is clearly recommended that at least 120 subjects are required to calculate the CIs of the lower and upper RIs in this guideline (13). Horn and Pesce (43) proposed a 'robust method' method based on transformation of the original data according to Box and Cox (44) followed by a 'robust' algorithm giving different weights to the data, depending upon their distance from the mean. This method can provide the reference limits from a limited number of observations using only 20 subjects (45). However, a robust method with such a small number of reference subjects (e.g. N = 20) cannot provide an acceptably narrow set of confidence limits. A small number of subjects can lead to uncertainty of calculated reference limits revealed by the width of its CIs. To calculate the 90% CIs around

the limits, it is possible to use 'the bootstrap method' which is a 'resampling' method and creates a 'pseudosample' from the data. The RI is derived from each pseudosample and the process is repeated many times (1000 - 2000) yielding a distribution of lower and upper RIs (43). From this distribution, 5th and the 95th quantiles may be used to determine the 90% CI for each limit. A critical drawback of this approach is that the 90% CIs can be very wide if the sample size is small (at least 80 individuals are needed to obtain acceptably small 90% CIs) (14).

If a clinical laboratory changes the method used or wishes to apply RIs established by another laboratory which has used a different method, transference of the RIs can be implemented, rather than collecting samples from reference individuals to establish a RI for the new method. If the new method has similar imprecision and known interferences, uses the same or comparable standards or calibrators, and provides values that are acceptably comparable, the RIs can be transferred by method comparison based on linear regression analysis. In addition, the question of transference becomes one of comparability of the reference population (46).

The C28-A3 guideline allows for subjective validation of a RI by laboratory assessment of population demographics and pre-analytical and analytical parameters. This guideline recommends that each laboratory adopts existing RI values by performing an analysis to validate the transference of a RI reported by a manufacturer or other donor laboratory. The acceptability of the transfer may be assessed by examining a small number of reference individuals (N = 20) from the receiving laboratory's own subject population and comparing these reference values to the larger, more adequate original study (13). If no more than 2 of the 20 samples (or 10% of the test results) fall out of the range of the existing RI, it may be adopted for use, at least provisionally (13). If more than 2 of the 20 samples fall outside these limits, a second 20 reference specimens 20 should be obtained. If no more than 2 of the 20 samples fall out of the range of the existing RI, it may be adopted for use. If three or more again fall outside these limits, the

user should re-examine the analytical procedures used and consider possible differences in the biological characteristics of the two populations sampled (13).

Intra- and inter-individual biological variability of the subjects within the reference population may influence the determination of that RI. In 1974, it was demonstrated by Eugene Harris that only when intra-individual variability (CV_I) is greater than inter-individual variability (CV_G), (e.g. $CV_I / CV_G > 1.4$) does the distribution of values from any individual cover much of the RI (47). In contrast, with the common occurrence of CV_I / CV_G is < 0.6, the dispersion of values for an individual will span only a part of the population-based RIs. Thus, the RI will not be sensitive to changes for that individual and, on average, for any individual and in this case, subject-based RIs are considered. In cases where the reference population is a single subject, that person may serve as a reference for himself or herself, and these are known as 'individual RIs'. This approach is quite simple and requires the collection of several samples from the same individual (48). The "reference sample" is now replaced by a set of results belonging to the single individual, assumed to have been collected when he or she was in a steady state (49). However, data for statistical analysis are very different: in the individual approach, few observations are usually available. In addition, they may be collected in a defined order and may not be mutually independent. The results of measurements on these samples for a given analyte will produce a temporal series, forming a baseline against which future results will be judged. A fundamental issue is the number of samples needed to define the baseline value. This depends upon the biological variability of the analyte, its analytical reproducibility and the applied mathematical procedures (49).

Multicenter reference interval studies

The requirement that each clinical laboratory produce its own RIs is practically impossible for most clinical laboratories. The selection and recruitment of a sufficient number of reference subjects is difficult, time-consuming, and costly. Although some

laboratories have performed local studies for their own use, there have also been multicenter studies performed with considerable numbers of subjects to establish useful RIs by laboratories in the Nordic countries (50,51), Spain (52), Australia (53,54), Asia (55,56) and Turkey (57). As common standardization and traceability are crucial during production of reference values, each step of pre-analytical, analytical and statistical application follows a well-defined protocol. In recent years, C-RIDL has been devoted to the determination of Common or Harmonized RIs (58). A study was made of the measurement of three enzymes (aspartate aminotransferase - AST, alanine aminotransferase - ALT and gamma-glutamyltransferase - GGT) measured with commercial analytical systems according to the standard methods recommended by the IFCC (59). Analysis was made of patient's sera from Italy, China, Turkey and the Nordic countries, and it was concluded that for AST and ALT, the use of common RIs appears possible. However, significant differences were observed in GGT between populations, so worldwide RIs for GGT would not seem to be applicable (59).

The ongoing Worldwide Project involving many countries, which was initiated by C-RIDL, aimed to derive reliable country-specific RIs through multicenter studies. With the implementation of a common protocol and SOPs, the utility of a panel of sera was indicated for the alignment of test results among laboratories in multicenter studies (60). The two most recent papers published by the C-RIDL include this strategy for the alignment of test results for the derivation of RIs (18,19). The requirements for conducting the multicenter study, phase by phase, are described in a new protocol which recommends that a practically attainable target sample size from each country is set at a minimum of 500, which is more than double the previously recommended minimum in the C28-A3 (240; 120 male and 120 female). The other prerequisites of multicenter studies can be summarized as *a priori* selection of reference subjects (i.e. inclusion-exclusion criteria, ethnicity and questionnaire), a clear definition of pre-analytical phases (i.e. blood collection, sample proccesing, storage and transportation), a clear definition of analytical phases

(i.e. requirements for the central laboratories and the measurements, quality control, standardization of the assay and cross-comparison of values) and the statistical procedures for data analysis and reports of the results (i.e. validation of data, analyses of sources of variation, partitioning criteria and derivation of RI) (18). This should ensure that country-specific RIs are obtained in a more reproducible manner. In addition to ethnic origins, other items were included in the questionnaire to obtain more quantitative information regarding alcohol consumption, physical activities, menstrual cycle, and medications to ascertain how these factors influence test values. Overall, the procedure for standardization of test results is of the utmost importance, and all centers need to comply when dealing with standardized analytes. The requirements of the central laboratory are also described in detail, including the method of cross-check testing between the central laboratory of each country and the local laboratories before the RIs can be applied (18). In the protocol for multicenter studies (18), cross-check testing is recommended to convert the RIs obtained from the multicenter study by the centralized assay to the values of each participating laboratory. The linear structural relationship (reduced major axis regression) is used to convert the RIs. A cross comparison study with another laboratory is an approach to compare the laboratories participating in the multicenter study, using a panel of sera from healthy individuals, and recalibrating the results based on regression analysis, especially in cases where there are no standardized materials for harmonization of test results (18,19,61). The steps for the scheme of a multicenter study when all the samples from healthy individuals are collected in the participating laboratories and sent to the central laboratory for analysis are summarised in Table 1. It should be noted that when each participating laboratory acts as a central laboratory, and collects and analyzes the samples, all actions including the standardization of the analytical phase should be the responsibility of each laboratory.

Once RIs have been obtained from a multicenter study, the next step in the transference process and potential adoption of an interval is validation

of the proposed RI, which takes account of pre-analytical, analytical, and local population differences (61). An alternative approach for adopting a RI can be done by the indirect method using the laboratory's existing data which is verification of a RI. This approach can be a potential tool for further harmonization of RIs (62).

Reference intervals and clinical decision limits

The RIs are descriptive of a specific population and are derived from a reference distribution (usually 95% interval), whereas CDLs are thresholds above or below which a specific medical decision is recommended and are derived from Receiver Operating Characteristic (ROC) curves and predictive values (63). CDLs are based on the diagnostic question and are obtained from specific clinical studies to define the probability of the presence of a certain disease or a different outcome. These limits lead to the decision that individuals with values above or below the decision limit should be treated differently. CDLs are defined by consensus and vary among different populations. It is important that RIs are not confused with CDLs (64). To avoid confusion, the C28-A3 recommended reporting decision limits or RIs but not both, with a clear indication of which has been used. However, in the report example of the C28-A3, in the section of the medical decision limits, the CDLs of total cholesterol and high-density lipoprotein (HDL) cholesterol have been given in the same column as the RIs, which is confusing on the basis of terminology (13). Although highlighted at the bottom of the report sample of C28-A3, it has been noted that HDL cholesterol > 1.04 mmol/L and total cholesterol < 5.17 mmol/L are the recommendations by National Cholesterol Education Program. However, it may still be confusing because the given values are in the RIs column and they are CDLs, but not RIs. It is known that lipids (e.g. total cholesterol, HDL cholesterol) have well-defined CDLs. In the case of HDL cholesterol, decision limits can be used to categorize people as having increased risk (< 1.036 mmol/L) or decreased risk (> 1.554 mmol/L) for coronary artery disease based on data from large population studies (65). However, in recent years, C-RIDL has also supported the estimation of RIs for parameters which have clearly-defined CDLs, since RIs are specific for the characteristics of the population. For example, it is well-established that the Turkish population has a high prevalence of coronary heart disease associated with some known risk factors (66). Turks have distinctively low concentrations of HDL cholesterol, associated with elevated hepatic lipase activity and fasting triglyceridemia (67). It has also been reported that genetic and environmental factors are important in modulating HDL cholesterol concentrations in Turks (68). Therefore, it would be better to report the population-based RIs only in the RI column in the laboratory results, and to state the CDL clearly as a comment in the laboratory report, for example at the bottom of the report when a parameter has well-defined CDLs in the report.

Pediatric and geriatric reference intervals

As the concentrations of many routinely measured analytes vary significantly with growth and development, the use of inappropriate pediatric RIs can result in mis-diagnosis and mis-classification of disease. Establishing RIs can be challenging as the ideal RIs should be established based on a healthy population and stratified for key covariates including age, gender and ethnicity, but this requires the collection of large numbers of samples from healthy individuals (69). It is well known that the determination of pediatric RIs is an extremely difficult task, primarily because of ethical limitations related to blood drawing in very young children and neonates. The most significant step in this area has been taken by Adeli *et al.* in the CALIPER (CAnadian Laboratory Initiative in PEdiatric Reference Intervals) Project, which is a collaboration between multiple pediatric centers across Canada, that aims to address the current gaps in pediatric RIs and has established a database of age- and gender-specific pediatric RIs (70). Recently, the CALIPER study demonstrated the relationship between Abbott Architect assays and four other commonly used assays (Beckman Coulter, Ortho Vitros, Roche Cobas, and Siemens Vista) for a wide

TABLE 1. The scheme for multicenter reference interval studies.

The total testing process The process phases	Responsibility	Action
Preanalytical phase	Central laboratory#, All participating laboratories*	1. Organization of the study Establishing the inclusion and exclusion criteria Informing each laboratory of the sample size Deciding the essential items in the questionnaire and distributing to all participating laboratories Communicating with the participating laboratories and informing about the procedures (e.g. blood collection and preparation of the serum samples) to standardize the pre-analytical phase in each laboratory # 2. Using the same protocol in each laboratory (e.g. selection and preparation of the volunteers, blood collection and sample preparation) #,* 3. Transportation of the samples to central laboratory * 4. Storing of the remaining samples for cross-check study in each laboratory*
Analytical phase	Central laboratory#, All participating laboratories*	1. Using a reference measurement system, certificated reference materials/value assigned sera, standardization of the assays # 2. Quality control of the assays #,* 3. Analyzing all the samples # 4. Analyzing the cross-check samples *
Postanalytical phase	Central laboratory#, All participating laboratories*	1. Reporting of the test results to each laboratory # 2. Data analysis and derivation of the RIs (common RIs, if it's possible) # 3. Reporting of cross-check results and the RIs for each of the local laboratory # 4. Using the calculated RIs #,*

RIs - Reference intervals.

spectrum of biochemical markers (71). The pediatric health survey in Germany (KiGGS) is an another excellent example in this area (72). As these direct studies were well conducted and of large sample size, the current problems in pediatric RIs could be resolved through evaluation and application of the findings. However, as an alternative, indirect methods can be used for the pediatric group as recommended in the C28-A3 (13,73).

The major difficulty in obtaining geriatric RIs is in the selection of healthy individuals as most seniors do not meet the criteria of the C28-A3. The width of the reference range is altered by factors such as the regular use of medications or unrecognized subclinical diseases. Therefore, it becomes very difficult to differentiate the effects of age, aging or a pathological condition. Although there

has been increasing interest in this subject (74,75), this issue remains incomplete in the same way as for pediatric RIs. To overcome this problem, a multicenter study which has extensive sample size from the pediatric, adult and geriatric age groups, is the best way to establish and harmonize the RIs across a country (76,77).

Laboratory RIs during pregnancy, delivery and the early postpartum period are another specific group as physiological changes during pregnancy may affect laboratory parameters and there is a need to establish reference values during pregnancy in order to recognize pathological conditions (78). Reporting the correct gestational age-specific reference values can also improve the sensitivity of the RIs as mentioned before in this review by stratification of age and gender.

Partitioning by genetic effects on reference intervals

Integrating genetic and laboratory information would increase the accuracy of RIs by eliminating extreme results related to genetic variation. It has been reported that the use of genetic information to partition RIs could reduce the between-person variation and therefore with the reduced variance obtained from partitioning based on genetic differences, there could be potentially less mis-identification of unusual test results caused by non-disease associated genetic variations (79). The genetic information was used for subgroup stratification for ApoE (80) and more recently for haptoglobin (81). Ozarda *et al.* published a paper on methylenetetrahydrofolate reductase (MTHFR), and reported that serum folate and homocysteine status are impaired by subgroup stratification of the rate of MTHFR 677C > T i 1298A > C (82). However, the extent of biological variability induced by genetic variants is often low and there is often a lack of knowledge of the genetic status of the reference individuals. As whole-genome data becomes clinically available and more associations between genetic polymorphisms and laboratory test results are discovered, it will become possible to integrate the genetic information with RI values.

The RIs for uncommon sample types [e.g. cerebrospinal fluids (CSF), amniotic fluids] are usually interpreted on the basis of values reported in reference texts or handbooks; however, current reference texts either present normal CSF parameters without citation or cite studies with significant limitations. Recent developments to determine accurate, age-specific reference values for glucose, protein concentrations and white blood cell counts in CSF, amniotic fluids and aspirations in a large population of neonates and young infants will bring literature up to date at a time when molecular tools are commonly used in clinical practice (83,84).

Conclusion

Due to the increasing numbers of multi-centric studies in recent years, there was seen to be a need for a detailed protocol. IFCC, C-RIDL met this need with the publication of a very detailed protocol in 2014, which can be used when conducting multicentric studies. Based on this protocol a number of multicentric RI studies have been performed and common RIs have been reported. The common RIs reported in the multicenter study should be validated locally, using reference specimens from healthy individuials in the local population as recommended by C28-A3 and recent C-RIDL studies.

Although indirect methods can be used as an alternative, the problem of valid RIs for specific age groups (e.g. pediatric, geriatric) has not yet been resolved. Specific RI values for pregnant women and for uncommon samples are also necessary. It is vital that a clear distinction is made between RIs and CDLs to allow optimal use of laboratory tests and avoid misdiagnosis. Future studies should focus more specifically on the genetic effects on RIs and generate genotype-specific RIs.

Potential conflict of interest

None declared.

References

1. Gräsbeck R, Fellman J. Normal values and statistics. Scand J Clin Lab Invest 1968;21:193-5. http://dx.doi.org/10.3109/00365516809076984.
2. Gräsbeck R, Saris NE. Establishment and use of normal values. Scand J Clin Lab Invest 1969;Suppl 110:62-3.
3. Solberg HE. International Federation of Clinical Chemistry. Scientific committee, Clinical Section. Expert Panel on Theory of Reference Values and International Committee for Standardization in Haematology Standing Committee on Reference Values. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. Clin Chim Acta 1987;165:111-8. http://dx.doi.org/10.1016/0009-8981(87)90224-5.

4. Petit Clerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 2. Selection of individual for the production of reference value. Clin Chim Acta 1987;170:S1-S12. http://dx.doi.org/10.1016/0009-8981(87)90150-1.

5. Solberg HE, Petit Clerc C. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 3. Preparation of individuals and collection of the specimens for the production of reference values. Clin Chim Acta 1988;177:S3-S11. http://dx.doi.org/10.1016/0009-8981(88)90074-5.

6. Solberg HE, Stamm D. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. Clin Chim Acta 1991;202:S5-S11. http://dx.doi.org/10.1016/0009-8981(91)90266-F.

7. Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 5. Statistical treatment of collected reference values. Clin Chim Acta 1987;170:S13-S32. http://dx.doi.org/10.1016/0009-8981(87)90151-3.

8. Dybkaer R, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 6. Presentation of observed values related to reference values. Clin Chim Acta 1987;170:S33-S42. http://dx.doi.org/10.1016/0009-8981(87)90152-5.

9. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. Clin Chem 1990;36:265-70.

10. Harris EK, Wong ET, Shaw ST, Jr. Statistical criteria for separate reference intervals: race and gender groups in creatine kinase. Clin Chem 1991;37:1580-2.

11. Fuentes-Arderiu X. Biological reference intervals and ISO15189. Clin Chim Acta 2006;364:365-6. http://dx.doi.org/10.1016/j.cca.2005.07.014.

12. Henny J, Petitclerc C, Fuentes-Arderiu X, Petersen PH, Queralto JM, Schiele F, et al. Need for revisiting the concept of reference values. Clin Chem Lab Med 2000;38:589-95. http://dx.doi.org/10.1515/CCLM.2000.085.

13. CLSI and IFCC, C28-A3 document; Defining, establishing and verifying reference intervals in the clinical laboratory: approved guideline-third edition, 2008;28:1-76.

14. Ceriotti F, Hinzmann R and Panteghini M. Reference intervals: the way forward. Ann Clin Biochem 2009;46:8-17. http://dx.doi.org/10.1258/acb.2008.008170.

15. Directive 98/79/EC of European Parliamant and the Council of 27 October 1998 on in vitro diagnostic medical devices. Offical J Eur Commun 1998;L331/1-L331/37.

16. International Organization for Standardization. ISO 15189:2012 Medical laboratories -- Requirements for quality and competence, Third edition, 2012.

17. Ceriotti F. Common reference intervals the IFCC position. Clin Biochem 2009;42:297. http://dx.doi.org/10.1016/j.clinbiochem.2008.09.017.

18. Ozarda Y, Ichihara K, Barth JH, Klee G. Committe on Reference Intervals and Decision Limits (C-RIDL), International Federation of Clinical Chemistry. Protocol and standard operating procedures for common use in wordwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1027-40. http://dx.doi.org/10.1515/cclm-2013-0249.

19. Ichihara K, Ozarda Y, Klee G, Straseski J, Barth JH, Baumann N, et al. Utility of panel of sera for alignment of test results in the worldwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1007-20. http://dx.doi.org/10.1515/cclm-2013-0248.

20. Siest G, Henny J, Gräsbeck R, Wilding P, Petitclerc C, Josep M, et al. The theory of reference values: an unfinished symphony. Clin Chem Lab Med 2013;51:47-64. http://dx.doi.org/10.1515/cclm-2012-0682.

21. Ceriotti F. Establishing pediatric reference intervals: a challenging task. Clin Chem 2009;58:5.

22. Ceriotti F, Henny J. "Are my laboratory results normal?" Considerations to be Made Concerning Reference Intervals and Decision Limits. eJIFCC 2008;19:1-9.

23. Grossi E, Colombo R, Cavuta S, Franzini C. The RELAB Project: a new method for the formulation of reference intervals based on current data. Clin Chem 2005;51:1232-40. http://dx.doi.org/10.1373/clinchem.2005.047787.

24. Bhattacharya C. A simple method of resolution of a distribution into Gaussian components. Biometrics 1967;23:115-35. http://dx.doi.org/10.2307/2528285.

25. Baadenhuijsen H, Smith CJ. Indirect estimation of clinical chemical reference intervals from hospital patient data. J Clin Chem Clin Biochem 1985;23:829-39.

26. Concordet D, Geffré A, Braun JP, Trumel C. A new approach for the determination of reference intervals from hospital-based data. Clin Chim Acta 2009;405:43-8. http://dx.doi.org/10.1016/j.cca.2009.03.057.

27. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev 2007;28:115-21.

28. Ozarda Ilcol Y, Diler A. Use of total patient data for indirect estimation of reference intervals for 40 clinical chemical analytes in Turkey. Clin Chem Lab Med 2006;44:867-76. http://dx.doi.org/10.1515/CCLM.2006.139.

29. Oosterhuis WP, Modderman TA, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC? Ann Clin Biochem 1990;27:359-65. http://dx.doi.org/10.1177/000456329002700413.

30. Plebani M, Sciacovelli L, Aita A, Chiozza ML. Harmonization of pre-analytical quality indicators. Biochem Med (Zagreb) 2014:24:105-13. http://dx.doi.org/10.11613/BM.2014.012.

31. Lippi G, Banfi G, Churcha S, Cornesa M, De Carli G, Grankvista K, et al. Preanalytical quality improvement. In pursuit of harmony, on behalf of European Federation for Clinical Chemistry and Laboratory Medicine (EFLM) Working group for Preanalytical Phase (WG-PRE) Clin Chem Lab Med 2015;53:357-70. http://dx.doi.org/10.1515/cclm-2014-1051.

32. Simundic AM, Cornes MP, Grankvist K, Lippi G, Nybo M, Ceriotti F, et al. Colour coding for blood collection tube closures-a call for harmonisation. Clin Chem Lab Med 2015;5:371-6. http://dx.doi.org/10.1515/cclm-2014-0927.

33. Jones G, Barker A. Reference Intervals. Clin Biochem Rev 2008;29:S93-7.

34. Dixon WJ. Prossesing data for outliers Biometrics 1953;9:74-89. http://dx.doi.org/10.2307/3001634.

35. Horn PS, Feng L, Li Y, Pesce AJ. Effects of outliers and non-healthy individuals on reference interval estimation. Clin Chem 2001;47:2137-45.

36. Tukey JW. Exploratory data analysis. Reading, MA, USA: Addison-Wesley, 1977:688.

37. Ichihara K, Boyd J. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537-51. http://dx.doi.org/10.1515/CCLM.2010.319.

38. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. Clin Chem 1990;36:265-70.

39. Lahti A, Hyltoft Peterson P, Boyd JC. Impact of subgroup prevalances on partitioning Gaussian distributed reference values. Clin Chem 2002;48:1987-99.

40. Lahti A, Hyltoft Peterson P, Boyd JC, Rustat P, Laake P, Solberg HE. Partitioning of nongaussian distributed biochemical reference data into subgroups. Clin Chem 2004;50:891-900. http://dx.doi.org/10.1373/clinchem.2003.027953.

41. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. Crit Rev Clin Lab Sci 1989;27:409-37. http://dx.doi.org/10.3109/10408368909106595.

42. Reed AH, Henry RJ, Manson WB. Influence of statistical method used on the resulting estimate on normal range. Clin Chem 1971;17:275-84.

43. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. Clin Chem 1998;44:622-31.

44. Box GEP, Cox DR. An analysis of transformations. JR Stat Soc 1964; B26:211-52.

45. Efron B. The jacknife, the bootstrap, and other resampling plans. Philadelphia, PA: CBMS-NSF regional conference series in applied mathematics. 1982:92. http://dx.doi.org/10.1137/1.9781611970319.

46. Horn PS, Pesce AJ. Reference intervals an update. Clin Chim Acta 2003;334:5-23. http://dx.doi.org/10.1016/S0009-8981(03)00133-5.

47. Harris EK. Effects of intra- and interindividual variation on the appropriate use of normal ranges. Clin Chem 1974;20:1535-42.

48. Fraser CG. Inherent biological variation and reference values. Clin Chem Lab Med 2004;42:758-64. http://dx.doi.org/10.1515/CCLM.2004.128.

49. Queralto JM. Intraindividual reference values. Clin Chem Lab Med 2004;42:765-7. http://dx.doi.org/10.1515/CCLM.2004.129.

50. Rustad P, Felding P, Lahti A, Petersen PH. The Nordic Reference Interval Project 2000. Scand J Clin Lab Invest 2004;64:271-84. http://dx.doi.org/10.1080/00365510410006324.

51. Nordin G, Mårtensson A, Swolin B, Sandberg S, Christensen NJ, Thorsteinsson V. A multicentre study of reference intervals for haemoglobin, basic blood cell counts and erythrocyte indices in the adult population of the Nordic countries. Scand J Clin Lab Invest 2004;64:385-98. http://dx.doi.org/10.1080/00365510410002797.

52. Fuentes-Arderiu X, Ferrer-Masferrer M, Gonzales-Alba JM, Villarino-Gonzales MI, Arrimadas-Esteban E, Cabrero-Olive D, et al. Multicentric reference values for some quantities measured with the Elecsys 2010 R analyser. Clin Chim Acta 2001;304:143-6. http://dx.doi.org/10.1016/S0009-8981(00)00406-X.

53. Jones GR, Barker A, Tate J, Lim C-F, Robertson K. The case for common reference intervals. Clin Biochem Rev 2004;25:99-104.

54. Koerbin G, Cavanaugh JA, Potter JM, Abhayaratna WP, West NP, Glasgow N, et al. 'Aussie normals': an a priori study to develop clinical chemistry reference intervals in a healthy Australian population. Pathology 2015;47:138-44. http://dx.doi.org/10.1097/pat.0000000000000227.

55. Ichihara K, Ceriotti F, Tam TH, Sueyoshi S, Poon PM, Thong ML, et al. Committee on Reference Intervals and Decision Limits, International Federation for Clinical Chemistry and Laboratory Medicine, and the Science Committee for the Asia-Pacific Federation of Clinical Biochemistry. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. Clin Chem Lab Med 2013;51:1429-42. http://dx.doi.org/10.1515/cclm-2012-0421.

56. Ichihara K, Ceriotti F, Kazuo M, Huang YY, Shimizu Y, Suzuki H, et al. Committee on Reference Intervals and Decision Limits, International Federation for Clinical Chemistry and Laboratory Medicine, and the Science Committee for the Asia-Pacific Federation of Clinical Biochemistry. The Asian project for collaborative derivation of reference intervals: (2) results of non-standardized analytes and transference of reference intervals to the participating laboratories on the basis of cross-comparison of test results. Clin Chem Lab Med 2013;51:1443-57. http://dx.doi.org/10.1515/cclm-2012-0422.

57. Ozarda Y, Ichihara K, Aslan D, Aybek H, Ari Z, Taneli F, et al. A multicenter nationwide reference intervals study for common biochemical analytes in Turkey using Abbott analyzers. Clin Chem Lab Med 2014;52:1823-33. http://dx.doi.org/10.1515/cclm-2014-0228.

58. Ceriotti F, Boyd JC, Klein G, Henny J, Queralto J, Kairisto V, et al. IFCC, Committee on Reference Intervals and Decision Limits (C-RIDL). Reference Intervals for Serum Creatinine Concentrations: Assessment of Available Data for Global Application Clin Chem 2008;54:559-66.

59. Ceriotti F, Henny J, Queralto J, Ziyu S, Ozarda Y, Chen B, et al. IFCC Committee on Reference Intervals and Decision Limits (C-RIDL). Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyl transferase (GGT) in serum: results from an IFCC multicenter study. Clin Chem Lab Med 2010;48:1593-601. http://dx.doi.org/10.1515/CCLM.2010.315.

60. Ichihara K. Statistical considerations for harmonization of the global multicenter study on reference values. Clin Chim Acta 2014;432:108-18. http://dx.doi.org/10.1016/j.cca.2014.01.025.

61. Tate J, Yen T, Jones GRD. Transference and Validation of Reference Intervals. Clin Chem 2015;61:1012-5. http://dx.doi.org/10.1373/clinchem.2015.243055.

62. Jones GRD. Validating common reference intervals in routine laboratories. Clin Chim Acta 2014;432:119-21. http://dx.doi.org/10.1016/j.cca.2013.10.005.

63. Sikaris K. Physiology and its Importance for Reference Intervals. Clin Biochem Rev 2014;35:3-14.

64. Sikaris K. Application of the Stockholm Hierarchy to Defining the Quality of Reference Intervals and Clinical Decision Limits. Clin Biochem Rev 2012;33:141-8.

65. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel of detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). JAMA 2001;285:2486-97. http://dx.doi.org/10.1001/jama.285.19.2486.

66. Onat A. Risk factors and cardiovascular disease in Turkey. Aterosclerosis 2001;156:1-10. http://dx.doi.org/10.1016/S0021-9150(01)00500-7.

67. Bersot TP, Vega GL, Grundy SM, Paloglu EK, Atagu nduz P, Ozbayrakci S, et al. Elevated hepatic lipase activity and low levels of high density lipoprotein in a normotriglyceridemic, nonobese Turkish population. J Lipid Res 1999;40:432-8.

68. Mahley RW, Can S, Ozbayrakci S, Bersot TP, Tanir S, Palaoglu KE et al. Modulation of high-density lipoproteins in a population in Istanbul, Turkey with low levels of high-density lipoproteins. Am J Cardiol 2005;96:547-55. http://dx.doi.org/10.1016/j.amjcard.2005.04.018.

69. Adeli K. Closing the gaps in pediatric reference intervals: the CALIPER initiative. Clin Biochem 2011;44:480-2. http://dx.doi.org/10.1016/j.clinbiochem.2011.02.017.

70. Jung B, Adeli K. Clinical laboratory reference intervals in pediatrics: the CALIPER initiative. Clin Biochem 2009;42:1589-95. http://dx.doi.org/10.1016/j.clinbiochem.2009.06.025.

71. Estey MP, Cohen AH , Colantonio DA, Chan MK , Marvasti TB, Randell E, et al. CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens Clinical Chemistry Assays: Direct validation using reference samples from the CALIPER cohort. Clin Biochem 2013:1197-219. http://dx.doi.org/10.1016/j.clinbiochem.2013.04.001.

72. Kohse KP, Thamm M. KiGGS—the German survey on children's health as data base for reference intervals. Clin Biochem 2011;44:479. http://dx.doi.org/10.1016/j.clinbiochem.2011.02.016.

73. Shaw JL, Cohen A, Konforte D, Binesh-Marvasti T, Colantonio DA, Adeli K. Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clin Biochem 2014;47:166-72. http://dx.doi.org/10.1016/j.clinbiochem.2013.11.008.

74. Nilsson SE, Evrin PE, Tryding N, Berg S, McClearn G, Johansson B. Biochemical values in persons older than 82 years of age: report from a population-based study of twins. Scand J Clin Lab Invest 2003;63:1-13. http://dx.doi.org/10.1080/00365510310000420.

75. Huber KR, Mostafaie N, Stangl G, Worofka B, Kittl E, Hofmann J, et al. Clinical chemistry reference values for 75-year-old apparently healthy persons. Clin Chem Lab Med 2006;44:1355-60. http://dx.doi.org/10.1515/CCLM.2006.247.

76. 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. http://dx.doi.org/10.1373/clinchem.2015.240531.

77. Tate J, Sikaris K, Jones GRD, Yen T, Koerbin G, Ryan J, et al. Harmonising Adult and Paediatric Reference Intervals in Australia and New Zealand: An Evidence-Based Approach for Establishing a First Panel of Chemistry Analytes. Clin Biochem Rev 2014;35:213-35.

78. Klajnbard A, Szecsi PB, Colov NP, Andersen MR, Jørgensen M, Bjørngaard B, et al. Laboratory reference intervals during pregnancy, delivery and the early postpartum period. Clin Chem Lab Med 2010;48:237-48. http://dx.doi.org/10.1515/CCLM.2010.033.

79. Shirts BH, Wilson AR, Jackson BR. Partitioning reference intervals by use of genetic information. Clin Chem 2011;57:475-81. http://dx.doi.org/10.1373/clinchem.2010.154005.

80. Vincent-Viry M, Schiele F, Gueguen R, Bohnet K, Visvikis S, Siest G. Biological variations and genetic reference values for apolipoprotein E serum concentrations: results from the Stanislas Cohort study. Clin Chem 1998;44:957-65.

81. Shahabi P, Siest G, Herbeth B, Ndiaye NC, Visvikis-Siest S. Clinical necessity of partitioning of human plasma haptoglobin reference intervals by recently-discovered rs2000999. Clin Chim Acta 2012;413:1618-24. http://dx.doi.org/10.1016/j.cca.2012.04.033.

82. Ozarda Y, Sucu DK, Hizli B, Aslan D. Rate of T alleles and TT genotype at MTHFR 677C- > T locus or C alleles and CC genotype at MTHFR 1298A- > C locus among healthy subjects in Turkey: impact on homocysteine and folic acid status and reference intervals. Cell Biochem Funct 2009;27:568-77. http://dx.doi.org/10.1002/cbf.1610.

83. Kestenbaum LA, Ebberson J, Zorc JJ, Hodinka RL, Shah SS. Defining Cerebrospinal Fluid White Blood Cell Count Reference Values in Neonates and Young Infants. Pediatrics 2010;125:257-64. http://dx.doi.org/10.1542/peds.2009-1181.

84. Renuart AJ, Mistry RD, Avery RA, Licht DJ, Seiden JA, Huh JW, et al. Reference Range for Cerebrospinal Fluid Protein Concentration in Children and Adolescents. Arch Pediatr Adolesc Med 2011;165:671-3. http://dx.doi.org/10.1001/archpediatrics.2011.89.