Opinion Paper

Yesim Ozarda*, Victoria Higgins and Khosrow Adeli

Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations

https://doi.org/10.1515/cclm-2018-0059
Received January 15, 2018; accepted April 3, 2018; previously published online May 5, 2018

Abstract: Reference intervals (RIs) are fundamental tools used by healthcare and laboratory professionals to interpret patient laboratory test results, ideally enabling differentiation of healthy and unhealthy individuals. Under optimal conditions, a laboratory should perform its own RI study to establish RIs specific for its method and local population. However, the process of developing RIs is often beyond the capabilities of an individual laboratory due to the complex, expensive and time-consuming process to develop them. Therefore, a laboratory can alternatively verify RIs established by an external source. Common RIs can be established by large, multicenter studies and can subsequently be received by local laboratories using various verification procedures. The standard approach to verify RIs recommended by the Clinical Laboratory Standards Institute (CLSI) EP28-A3c guideline for routine clinical laboratories is to collect and analyze a minimum of 20 samples from healthy subjects from the local population. Alternatively, "data mining" techniques using large amounts of patient test results can be used to verify RIs, considering both the laboratory method and local population. Although procedures for verifying RIs in the literature and guidelines are clear in theory, gaps remain for the implementation of these procedures in routine clinical laboratories. Pediatric and geriatric age-groups also continue to pose additional challenges in respect of acquiring and verifying RIs. In this article, we review the current guidelines/approaches and challenges to RI verification and provide a practical guide for routine implementation in clinical laboratories.

Keywords: clinical laboratories; CLSI EP28-A3c guideline; data mining; reference intervals; verification.

***Corresponding author: Yesim Ozarda,** MD, Department of Medical Biochemistry, Uludag University School of Medicine, Bursa, Turkey, Phone: +90-224-29-53917, Fax: +90-224-29-50019, E-mail: yesim@uludag.edu.tr
Victoria Higgins and Khosrow Adeli: Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

Introduction

Reference intervals (RIs) are most commonly defined as the central 95% of laboratory test results expected in a healthy reference population. Establishing accurate and robust RIs to interpret laboratory test results is a very important task [1]. According to the directive on *in vitro* diagnostic medical devices of the European Union, diagnostic manufacturers are required to supply their clients with appropriate RIs for use with their assay platforms and reagents [2], and the International Organization for Standardization (ISO) 15189 standard for clinical laboratory accreditation states that each laboratory should periodically reevaluate its own RIs [3]. Despite these requirements, RIs in most clinical laboratories remain out of date and incomplete due to the complex process of their establishment [4]. Therefore, instead of developing RIs directly from an apparently healthy population, most laboratories receive RIs for clinical use from various sources (e.g. manufacturers' package inserts, publications, text books, multicenter studies, published national or international expert panel recommendations, guidelines, local expert groups or data mining of existing data).

Several differences can exist between the sample collection procedures and laboratory operations of the RI study and the local laboratory receiving the RI. Therefore, it is of critical importance for a local laboratory to address the following question prior to receiving RIs from an external source: "Is this RI suitable for my laboratory's collection processes, method, and population?" [5]. The Clinical and Laboratory Standards Institute (CLSI, previously National Committee for Clinical Laboratory Standards [NCCLS]) EP28-A3c (formerly C28-A3) guideline for

Defining, Establishing and Verifying Reference Intervals in the Clinical Laboratory provides recommendations for transferring and verifying RIs established by external sources for a local laboratory [6]. This approach is advantageous for many laboratories as it does not require extensive recruitment of healthy reference individuals and is thus time and cost efficient. The process of transference and verification considers the consistency of preanalytical and analytical factors between those used in the direct *a priori* RI study and the receiving laboratory. If deemed to be similar, an RI can undergo transference through a method comparison analysis and subsequent determination of the best-fit regression line to subsequently calculate the transferred RI. Verification of transferred RIs is then recommended using a relatively small number of healthy reference individuals from the local population served by the laboratory [6]. Transferring and verifying common RIs obtained from multicenter studies, rather than a single laboratory, is occurring more frequently. Common RIs include both traceable RIs, defined by well-conducted multicenter studies using traceable materials [7, 8], and conventional RIs, defined by a group of experts [9]. When receiving common RIs, the most important question is still valid: "if methods and reference populations are the same/similar, can the same limits be received directly?".

The aims of this opinion paper were to clarify the use of the terms validation and verification of RIs, to describe the CLSI recommendations for verifying RIs, to discuss the development of currently available RIs, to describe the transferring/verifying processes of RIs obtained from multicenter studies, to introduce alternative verification methods, to discuss challenges for RI establishment/verification in pediatric and geriatric populations and, ultimately, to encourage routine laboratories to verify RIs.

Validation or verification?

Although the CLSI EP28-A3c guideline includes the term "verification" in the title of the guideline in Section 11 (11.1, 11.2 and 11.3), "validation" is used throughout the document [6]. In practice, there is frequent confusion about using the term validation or verification, although both refer to the same procedure with respect to assessing the validity of an RI [10]. The term "verification" means confirmation through the provision of objective evidence that specified requirements have been fulfilled. "Validation" also means confirmation through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled [3]. Thus, both terms have very similar definitions. However, in the field of laboratory medicine, validation of a measurement procedure for manufacturer and laboratory-driven tests is a more complex procedure, involving the collection of reliable and valid data on the performance characteristics and examination of the data to ascertain if the acceptance criteria have been met. The validation procedure also includes the determination of RIs through a direct RI study by the laboratory. However, when a laboratory transfers and receive an RI from the literature or another laboratory, the procedure of assessing and confirming its appropriateness for use is more likely to be defined as verification.

CLSI EP28-A3c guideline recommendations to transfer and verify RIs

It is the responsibility of individual laboratories or laboratory networks to use RIs that are appropriate for their methodologies and the population they serve. For this purpose, and to obviate the need to obtain samples from a large reference population, clinical laboratories may transfer adequate RIs from external sources. Assuming the original RI study was performed using robust methodology and statistical procedures, transferring an RI requires that certain conditions be fulfilled in order to be acceptable, prior to verifying and receiving an RI. There are two main scenarios in which RIs are transferred. First, reference values may originate from a different population/laboratory method than the receiving laboratory, and second, reference values may originate from a laboratory that shares the same laboratory method/population as the receiving laboratory. In the first instance, comparing the laboratory methods serves as an instructive early screening tool to assess the suitability of the reference values for the receiving laboratory. Laboratory methods can be compared by a method comparison study between the method used during the development of the RI and the method used by the receiving laboratory to determine the statistical validity of an RI transfer [11]. For a method comparison study, samples must be collected with an appropriate distribution of values spanning the RI, as an insufficient range may underestimate and a range too large may overestimate the strength of the correlation. The correlation between the two methods is subsequently analyzed and, if appropriate, linear regression analysis is performed to determine the slope and y-intercept values of the best-fit regression line [12]. These values are subsequently used to transfer the RI. According to the CLSI EP28-A3c guideline, the best-fit regression line should have a slope bias close to 1, a y-intercept close to 0 and a correlation coefficient (r^2) close to 1 [6]. Furthermore, according to CLSI EP09-A3

guidelines, the scatter and bias plots should be examined for constant scatter to ensure there are no dramatic differences between the variation at the upper and lower ends of the range of values [12]. To sufficiently assess the acceptability of the method bias, it is also important that the magnitude of the y-intercept is small compared to the range of the data and the RI. If the y-intercept is large compared to the RI, it is recommended to reject transference and establish an RI directly from a healthy reference population.

If the preanalytical processes (e.g. preparation of reference individuals, specimen collection, transportation, and handling), the laboratory methods and the populations (e.g. a relatively homogenous population within the same geographical region) are very similar to those of the laboratory where the RIs originated, the method comparison study is still recommended to confirm the comparability, although the bias between the laboratory methods is expected to be very small [12]. However, in this situation, subsequent verification using samples from healthy reference individuals may not be necessary, and the laboratory may opt to simply perform a subjective assessment by carefully inspecting the reference population demographics, geographic location, preanalytical and analytical procedures, analytical performance and the statistical methods used in the RI study. If these factors are all consistent with the receiving laboratory's population and procedures, the RI may be transferred without further verification [6].

Following transference, the CLSI EP28-A3c guideline recommends subsequently verifying the transferred RI. The guideline emphasizes that three approaches can be used to verify RIs: (1) a subjective assessment, (2) using a small number of reference individuals (e.g. n = 20) and (3) using a large number of reference individuals (e.g. n = 60, but fewer than 120) [6]. Using a large number of reference individuals is not generally preferred by routine laboratories, as this is nearly the same as the sample size required for an RI study. The standard approach recommended by the guideline for routine practice in laboratories is to collect and analyze samples from 20 healthy subjects per age and/or sex partition from the receiving laboratory's local population and to compare these reference values with the RI established from the larger, more robust, original study (Figure 1). The Reed/Dixon [13, 14] or Tukey methods [15] should be applied to test and subsequently remove outliers, and new specimens should be obtained to replace those removed. If no more than 2 of the 20 samples (i.e. 10% of the test results) fall outside the RI, it may be received for use, at least provisionally. If 3 or 4 of the 20 samples fall outside the RI, a second set of 20 reference specimens should be obtained. If again 3 or more

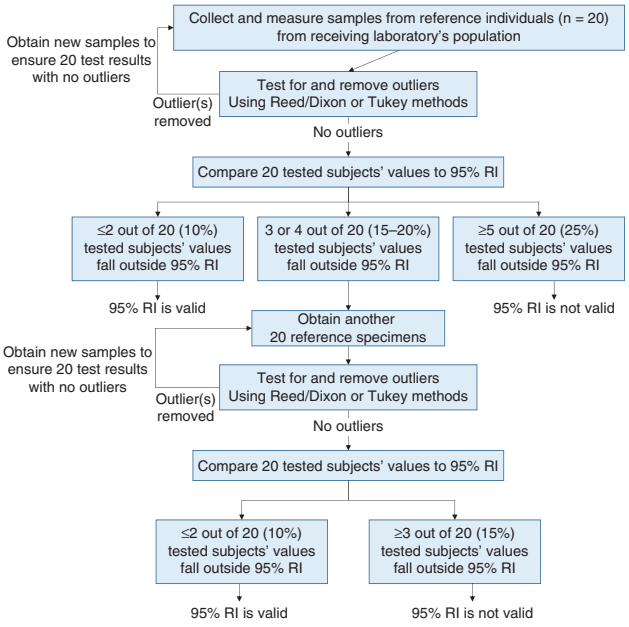

Figure 1: Verification of reference intervals as recommended by CLSI EP28-A3c guidelines using a statistical test on a relatively small number of reference individuals.
After confirmation that the preanalytical factors, analytical factors, and local populations are consistent between the reference interval study and receiving laboratory, this procedure to verify reference intervals can be followed. If the 95% reference interval (RI) is not considered to be valid, the analytical procedures used and population differences should be reexamined. If these factors differ substantially, the receiving laboratory should consider developing its own reference intervals (CLSI EP28-A3c).

of the new specimens (i.e. ≥10% of the test results) or 5 or more of the original 20 fall outside the RI, the user should reexamine the analytical procedures used and consider possible differences in the biological characteristics of the two populations sampled (Figure 1). As an example of this approach, after transferring RIs to additional analytical platforms in the study, these RIs were subsequently verified using samples from approximately 100 healthy reference individuals across all partitions, ensuring a minimum sample size of 20 in at least one partition [16, 17]. The verification procedure recommended by CLSI EP28-A3c guidelines was followed, with the exception of collecting additional RIs if verification of the initial set of reference samples failed. Thus, if the transferred RI failed the initial verification, it was simply concluded to not successfully verify and was not recommended for use on the analytical platform under investigation.

Although this guideline appears straightforward, RI partitions, the presence of outliers and initial unsuccessful verification can further complicate the verification process. In clinical laboratories, some parameters require

age and/or sex partitions, whereas others do not. The CLSI EP28-A3c guideline states that if one RI partition successfully verifies, there may not be a need to verify all other partitions. However, whether successfully verifying one partition is adequate to consider all partitions verified may depend on analyte levels, forms of the measured analyte and other matrix issues, which can vary across partitions [18]. Therefore, although it might be operationally impractical to verify all partitioned RIs, verifying each partition would provide the most confidence in accepting external RIs. If a laboratory decides to verify each partition with 20 samples from healthy reference individuals, the number of samples required for verification will differ substantially depending on the number of age and/or sex partitions required for a given analyte. Further clarity is also needed for the limit of reference values falling outside the transferred RI in subsequent partition(s), following successful verification of one partition. However, it is important to note that the distribution of reference samples used for verification is as imperative as the number of samples that fall outside of the RI. For example, if 2/20 of the samples fall outside one of the reference limits, with the remaining samples evidently biased towards the same side, the suitability of the RI for the receiving laboratory would also be called into question. Lastly, an RI should not automatically be accepted when ≥90% of healthy reference samples fall within the RI, as this could also be indicative of the RI being too wide. This again highlights the importance of considering the distribution of healthy reference samples when performing verification analysis.

Establishment of common RIs and examples for transferring and verifying RIs from multicenter studies

To enable laboratories to transfer and verify RIs for clinical use, accurate and robust RIs from large, direct *a priori* RIs are first required [6]. Health-associated RIs can be directly established using a healthy, reference population, defined using strict inclusion/exclusion criteria. Several statistical methods can be used to calculate RIs (e.g. parametric, non-parametric, robust), with their accuracy dependent on the sample size and underlying distribution of the data [19]. For simplicity, the CLSI EP28-A3c guideline recommends a minimum of 39 samples per RI partition using robust methods, a minimum of 120 samples are recommended to establish RIs using the non-parametric method, as this is the requirement to calculate 90% confidence intervals around the upper and lower reference limits [6].

Although direct RIs are most commonly established using a well-defined and representative reference population, with sample analysis performed by a single laboratory, RIs can also be determined with the intention of serving a much broader population demographic and/or geographic location with sample analysis performed by a single platform or multiple platforms, termed common RIs. Once common RIs have been established from a multicenter study, the process of transference and verification can be undertaken, on the consideration of preanalytical, analytical and local population differences to ultimately implement RIs in a clinical laboratory [20]. The Committee on Reference Intervals and Decision Limits (C-RIDL) conducted a global RI study on behalf of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) [7], with 12 participating countries worldwide. This is an example of a common RI study, in which each laboratory acts as a central laboratory and sample analysis is performed using multiple platforms. In this type of multicenter study, it is essential to perform rigorous quality control monitoring to detect analytical deviations and use internationally accepted reference materials for standardized analytes to ensure traceability in each center. In addition to internationally accepted reference materials, in the global IFCC, C-RIDL study, a serum panel with assigned values was measured by all laboratories, and the panel of sera strategy was used to compare results obtained from the participating centers [21]. The basic scheme for conducting the global study was to make test results comparable among countries based on the panel test results measured in each participating country [8]. This approach resulted in a method comparison and successful transference of the data obtained from the global study. As part of this global study, a multicenter study was also performed in Turkey, including seven geographical regions, using traceable materials and panel of sera from 40 reference individuals from the global study in the central laboratory, using a single platform (Abbott ARCHITECT) [22]. After obtaining common RIs for Turkey for biochemical parameters, Ozarda et al. performed additional "cross-check testing" using at least 20 samples to compare results among the participating laboratories in Turkey as recommended in the protocol for multicenter studies [23]. Thus, RIs were transferred from the multicenter study to each participating laboratory in Turkey using the linear regression slope and intercept [12].

The CALIPER [11, 24–27], the Canadian Health Measures Survey (CHMS) [28–30] and the Aussie Normals study

[31] are examples of studies where the measurements were performed in one center acting as the central laboratory. These were large, well-conducted, direct studies, with sample analysis performed on a single platform. CALIPER also performed studies to transfer the pediatric RI database from the initial single platform to several other commonly used platforms. In the first of several transference studies, CALIPER transferred pediatric RIs from the Abbott ARCHITECT assays/platforms to four other commonly used chemistry platforms by performing method comparisons using 200 pediatric serum specimens obtained from leftover pediatric sera that covered most age and sex partitions [11, 16]. RIs were successfully transferred if the methods sufficiently correlated (i.e. $r^2 > 0.70$), and the linear model was deemed appropriate in respect to bias and normality of residuals.

The United Kingdom (UK) Pathology Harmony Reference Interval project [32] and the RI harmonization initiative led by the Australasian Association of Clinical Biochemists (AACB) Committee for Common Reference Intervals [33] are examples of an expert panel determining conventional RIs using an evidence-based approach. For example, the AACB Committee for Common Reference Intervals developed and followed a standard protocol to develop harmonized RIs for 11 and 10 chemistry analytes for adults and pediatrics, respectively [9]. RIs in adults were first obtained from a local RI study (i.e. the Australian Aussie Normals study) then compared with RIs obtained from various sources, including published direct RI studies and indirect data mining of large databases (e.g. Sonic Healthcare). Method differences were first assessed for clinical acceptability using various approaches, including a method bias study using commutable samples measured across 24 laboratories and eight major chemistry platforms in use throughout Australia and New Zealand, examination of the reference measurement system/procedure in place for each analyte and the average of normals obtained from participating laboratories. In the same study, harmonized pediatric RIs were established based on data mining and consensus agreement across 31 pediatric laboratories within Australasia [9].

Additional approaches to verify RIs

There are additional approaches to verify RIs other than those recommended by the CLSI EP28-A3c guidelines. An approach using healthy reference individuals would be to perform a local RI study with sufficient power to establish confidence intervals around the upper and lower limits. A more robust RI established from an external source can then be verified if its reference limits are within the confidence intervals established by the local RI study [34].

Furthermore, indirect data mining methods can be applied to the laboratory's existing data to verify RIs established by an external source. Advantages of the indirect method include large data sources that are more easily accessible, analysis that is directly targeting the local population and preanalytical and analytical factors that reflect those used in the local laboratory [35]. However, the population also inevitably includes unhealthy subjects, and a verification concept has not yet been fully developed for indirect methods [36]. Therefore, valuable efforts are being made to develop robust methods and criteria to apply indirect techniques for verification purposes [36]. The most useful parameter is the midpoint (i.e. median) of the extracted data, which can be used to assess analytical or population bias through comparison with the corresponding midpoint of the data used to set the RI, if data are skewed, or the midpoint of the RI itself, if data are Gaussian distributed [35]. The midpoint is highly resistant to the presence of results from unhealthy subjects and has less uncertainty compared to reference limits [35]. Data mining of local population values also allows for an assessment of the number of results outside the RI and monitoring of the percentage of abnormal results, which would typically be flagged by the laboratory information system [9]. The laboratory can then compare the expected flagging rates with their current rates derived from the original indirect study calculations. This method can be programmed in the laboratory information system as a continuous quality control monitoring measure. When the increase of flagging in any direction does not exceed the predefined quality goals, the RI under evaluation can be considered acceptable for use [37]. Based on the principle of minimum, desirable and optimal categories used to define allowable bias limits, flagging rates may range from 1.0% to 1.8% and 5.7% to 3.3% for low and high flagging rates, respectively [38]. It is important to note that these flagging rates may be expected to be exceeded when applying them to a pathology population. Furthermore, as common RIs suitable for use in numerous laboratories are generally wider than those expected for a single laboratory, this may lead to lower than expected flagging rates in individual laboratories.

Verification of RIs in pediatric and geriatric populations

It is well known that the determination of pediatric RIs is an extremely difficult task, primarily due to ethical

limitations related to obtaining blood samples from neonates and very young children. The most significant milestones in this area have been achieved by the CALIPER Project, a Canada-wide initiative that aims to address this current gap by establishing a comprehensive database of age- and sex-specific pediatric RI (www.caliperproject.ca). Through robust transference and verification studies, the CALIPER database is now applicable for all five commonly used analytical platforms [11].

The major difficulty in obtaining geriatric RIs is the selection of healthy individuals, as most elderly subjects do not meet the criteria of the CLSI EP28-A3c guideline for inclusion in a healthy reference population [39]. The width of the RI 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, this issue remains inadequately addressed [40]. It would be of great benefit to conduct a large, multicenter study with pediatric, adult and geriatric reference individuals to develop common RIs, subsequently transfer them to local laboratories and verify them with respect to these specific age-groups using a limited number of healthy subjects and/or existing laboratory data [11]. The most significant contribution in this area has been from the CHMS study, in which RIs were established for Canadians 3–<80 years of age [28–30]. The same exclusion criteria to ensure the reference population is void of unhealthy subjects was used for all age-groups, which may not be appropriate for the geriatric population, in which the use of medications and certain health conditions are much more prevalent. Furthermore, partitioning was determined based on age and sex, without taking into account the difference between chronological and biological age in the geriatric population.

Conclusions

Clinical laboratories are recommended to determine their own RIs. However, this could be difficult in routine practice due to the heavy workload of clinical laboratories. Therefore, it is important that routine laboratories transfer and verify their RIs from an external source before applying them for clinical use. The RIs can be received from various external sources, including manufacturers' package inserts, studies published in literature, local RI studies, multicenter studies, laboratory surveys, relevant guidelines, consensus statements and mining of databases. If there are available RIs obtained from local/multicenter studies originating from the same population, the laboratories are first encouraged to receive these RIs. As multicenter studies are performed using large, well-defined reference individuals, and consequently result in reliable and robust data with narrow confidence intervals, transferring and verifying common RIs obtained from multicenter studies is now recognized as an important step in receiving RIs for use in another laboratory. If similar methods and reference populations have been used, the CLSI guidelines and literature encourage a method comparison with values spanning the width of the RIs to receive the same RI at a local or national level. If no bias has been detected by the method comparison study, verification of the transferred RI using samples from healthy individuals from the local population is recommended. The standard verification procedure outlined in the CLSI EP28-A3c requires a minimum of 20 samples from healthy subjects and is very clear in theory, although gaps remain for the implementation of these procedures in routine clinical laboratories. Additional verification procedures, including indirect methods, have become more widely used in recent years in this field, as they are inexpensive and easy to perform. IFCC, C-RIDL has also been very eager to use data mining approaches for assessment and verification of RIs in recent years.

The ISO 15189 Standard for clinical laboratory accreditation states that each laboratory should periodically re-evaluate its own RIs, although no consensus is provided on how often they should be reevaluated in routine laboratories. Should they be reevaluated annually or every few years? It should also be noted that gaps remain in pediatric and geriatric RIs, including specific guidelines for their establishment and verification. Taken together, a detailed protocol for transference and verification of RIs is warranted for routine clinical laboratories.

Acknowledgments: We had invaluable input from members of the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL).
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. Horn PS, Pesce AJ. Reference intervals an update. Clin Chim Acta 2003;334:5–23.
2. Directive 98/79/EC of European Parliament and the Council of 27 October 1998 on in vitro diagnostic medical devices. Offical J Eur Commun 7 December 1998; L331/1-L331/37.
3. International Organization for Standardization. ISO 15189:2012 (E) Medical laboratories – Requirements for quality and competence, Third edition, 2012-11-01.
4. Ceriotti F, Hinzmann R, Panteghini M. Reference intervals: the way forward. Ann Clin Biochem 2009;46:8–17.
5. Tate JR, Yen T, Jones GR. Transference and validation of reference intervals. Clin Chem 2015;61:1012–5.
6. CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; Approved Guideline – Third Edition. CLSI EPC28-A3c. Wayne (PA): Clinical and Laboratory Standards Institute, 2010.
7. Ichihara K, Ozarda Y, Barth JH, Klee G, Shimizu Y, Xia L, et al.; Committee on Reference Intervals and Decision Limits, International Federation of Clinical Chemistry and Laboratory Medicine and Science Committee, Asia-Pacific Federation for Clinical Biochemistry. A global multicenter study on reference values: 2. Exploration of sources of variation across the countries. Clin Chim Acta 2017;467:83–97.
8. Ichihara K, Ozarda Y, Barth JH, Klee G, Qiu L, Erasmus R, et al.; Committee on Reference Intervals and Decision Limits, International Federation of Clinical Chemistry and Laboratory Medicine. A global multicenter study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. Clin Chim Acta 2017;467:70–82.
9. Tate JR, Sikaris KA, Jones GR, Yen T, Koerbin G, Ryan J, et al.; on behalf of the AACB Committee for Common Reference Intervals. 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.
10. Nichols JH. Verification of method performance for clinical laboratories. Adv Clin Chem 2009;47:121–37.
11. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian Laboratory Initiative on Pediatric Reference Intervals: a CALIPER white paper. Crit Rev Clin Lab Sci 2017;54:358–413.
12. CLSI. Method procedure comparison and bias estimation using patient samples, approved guideline, 3rd ed. Wayne (PA): CLSI, 2013. CLSI document EP09-A3.
13. Dixon WJ. Processing data for outliers. Biometrics 1953;9:74–89.
14. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. Clin Chem 1971;17:275–84.
15. Tukey JW. Exploratory data analysis. Boston (MA): Addison-Wesley, Reading (PA), 1977:1–688.
16. 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;46:1197–219.
17. Higgins V, Chan MK, Nieuwesteeg M, Hoffman BR, Bromberg IL, Gornall D, et al. Transference of CALIPER pediatric reference intervals to biochemical assays on the Roche cobas 6000 and the Roche Modular P. Clin Biochem 2016;49:139–49.
18. Sikaris KA, Yen T. CALIPER: Editorial. Supporting the steps forward in pediatric laboratory measurement. Clin Biochem 2013;46:1195–6.
19. Daly CH, Higgins V, Adeli K, Grey VL, Hamid JS. Reference interval estimation: methodological comparison using extensive simulations and empirical data. Clin Biochem 2017;50:1145–58.
20. Ozarda Y. Reference intervals: current status, recent developments and future considerations. Biochem Med 2016;26:5–16.
21. 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.
22. 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.
23. Ozarda Y, Ichihara K, Barth JH, Klee G; on behalf of the CRIDL on Reference intervals and Decision Limits, IFCC. Protocol and standard operating procedures for common use in worldwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1027–40.
24. Adeli K. Closing the gaps in pediatric reference intervals: the CALIPER initiative. Clin Biochem 2011;44:480–2.
25. Colantonio DA, Kyriakopoulou L, Chan MK, Daly CH, Brinc D, Venner AA, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem 2012;58:854–68.
26. Bailey D, Colantonio D, Kyriakopoulou L, Cohen AH, Chan MK, Armbruster D, et al. Marked biological variance in endocrine and biochemical markers in childhood: establishment of pediatric reference intervals using healthy community children from the CALIPER cohort. Clin Chem 2013;59:1393–405.
27. Karbasy K, Lin DC, Stoianov A, Chan MK, Bevilacqua V, Chen Y, et al. Pediatric reference value distributions and covariate-stratified reference intervals for 29 endocrine and special chemistry biomarkers on the Beckman Coulter Immunoassay Systems: a CALIPER study of healthy community children. Clin Chem Lab Med 2016;54:643–57.
28. 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.
29. 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.
30. 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.
31. 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.
32. Berg J, Lane V. Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem 2011;48:195–7.
33. Koerbin G, Sikaris KA, Jones GR, Ryan J, Reed M, Tate J; on behalf of the AACB Committee for Common Reference Intervals. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014;432:99–107.
34. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev 2007;28:115–21.
35. Jones GR. Validating common reference intervals in routine laboratories. Clin Chim Acta 2014;432:119–21.
36. Haeckel R, Wosniok W, Arzideh F, Zierk J, Gurr E, Streichert T; on behalf of the working group Guide Values (Richtwerte) of the German Society for Clinical Chemistry and Laboratory Medicine (DGKL-WG). Critical comments to a recent EFLM recommendation for the review of reference intervals. Clin Chem Lab Med 2017;55:341–7.
37. Fraser CG. Biological variation: from principles to practice. Washington, DC: AACC Press, 2001.
38. Tate J, Koerbin G, Adeli K. Opinion paper: deriving harmonised reference intervals – global activities. eJIFCC 2016;27:48–65.
39. 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.
40. 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.