Reference values: a review

Anne Geffré[1], Kristen Friedrichs[2], Kendal Harr[3], Didier Concordet[4], Catherine Trumel[1], Jean-Pierre Braun[1,4]

[1]Département des Sciences Cliniques, Ecole Nationale Vétérinaire de Toulouse, Toulouse, France; [2]Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, WI, USA; [3]IDEXX Laboratories, Vancouver, British Columbia, Canada; and [4]UMR181 Physiopathologie & Toxicologie Expérimentales, INRA, ENVT, Toulouse, France

Key Words
Healthy population, nonparametric, normality, reference individual, reference interval, selection criteria

Abstract: Reference values are used to describe the dispersion of variables in healthy individuals. They are usually reported as population-based reference intervals (RIs) comprising 95% of the healthy population. International recommendations state the preferred method as a priori nonparametric determination from at least 120 reference individuals, but acceptable alternative methods include transference or validation from previously established RIs. The most critical steps in the determination of reference values are the selection of reference individuals based on extensively documented inclusion and exclusion criteria and the use of quality-controlled analytical procedures. When only small numbers of values are available, RIs can be estimated by new methods, but reference limits thus obtained may be highly imprecise. These recommendations are a challenge in veterinary clinical pathology, especially when only small numbers of reference individuals are available.

Introduction

The concept of reference values was introduced in 1969 by Grasbeck and Saris[1] to describe fluctuations of blood analyte concentrations in well-characterized groups of individuals. It was intended to replace the more ambiguous concept of normal values,[2,3] and to "establish a well-defined nomenclature and recommended procedures in the field."[1] In this first publication, there was a clear distinction between healthy reference values measured in healthy populations or individuals and patient reference values measured in patients having various diseases. It is now commonly accepted that reference values describe fluctuations observed in healthy populations or individuals, which makes the definition of health or characterization of health status a critical step.

Reference values, first introduced as a philosophy, have gained universal acceptance as one of the most powerful tools in laboratory medicine to aid in the clinical decision-making process.[3–5] However, the recommendations for establishing *reference interval*s (RIs) described in the original series of articles published by the International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine[6–11] were sometimes considered too complicated to be applicable in practice; and thus, they have been used erroneously, if used at all. For instance, a recent survey of RIs for serum

creatinine in humans identified 37 reports of which only 6 met IFCC criteria.[12] These difficulties have led to a necessary revision of the original recommendations[13,14] and the publication of common IFCC and Clinical Laboratory and Standards Institute (CLSI) guidelines (C28-A3) in 2008.[5] In the latter document, previous recommendations are reinforced, which were to establish RIs with at least 120 reference individuals using the nonparametric ranking method. However, it is also acknowledged that RI determination is difficult, time-consuming, and expensive, and therefore, "it is unrealistic to expect each laboratory to develop its own RIs." The new document now allows individual laboratories to adopt, by transference and verification, RIs established elsewhere. Additionally, alternate statistical approaches, such as the robust method, make it possible to establish RIs using smaller reference sample sizes; however, "the working group is hesitant to recommend that it be done (with fewer than 80 observations), except in the most extreme instances."[5]

The present review is based on C28-A3 and a MEDLINE search on the theory and production of reference values in humans and animals. Only a few of the numerous articles on this subject (126,242 hits for "reference values" in February 2009) have been selected for this review. General information on reference values can be found in textbooks,[15,16] chapters in human[17] and veterinary[18] clinical pathology texts, a special issue of *Clinical Chemistry and Laboratory Medicine*,[7] and the RefVal computer program of statistical calculations.[19,20]

Nomenclature

Standard terms and definitions

The latest definitions cited from C28-A3[5] differ slightly from the previous IFCC document,[6] but the overall relationships between the terms are the same (Figure 1).

A *reference individual* is a person selected for testing on the basis of well-defined criteria. Reference individuals are generally assumed to be "healthy"; however, health is relative and lacks a precise and quantifiable definition. Therefore, reference individuals are selected using "well-defined criteria," ie, inclusion and exclusion criteria, which approximate health. Inclusion and exclusion criteria should be defined precisely, according to the aims of the study, and may differ from one study to another. The RI determined from the individuals selected according to the given criteria will be applicable only to similar individuals, ie, only to individuals fulfilling the same criteria.

A *reference population* is a group consisting of all possible reference individuals.

A *reference sample group* is an adequate number of persons selected to represent the reference population. Although meant to be representative, the characteristics of a reference sample group are not identical to the characteristics of the reference population for the following reasons. First, the reference population is hypothetical because the number of individuals it comprises is unknown. Second, the reference sample group rarely is selected in a completely random manner.

A *reference value* is the value, or test result, obtained by the observation or measurement of a particular type of quantity on a reference individual. A "particular type of quantity"[21] ("measurand" in metrology and "component" or "analyte" in laboratory medicine)[22] implies that most of the theory and application of reference values deals with univariate RIs, ie, only 1 analyte at a time, whereas interpretation of results is mostly multivariate. This has led some authors to study multivariate reference regions, which at this time have only limited development.[23,24] A reference value, which represents 1 value obtained in 1 reference individual, is not synonymous with a reference limit, which is a value derived from all results obtained in the reference sample group. The term reference value should not be used to denote a limit of the RI.

A *reference distribution* is the distribution of reference values.

Reference limits are the values derived from the reference distribution and are used for descriptive purposes. Reference limits should not be confused with decision limits, which are defined below.

An RI is the interval between, and including, 2 reference limits. The RI comprises only a fraction of the values measured in reference individuals, most frequently the central 95% of the distribution located between the 0.025 and 0.975 fractiles as defined by ISO 15189 and IFCC.[10,25] As a consequence, 5% of healthy individuals have observed values above or below these reference limits. In other words, it is perfectly normal to observe abnormal results in healthy individuals – it just is not frequent. The term "reference range," often used as a synonym for RI, is not defined in C28-A3 and therefore should not be used interchangeably.

An *observed value*, or patient laboratory test result, is the value obtained in a test subject that is compared with reference values, reference distributions, reference limits, or RIs.

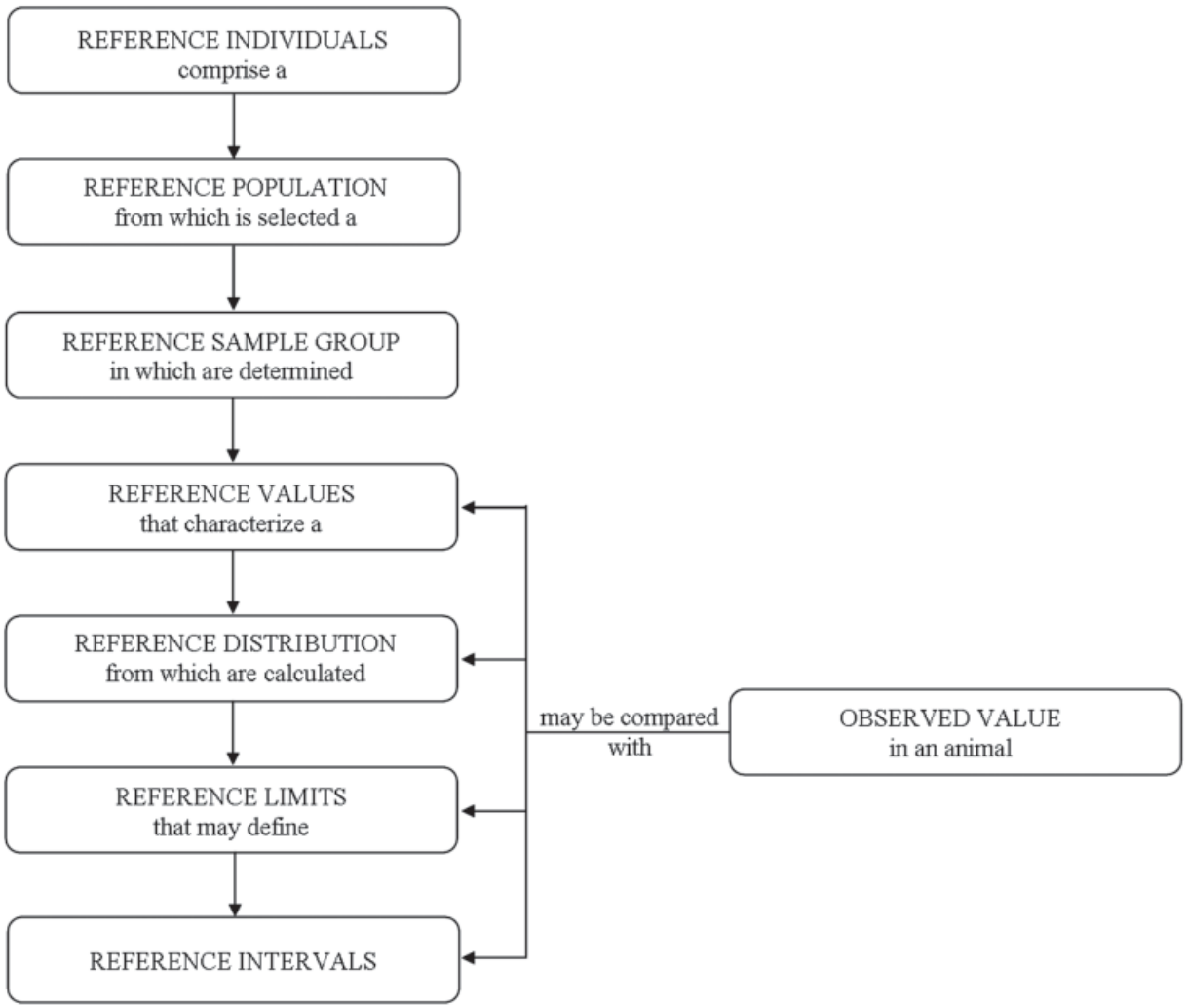

Figure 1. Relationships between the terms related to reference values according to the Clinical Laboratory and Standards Institute (CLSI) and International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine document C28-A3.[5]

Other potentially confusing terms

Individual RIs are derived from a single individual and are narrower than population-based RIs.[26] Comparing repeated measurements to the individual RI allows more efficient interpretation.

Reference change is the difference between 2 successive values that would be significant ($P \leq 05$) in 95% of such persons.[27] It is based on the "critical range"[28] (or critical difference) observed in an individual and encompasses both intra-individual and analytical variability. A reference change is the most effective approach by which to detect significant changes within an individual. Because population-based RIs primarily comprise interindividual variability, they are much too wide to detect reference changes in an individual.[29,30] Because unpredictable and extreme changes can occur in diseased individuals due to disease progression or resolution, critical differences, in combination with intraindividual reference values, usually are evaluated only in apparently healthy individuals.[26,31]

Decision limits (cut-offs, cut-points, or consensus values[32]) are thresholds used to classify patients into diseased vs. non-diseased states or to identify when medical action is advised, regardless of the reference limit.[33] Decision limits are commonly used in human medicine for the diagnosis of specific conditions or risk factors, eg, fasting plasma glucose concentration for the diagnosis of diabetes mellitus, or urine protein:creatinine ratio in dogs and cats.[34]

A *parameter* is a quantity that defines certain characteristics of a population (eg, the mean of a

population) and does not vary among individuals. Plasma glucose concentration and alkaline phosphatase activity are not parameters, whereas temperature is a parameter of phosphatase activity measurement. This word is unduly used in place of "variable."[35]

A *variable* is a quantity that varies within or between individuals and is often confused with parameter.[35] For instance, RBC or plasma cholesterol concentration are variables.

A *confidence interval* (CI) contains, within a given probability, the value of an unknown population parameter. Because reference limits are determined from only a sample of the population, they are estimates of the true limits, which cannot be known; CIs indicate the imprecision of that estimate. The larger the reference sample size, the more closely the reference sample group approximates the reference population and the narrower the CI.

Prediction interval, a statistical term that has the same meaning as RI, contains a given percentage of values of a variable that can be observed in individuals from a population.

A *tolerance interval* is an interval within which a specified proportion of a population falls with a specified confidence. It is based on the CIs of limits of a prediction interval. Tolerance interval, RI, and CI of limits are schematically compared in Figure 2.

Inclusion and exclusion criteria establish whether a subject is eligible to participate in an RI study. These criteria are chosen so that only healthy individuals are included; individuals that are diseased, or do not belong to the reference population for whom an RI is being established, are excluded. Some exclusion criteria, eg, pregnancy and age, can serve as partitioning criteria. For reference individuals, inclusion and exclusion criteria can be applied a priori, before the collection of samples, or a posteriori, after the collection of samples. Inclusion and exclusion criteria should be determined before selecting reference individuals or reference samples from a database. Conformity to selection criteria may be established by physical examination, certain measurements or diagnostic tests, and/or completion of a questionnaire by the person in charge of the study with the client.[5]

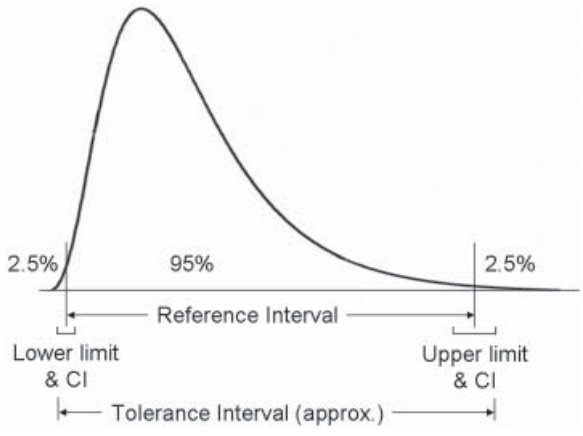

Figure 2. Schematic representation of a reference interval, reference limits, confidence intervals (CI) of the limits, and tolerance interval.

Reference Values and Health Status

From its inception, and according to IFCC definition, reference values are measured in a well-characterized population of individuals selected according to predefined criteria such as age, sex, breed, nutritional status, and diet. In addition, it is presumed that reference individuals are healthy, which raises the question of the definition of health. There is no accepted consensus on the definition of health. The World Health Organization definition[36] is inadequate even for humans and is not transferable to animals because it is impossible to define objective criteria to characterize "complete physical, mental, and social well-being." As a consequence, the initial and probably the most problematic step in the determination of an RI is defining the criteria used to characterize health.[37] These criteria must be clearly described and documented, "so others can evaluate the health status of the reference sample group."[5]

Determination of a Reference Interval

General approaches

There are 3 possible means by which to obtain the RI of a given analyte for a given population:

(1) determine the RI de novo from measurements made in reference individuals;
(2) transfer a pre-existing RI when a method/instrument is changed; or
(3) validate a previously established or transferred RI.

De novo determination of RIs is the most frequently used procedure in medical and veterinary laboratories, as indicated in the original IFCC recommendations. An a priori approach is recommended in which reference individuals are selected according to predefined criteria followed by determination of RIs from the reference values obtained. This approach is most often performed in a single laboratory, but a multicentric procedure also is possible if methods and populations are comparable. In some cases, an a posteriori approach is used in which pre-existing data is exploited to establish reference values. Because inclusion

and exclusion criteria are applied retrospectively, the necessary information regarding selection criteria may not be available.

Stepwise procedures for a priori determination of a reference interval

The details of the procedure are given in the IFCC-CLSI C28-A3 guidelines.[5] The 13 steps in that document can be summarized as follows below. All of the steps and procedures should be fully documented.

Fully document preanalytical, analytical, and biological factors of variation

The preanalytical, analytical, and biological factors of variation for each analyte should be determined by a literature search. Control of clinically meaningful factors of variation will minimize variability of the results obtained. Some factors of variability may be used as exclusion or partitioning criteria (eg, pregnancy). It may be difficult to control some preanalytical factors of variation in reference subjects, such as fasting (when animals are presented for a wellness examination) or stress in cats. It is difficult to objectively evaluate stress and to make decisions regarding the degree of stress that is tolerable in reference subjects. This is especially true for wild animals in which the level of stress is quite different, for example, in animals bred in zoos compared with those caught in the wild.

Establish inclusion and exclusion criteria and partitioning factors

The objective of the future use of the RI is critical, because it is the basis for defining the characteristics of the population to be studied and thus, for choosing inclusion, exclusion, and partitioning criteria for the selection of individuals. Minimal criteria of exclusion include any clinical sign of disease or administration of medications, perhaps with the exception of anthelminthics. Other quantifiable exclusion factors that indicate poor health or undue stress can be added, such as a body temperature, heart rate, or body condition score above a certain level. A questionnaire with simple questions requiring unambiguous answers can aid in categorizing individuals. An example of such a questionnaire for humans is given in C28-A3[5] and can be adapted to veterinary clinical pathology. Once the questionnaire is completed by the client and the investigator, the reference subjects undergo a physical examination and other testing as necessary or indicated by the selection protocol. Selected reference individuals are then categorized or excluded based on the exclusion criteria or evidence of poor health. The selection of reference individuals should not be too restrictive nor should reference individuals consist only of healthy young adult animals. All subjective and objective assessments should be recorded and included in the reference study document.

Decide on an appropriate number of reference individuals

The appropriate number of reference individuals should be determined according to the desired CI of the reference limits. One-hundred and twenty is the recommended minimum number of individuals in the reference sample group because it is the smallest number from which it is possible to estimate the 90% CIs of the reference limits using the nonparametric method.[10,38] The number of reference values necessary to achieve a given CI using nonparametric methods is much higher than by parametric methods, with the highest numbers required in cases of pronounced skewness.[39–41] In some animal populations (eg, exotic species) it is extremely difficult to achieve these recommended reference sample sizes; however, it is still advised that "the number of samples should be as high as possible"[5] without indicating a minimum number.

Prepare the reference individuals and collect and process the specimens

Preparation of selected reference animals, if necessary, should be made according to information collected in the first steps (preanalytical, analytical, and biological factors of variation and inclusion/exclusion criteria). Specimen type, collection method and specimen handling and processing should be standardized and the same as for patient specimens. Specimens handled improperly or of poor quality (eg, hemolyzed specimens or samples that have not been stored appropriately) should be rejected.

Analyze the specimens with a quality-controlled method

Reference specimens should analyzed in the same manner as for patient specimens. Quality management of analytical methods is critical for the reliability of the values obtained.[42,43] "The methods used must be described in detail, reporting between-run analytical imprecision, limit of detection, linearity, recovery, and interference characteristics, but especially its trueness and the demonstration of traceability of results provided to higher order methods or materials, when they exist."[5,44,45] Some experts advise that these specifications be communicated to clinicians to aid in the interpretation of results,[46] whereas CLSI only

recommends that this information be available upon request.

Inspect and analyze the data

The reference data should be inspected, the data, a histogram prepared, possible errors or outliers identified, and the reference limits and their CIs determined. It is generally agreed that the RI should cover the central 95% of the reference samples collected, limits thus being the 0.025 and 0.975 fractiles. However, some scientists suggest that alternative limits be considered, especially the 0.999 fractile for routine health evaluations, which would limit the number of false positive results.[47]

Reference limits should be determined by the nonparametric method. However, parametric estimation can be used when data fit or can be transformed to fit a Gaussian distribution.[41,48,49] Transformation is followed by a goodness-of-fit test, such as Anderson-Darling's.[50] Even with transformed data, parametric estimation of the 0.975 fractile may be biased when distributions are highly skewed to the right.[51] Some authors recommend comparing RIs obtained by several statistical approaches, for example, the nonparametric method, a parametric method with transformed data, and other methods, such as robust or bootstrap. If estimates of reference limits are dissimilar, the data set may be heterogeneous, ie, contain individuals that do not belong to the underlying population.[52,53]

There are especially 2 difficult issues that need to be addressed at this stage: outliers and partitioning. Outliers are values that do not truly belong to the reference distribution. Their detection and removal is critical because "unless the number of samples is extremely large, normal range estimation by nonparametric methods almost entirely depends on one or two lowest and highest values."[38] However, one has to be careful to avoid the temptation to eliminate too many values just to smooth a curve, because the deleted values may belong to the underlying distribution. "Unless outliers are known to be aberrant observations. . .the emphasis should be on retaining rather than deleting them."[5] In addition to visual examination of the histogram, the most frequently used outlier tests are Dixon–Reed's and Tukey's,[38] which are relatively straightforward. Tukey's method can be performed accurately in the presence of multiple outliers, whereas the Dixon–Reed test can only be used when 1 outlier is suspected.[5] However, some authors believe even these methods are insufficient, and that no method optimally detects all outliers.[54,55]

Partitioning of RIs into subclasses, often based on sex or age, should be considered if it is clinically useful or based on physiology.[5] However, "any observed difference, no matter how small or how questionable its clinical significance, can be statistically significant if the sample sizes are large enough."[56] The shape of the distribution[57] and/or the prevalence of the subclasses[58] also may contribute to significant differences between subclasses even when the means are identical. There is no consensus on the criteria used to decide whether partitioning is or is not relevant.[59,60] The IFCC-CLSI C28-A3 guideline recommends use of Harris-Boyd's *z*-test although it is limited to comparisons between 2 subclasses.[5,61] An alternative method for the production of covariate-dependent RIs (eg, the effect of age) is the use of regression-based reference limits, which require very large sample sizes but avoid dividing the values into subclasses.[62,63] These have not been used in veterinary clinical pathology to our knowledge.

Other options for the determination of a reference interval

A posteriori determination

When it is too difficult to apply the full a priori procedure, it may be necessary to use values selected from a databank. However, the same preanalytical, analytical, and selection factors outlined above should be applied, and all population and health data should be available for inspection. The only difference is that the selection of reference individuals is made after analysis has been performed.

Indirect determination

As in the a posteriori method, the indirect sampling technique relies on large databases consisting of both healthy and diseased individuals, such as hospital records.[64] Indirect methods should not be used except when no other option is available, due to the likelihood of erroneous values. Extreme caution must be used, and clinicians should be warned that these RIs are more likely to contain abnormal values due to generation from patient databases that contain diseased individuals.

Indirect determination of RIs is based on mathematical methods that separate, as efficiently as possible, healthy from unhealthy individuals. Extracted data then are used to estimate RIs. This approach is less reliable when distributions have large skewness and/or kurtosis.[65] Indirect methods probably will have limited use in veterinary clinical pathology where only few large databanks are available. To our knowledge, this has only been used once for serum biochemistry in sheep,[66] and a new method has been proposed recently.[67]

Estimation from small sample sizes

Small sample sizes are frequently used to estimate RIs in veterinary clinical pathology; it is a very problematic issue. Different methods have been proposed to deal with small sample groups. The IFCC-CLSI guideline recommends Horn's robust method involving iterative processes for identifying the location of the median and spread of the distribution.[68] In the examples of serum calcium and alanine aminotransferase in men and women, estimates of RIs by the robust method in sets of 80 individuals were close to the reference limits and CIs that were obtained nonparametrically with the full reference sample group of 120.[5] Although some publications demonstrate robust methods on smaller sample sizes,[65] the IFCC-CLSI working group "is hesitant to recommend" the robust method with sample sizes of fewer than 80 individuals.[5] In a study of canine plasma creatinine using multiple small subsets ($n = 27$) randomly selected from 1439 reference samples, it was shown that the robust method could only be applied appropriately after transformation of the data to fit a Gaussian distribution. Depending on the subset selected, the reference limits may be quite different from those estimated from the entire reference sample group.[69] When reference limits are estimated from small sample sizes, imprecision of the limits may be very high. In addition, when nonparametric methods are used, CIs of the limits are not easily estimated.[70] Other methods for estimating RIs in small samples sizes based on variance component analysis have been used in human clinical pathology.[71]

Multicenter reference intervals

The creation of multicenter RIs from the contributions of multiple laboratories has been successfully established and used clinically in human laboratory medicine.[72–75] The development of common or shared RIs was propelled by the necessity to share workload, augment sample size, and increase the number of analytes available for diagnostic use. Determination of common RIs is possible only when there is sufficient comparability of all preanalytical and analytical conditions and when the reference populations of the different laboratories are similar.[76] Common RIs should be validated or verified in each laboratory. However, a recent study in human clinical pathology revealed that adoption of common RIs should be performed with caution.[77,78] Common RIs have yet to be used in veterinary clinical pathology to our knowledge. However, it may be a practical option in the future, especially for exotic species or groups of animals for which only small sample sizes can be obtained. Common RIs require large databases with particular attention to analytical procedures and method accuracy.

Transference of a reference interval

Transference has been used for decades in many laboratories when a new instrument or technique is introduced, but is now accepted by IFCC-CLSI for broader application.[5,79]

The following 3 conditions should be fulfilled in order for transference to be acceptable:

1. The RI to be transferred must have been obtained properly and its generation and other validation procedures must be fully documented and available for review. In veterinary clinical pathology laboratories, some RI lack complete documentation of the reference population parameters or analytical specifications, a situation that should be rectified in the future.
2. The analytical systems must be comparable. A classical procedure for the comparison of methods (see a review in veterinary clinical pathology[80] or CLSI EP9-A2[81]) is used to determine whether correlation between the analytical systems is sufficiently high to use regression statistics to calculate a new RI from the preceding one. Even when correlation is excellent ($r^2 > .9$), there may be a significant difference between results of the existing and new systems due to bias, which may result in differences between the old and new RIs. For regression methods to be used properly, test values should have a large enough range ratio and the intercept should be small relative to the RI; even then, regression methods may not be suitable.[5,80]
3. The patient populations must be comparable. This implies that complete demographic information on the original reference sample group is available and corresponds to the demographics of the new population. This is not an issue when a method is changed within the same laboratory but may be highly significant when RIs are transferred to different regions or different countries.

Validation of a reference interval

Validating a pre-existing or a transferred RI avoids the enormous amount of work and expense necessitated by a priori determination of an RI. RI validation has been proposed for more than 15 years[82] and, according to C28-A3, is acceptable by adhering to one of the following 3 procedures.[5]

Subjective assessment

Acceptability is based on an expert opinion after careful examination of all conditions by which the RI was initially determined. These conditions must be matched by those in the receiving laboratory. Because this procedure is subjective, it comprises too many risks to be recommended in veterinary clinical pathology.

Validation using small numbers of reference individuals

Acceptability is based on "examining a small number of reference individuals ($n = 20$) from the receiving laboratory's own population and comparing these reference values to the larger, more comprehensive original study." The probability of false rejection of an RI by this method is $< 1\%$ when 1 or more sets of 20 reference individuals is used (binomial test).[5] However, this method cannot accurately identify RIs that are too wide for the new population. A schematic representation of the procedure is demonstrated in Figure 3.

Validation using large numbers of reference individuals

This procedure is roughly analogous to the a priori determination of an RI, except that the number of reference individuals is < 120. In this case, as stated in the IFCC-CLSI guidelines, "the availability of robust statistical techniques provides another alternative."[5]

Conclusions

The general recommendations for the determination of RIs in medical laboratories are applicable to veterinary clinical pathology. The first step in advancing the science of RI determination in veterinary clinical pathology is to speak the same language, ie, to use the correct terms according to internationally accepted definitions. The second step is to understand the importance of and implement the recommendations for reference subject selection and quality method performance. Collection of as many reference samples as possible from well-defined reference subjects is invaluable in the determination of accurate RIs. This will do more to optimize RIs than the selection of statistical

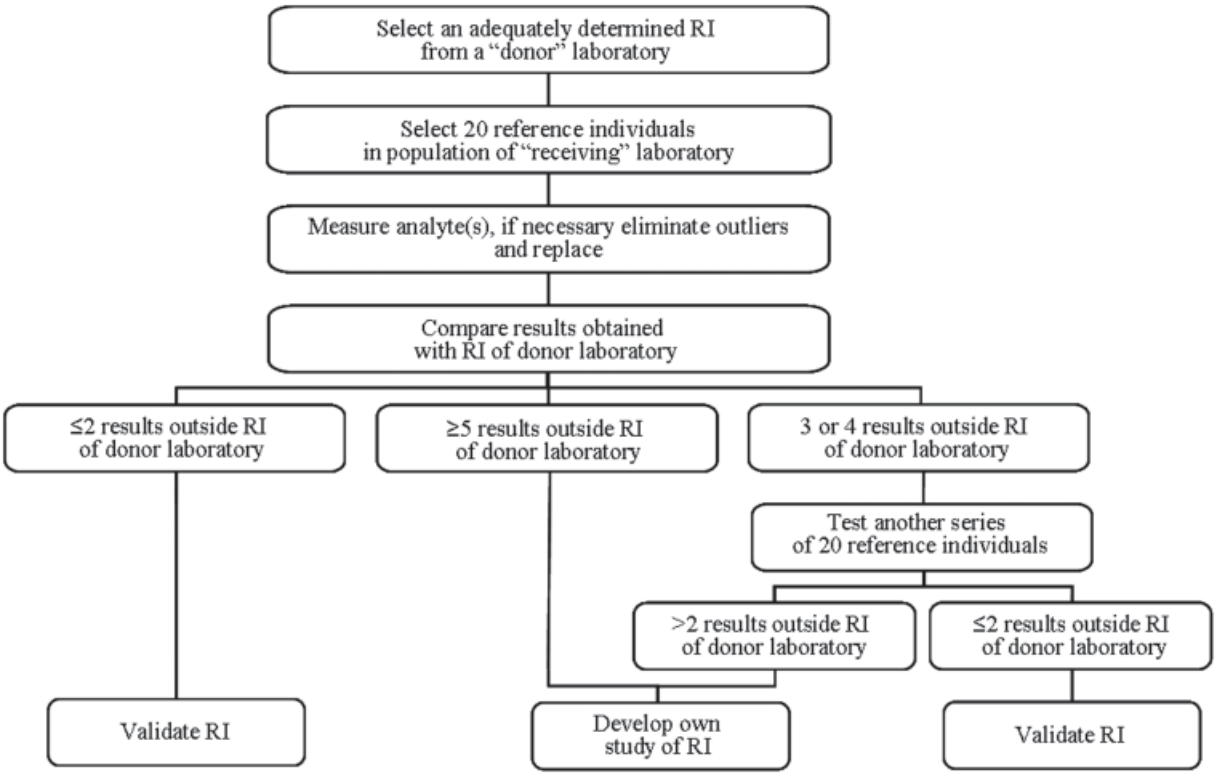

Figure 3. Algorithm of actions to validate a pre-existing reference interval according to the Clinical Laboratory and Standards Institute (CLSI) and International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine document C28-A3.[5]

methods, even though correct selection of the latter may also improve the accuracy of RIs, especially when collection of large numbers of specimens is not possible. Currently, most RIs published in veterinary clinical pathology do not meet the criteria discussed in this review. The challenge for the future is to make reasonable and applicable recommendations, especially for small samples, based on C28-A3, which can be used as a guideline in veterinary medicine.

References

1. Grasbeck R, Saris NE. Establishment and use of normal values. *Scand J Clin Lab Invest*. 1969;26(Suppl 110): 62–63.
2. Sunderman FW Jr. Current concepts of "normal values," "reference values," and "discrimination values" in clinical chemistry. *Clin Chem*. 1975;21:1873–1877.
3. Grasbeck R. Reference value philosophy. *Bull Mol Biol Med*. 1983;8:1–11.
4. Henny J, Hyltoft Petersen P. Reference values: from philosophy to a tool for laboratory medicine. *Clin Chem Lab Med*. 2004;42:686–691.
5. Clinical and Laboratory Standards Institute. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline*. 3rd ed. Wayne, PA: CLSI; 2008.
6. Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem*. 1987;25:337–342.
7. PetitClerc C, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem*. 1987;25:639–644.
8. Solberg HE, Petitclerc C. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem*. 1988;26:593–598.
9. Solberg HE, Stamm D. Approved recommendation on the theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. *Eur J Clin Chem Clin Biochem*. 1991;29:531–535.
10. Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem*. 1987;25:645–656.
11. Dybkaer R, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. *J Clin Chem Clin Biochem*. 1987;25:657–662.
12. Ceriotti F, Boyd JC, Klein G, et al. Reference intervals for serum creatinine concentrations: assessment of available data for global application. *Clin Chem*. 2008;54:559–566.
13. Henny J, Petitclerc C, Fuentes-Arderiu X, et al. Need for revisiting the concept of reference values. *Clin Chem Lab Med*. 2000;38:589–595.
14. Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta*. 2003;334:5–23.
15. Horn PS, Pesce AJ. *Reference Intervals. A User's Guide*. Washington: AACC Press; 2005.
16. Harris EK, Boyd JC. *Statistical Basis of Reference Values in Laboratory Medicine*. New York: Marcel Dekker; 1995.
17. Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER, eds. *Tietz Textbook of Clinical Chemistry*. 3rd ed. Philadelphia, PA: Saunders; 1999:336–356.
18. Farver TB. Concepts of normality in clinical biochemistry. In: Kaneko JJ, Harvey JW, Bruss ML, eds. *Clinical Biochemistry of Domestic Animals*. 6th ed. Amsterdam: Elsevier; 2008:1–25.
19. Solberg HE. RefVal: a program implementing the recommendations of the International Federation of Clinical Chemistry on the statistical treatment of reference values. *Comput Methods Programs Biomed*. 1995;48:247–256.
20. Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal program. *Clin Chem Lab Med*. 2004;42:710–714.
21. Kenny D, Olesen H. International Union of Pure and Applied Chemistry (IUPAC). International Federation of Clinical Chemistry (IFCC). Committee on nomenclature, properties and units: properties and units in the clinical laboratory sciences. II. Kinds-of-property. *Eur J Clin Chem Clin Biochem*. 1997;35:317–344.
22. Dybkaer R. Vocabulary for use in measurement procedures and description of reference materials in laboratory medicine. *Eur J Clin Chem Clin Biochem*. 1997;35:141–173.
23. Harris EK, Yasaka T, Horton MR, Shakarji G. Comparing multivariate and univariate subject-specific reference regions for blood constituents in healthy persons. *Clin Chem*. 1982;28:422–426.
24. Boyd JC. Reference regions of two or more dimensions. *Clin Chem Lab Med*. 2004;42:739–746.
25. International Organisation for Standardization. 15189 I. Medical laboratories—Particular requirements for quality and competence. 2003.

○

26. Queralto JM. Intraindividual reference values. *Clin Chem Lab Med*. 2004;42:765–777.
27. Harris EK, Yasaka T. On the calculation of a "reference change" for comparing two consecutive measurements. *Clin Chem*. 1983;29:25–30.
28. Harris EK. Some theory of reference values. II. Comparison of some statistical models of intraindividual variation in blood constituents. *Clin Chem*. 1976;22:1343–1350.
29. Jensen AL, Aaes H. Critical differences of clinical chemical parameters in blood from dogs. *Res Vet Sci*. 1993;54:10–14.
30. Jensen AL, Houe H, Nielsen CG. Critical differences of clinical chemical components in blood from Red Danish dairy cows based on weekly measurements. *J Comp Pathol*. 1992;107:373–378.
31. Nabity MB, Boggess MM, Kashtan CE, Lees GE. Day-to-day variation of the urine protein: creatinine ratio in female dogs with stable glomerular proteinuria caused by X-linked hereditary nephropathy. *J Vet Intern Med*. 2007;21:425–430.
32. Vickers AJ, Lilja H. Cutpoints in Clinical Chemistry: time for fundamental reassessment. *Clin Chem*. 2009;55:15–17.
33. Petitclerc C. Normality: the unreachable star? *Clin Chem Lab Med*. 2004;42:698–701.
34. Lees GE, Brown SA, Elliott J, et al. Assessment and management of proteinuria in dogs and cats: 2004 ACVIM Forum Consensus Statement (small animal). *J Vet Intern Med*. 2005;19:377–385.
35. Altman DG, Bland JM. Statistics notes: variables and parameters. *BMJ*. 1999;318:1667.
36. Anonymous. *Constitution of the World Health Organization*. New York: United Nations; 1946:19.
37. Walton RM. Establishing reference intervals: health as a relative concept. *Semin Avian Exotic Pet Ped*. 2001;10:67–71.
38. Reed AH, Henry RJ, Mason WB. Influence of Statistical Method Used on the Resulting Estimate of Normal Range. *Clin Chem*. 1971;17:275–284.
39. Jennen-Steinmetz C, Wellek S. A new approach to sample size calculation for reference interval studies. *Stat Med*. 2005;24:3199–3212.
40. Poulsen O, Holst E, Christensen JM. Calculation and application of coverage intervals for biological reference values (IUPAC technical report). *Pure Appl Chem*. 1997;69:1601–1611.
41. Linnet K. Two-stage transformation systems for normalization of reference distributions evaluated. *Clin Chem*. 1987;33:381–386.
42. Klein G, Junge W. Creation of the necessary analytical quality for generating and using reference intervals. *Clin Chem Lab Med*. 2004;42:851–857.
43. Thienpont LM, Van Uytfanghe K, Rodriguez Cabaleiro D. Metrological traceability of calibration in the estimation and use of common medical decision-making criteria. *Clin Chem Lab Med*. 2004;42:842–850.
44. Ricos C, Domenech MV, Perich C. Analytical quality specifications for common reference intervals. *Clin Chem Lab Med*. 2004;42:858–862.
45. Westgard JO. Design of internal quality control for reference value studies. *Clin Chem Lab Med*. 2004;42:863–867.
46. Plebani M. What information on quality specifications should be communicated to clinicians, and how? *Clin Chim Acta*. 2004;346:25–35.
47. Jorgensen LG, Brandslund I, Hyltoft Petersen P. Should we maintain the 95 percent reference intervals in the era of wellness testing? A concept paper. *Clin Chem Lab Med*. 2004;42:747–751.
48. Hyltoft Petersen P, Blaabjerg O, Andersen M, Jorgensen LG, Schousboe K, Jensen E. Graphical interpretation of confidence curves in rankit plots. *Clin Chem Lab Med*. 2004;42:715–724.
49. Lumsden JH, Mullen K. On establishing reference values. *Can J Comp Med*. 1978;42:293–301.
50. Solberg HE. Statistical treatment of reference values in laboratory medicine: testing the goodness-of-fit of an observed distribution to the Gaussian distribution. *Scand J Clin Lab Invest Suppl*. 1986;184:125–132.
51. Koduah M, Iles TC, Nix BJ. Centile charts I: new method of assessment for univariate reference limits. *Clin Chem*. 2004;50:901–906.
52. Horn PS, Pesce AJ, Copeland BE. Reference interval computation using robust vs parametric vs nonparametric analyses. *Clin Chem*. 1999;45: 2284–2285.
53. Linnet K. Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clin Chem*. 2000;46:867–869.
54. Solberg HE, Lahti A. Detection of outliers in reference distributions: performance of Horn's algorithm. *Clin Chem*. 2005;51:2326–2332.
55. Horn PS, Feng L, Li Y, Pesce AJ. Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem*. 2001;47:2137–2145.
56. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem*. 1990;36:265–270.
57. Lahti A, Petersen PH, Boyd JC, Rustad P, Laake P, Solberg HE. Partitioning of nongaussian-distributed

biochemical reference data into subgroups. *Clin Chem*. 2004;50:891–900.

58. Lahti A, Hyltoft Petersen P, Boyd JC. Impact of subgroup prevalences on partitioning of Gaussian-distributed reference values. *Clin Chem*. 2002; 48:1987–1999.
59. Lahti A, Hyltoft Petersen P, Boyd JC, Fraser CG, Jorgensen N. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. *Clin Chem*. 2002;48:338–352.
60. Lahti A. Partitioning biochemical reference data into subgroups: comparison of existing methods. *Clin Chem Lab Med*. 2004;42:725–733.
61. Fuentes-Arderiu X, Ferre-Masferrer M, Alvarez-Funes V. Harris & Boyd's test for partitioning the reference values. *Eur J Clin Chem Clin Biochem*. 1997;35:733.
62. Virtanen A, Kairisto V, Uusipaikka E. Regression-based reference limits: determination of sufficient sample size. *Clin Chem*. 1998;44:2353–2358.
63. Virtanen A, Kairisto V, Uusipaikka E. Parametric methods for estimating covariate-dependent reference limits. *Clin Chem Lab Med*. 2004;42:734–738.
64. Solberg HE. Using a hospitalized population to establish reference intervals: pros and cons. *Clin Chem*. 1994;40:2205–2206.
65. Ferre-Masferrer M, Fuentes-Arderiu X, Puchal-Ane R. Indirect reference limits estimated from patients' results by three mathematical procedures. *Clin Chim Acta*. 1999;279:97–105.
66. Dimauro C, Bonelli P, Nicolussi P, Rassu SP, Cappio-Borlino A, Pulina G. Estimating clinical chemistry reference values based on an existing data set of unselected animals. *Vet J*. 2007;178:278–281.
67. 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: 43–48.
68. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44:622–631.
69. Geffré A, Braun JP, Trumel C, Concordet D. Estimation of reference intrvals from small samples: an example with canine plasma creatinine. *Vet Clin Pathol*. E-pub ahead of print: DOI: 10.1111/j.1939–165X.2009.00155.x.
70. van der Meulen EA, Boogaard PJ, van Sittert NJ. Use of small-sample-based reference limits on a group basis. *Clin Chem*. 1994;40:1698–1702.
71. Hansen AM, Garde AH, Eller NH. Estimation of individual reference intervals in small sample sizes. *Int J Hyg Environ Health*. 2007;210:471–478.
72. Rustad P, Felding P, Lahti A. Proposal for guidelines to establish common biological reference intervals in large geographical areas for biochemical quantities measured frequently in serum and plasma. *Clin Chem Lab Med*. 2004;42:783–791.
73. Fuentes-Arderiu X, Mas-Serra R, Aluma-Trullas A, Marti-Marcet MI, Dot-Bach D. Guideline for the production of multicentre physiological reference values using the same measurement system. A proposal of the Catalan Association for Clinical Laboratory Sciences. *Clin Chem Lab Med*. 2004;42:778–782.
74. Jones GR, Barker A, Tate J, Lim CF, Robertson K. The Case for Common Reference Intervals. *Clin Biochem Rev*. 2004;25:99–104.
75. Stromme JH, Rustad P, Steensland H, Theodorsen L, Urdal P. Reference intervals for eight enzymes in blood of adult females and males measured in accordance with the International Federation of Clinical Chemistry reference system at 37 degrees C: part of the Nordic Reference Interval Project. *Scand J Clin Lab Invest*. 2004;64:371–384.
76. Ceriotti F. Prerequisites for use of common reference intervals. *Clin Biochem Rev*. 2007;28:115–121.
77. Boyd JC. Cautions in the adoption of common reference intervals. *Clin Chem*. 2008;54:238–239.
78. Ichihara K, Itoh Y, Lam CW, et al. Sources of variation of commonly measured serum analytes in 6 Asian cities and consideration of common reference intervals. *Clin Chem*. 2008;54:356–365.
79. Petitclerc C, Morasse P, Carrier G. Transferability of reference values. Practical experience with alkaline phosphatase. *Bull Mol Biol Med*. 1983;8: 41–52.
80. Jensen AL, Kjelgaard-Hansen M. Method comparison in the clinical laboratory. *Vet Clin Pathol*. 2006;35: 276–286.
81. National Committee for Clinical Laboratory Standards. *Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline. Document EP9-A2*. Wayne, PA: NCCLS; 2002.
82. Holmes EW, Kahn SE, Molnar PA, Bermes EW Jr Verification of reference ranges by using a Monte Carlo sampling technique. *Clin Chem*. 1994;40: 2216–2222.