THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Pathologie, Toxicologie, Génétique & Nutrition*

Présentée et soutenue par *Anne GEFFRÉ*
Le *1er juillet 2011*

Titre : *Nouvelles approches de la production d'intervalles de référence de populations*

JURY
Bureau Fabrice, Professeur d'Université, Rapporteur
Queralto Josep, Professeur d'Université, Rapporteur
Cadoré Jean-Luc, Professeur VetAgro Sup, Examinateur
Concordet Didier, Professeur INPT-ENVT, Examinateur
Courtade-Saïdi Monique, Professeur UPS, Examinateur
Henny Joseph, Expert, Examinateur
Braun Jean-Pierre, Professeur INPT-ENVT, Codirecteur de thèse
Trumel Catherine, Professeur INPT-ENVT, Codirecteur de thèse

Ecole doctorale : *Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries*
Unité de recherche : *UMR1331 Toxalim, INRA-INPT*
Directeur(s) de Thèse : *Prs JP. Braun & C. Trumel*

REMERCIEMENTS

Je tiens à remercier chaleureusement les Professeurs Jean-Pierre Braun et Catherine Trumel qui ont codirigé ce travail de thèse. Je remercie également les membres du jury : le Professeur Fabrice Bureau de l'Université de Liège, le Professeur Jean-Luc Cadoré de VetAgro Sup, le Professeur Didier Concordet de l'Ecole Nationale Vétérinaire de Toulouse, le Professeur Monique Courtade-Saïdi de l'Université Paul Sabatier, le Docteur Joseph Henny du Centre de Médecine Préventive de Vandoeuvre-lès-Nancy, le Professeur Josep Queralto de l'Université de Barcelona.

TABLE DES MATIERES

La biologie médicale est une discipline médicale qui englobe toutes les sciences de laboratoire utilisées dans le domaine de la santé à l'exception de l'anatomo-pathologie, c'est-à-dire principalement l'hématologie, la biochimie clinique, la parasitologie, la bactériologie, la virologie, l'immunologie et la biologie moléculaire. En biologie humaine, son exercice est étroitement encadré par la loi [1] et pratiqué exclusivement par des spécialistes. En biologie animale, c'est une spécialité vétérinaire [2], [3] mais à l'heure actuelle, elle peut être pratiquée sans qualification réglementaire.

Les prélèvements biologiques servent de support à des analyses variées « qui concourent au diagnostic, au traitement ou à la prévention des maladies (…) ou qui font apparaître toute autre modification de l'état physiologique, ». Les cliniciens recourent quotidiennement aux analyses biologiques puis doivent interpréter leurs résultats, en les replaçant dans le contexte ayant justifié leur prescription. La question quasi instinctive que l'on se pose lors de cette étape est « le résultat de l'analyse est-il normal (ou bien trop élevé ou trop bas) ? » ou, autrement dit, « où se situe le résultat de l'analyse par rapport aux résultats obtenus pour des sujets sains? ». C'est pour essayer de répondre à cette question qu'a été développée la théorie des intervalles de référence. Elle vise à décrire de manière aussi précise que possible les variations des marqueurs biologiques utilisés en médecine chez des sujets supposés en bonne santé.

La notion de valeur de référence a été élaborée en médecine humaine à partir de la fin des années mille neuf cent soixante[4]. Cette question a été intensément travaillée par les experts de la Société Française de Biologie Clinique dont les conclusions [5], [6], [7], [8], [9], [10],[11], [12] ont été reprises à l'étranger par d'autres sociétés et à l'échelle internationale par l'International Federation of Clinical Chemistry (IFCC). Elles ont ensuite fait l'objet de mises à jour au fil des questions soulevées, pour aboutir à la dernière version des recommandations internationales par l'IFCC et le Clinical Laboratory and Standards Institute (CLSI) [13]. Les principaux facteurs de variation qui ont été les premiers étudiés semblent évidents, comme l'âge ou le sexe des individus, mais des facteurs ethniques ou culturels (habitudes alimentaires, par exemple) peuvent également compliquer la tâche.

Qu'en est-il alors en médecine vétérinaire où il faut considérer la multitude des espèces étudiées, des races, voire des souches pour les espèces de laboratoire ou des conditions d'élevage ? Ceci donne un premier aperçu des difficultés rencontrées, auxquelles vient s'ajouter l'absence de grands centres vétérinaires de santé qui

permettraient, en fournissant des milliers de valeurs de référence, de documenter chacun de ces facteurs de variation.

Le présent travail a visé à apporter des réponses à ces questions, et pour ce faire a consisté à :

- faire le point sur la littérature scientifique humaine et animale concernant la détermination des intervalles de référence,

- tester l'application des dernières recommandations internationales à des exemples concrets, dans des conditions optimales,

- tester de nouvelles méthodes en suivant principalement deux pistes : d'abord en travaillant sur des échantillons de référence de petite taille (soit en déterminant un intervalle de référence *de novo*, soit en transférant un intervalle de référence préexistant), ensuite en travaillant sur des échantillons de référence « pollués » par des valeurs issues d'individus non sains,

- mettre au point un outil statistique permettant de répondre rapidement et simplement aux questions précédemment posées.

Chacun de ces points fait l'objet d'une partie de ce mémoire qui est conclu par une discussion générale synthétique.

Partie 1
Faire le point sur la littérature scientifique humaine et animale

La notion de valeurs de référence, introduite dans les années mille neuf cent soixante [4], est aujourd'hui mondialement recommandée pour décrire les variations interindividuelles observées chez des sujets sains et pour remplacer des termes ambigus tels que « valeurs normales » ou « normes » qui ont d'autres significations [14]. Cependant, il faut reconnaître que ces derniers ont persisté dans l'usage courant pour le grand public, et parfois même chez les spécialistes.

La collecte et l'utilisation de valeurs de référence impose de définir préalablement l'état de « bonne santé » des sujets, ce qui est difficile en biologie humaine [15] et encore plus compliqué en biologie animale. Faute de pouvoir disposer de critères indiscutables, il est nécessaire de sélectionner les individus de référence selon des critères d'inclusion et d'exclusion bien définis et clairement décrits (cf. Figure 1, page 290 de l'article suivant). L'ensemble de ces individus constitue la population de référence, dont les limites sont inconnues mais à partir de laquelle on peut extraire de manière aléatoire un échantillon de référence sur lequel sont pratiquées les analyses. Les valeurs de référence sont les résultats des mesures pratiquées sur cet échantillon d'individus. Leur distribution peut ensuite être facilement visualisée sous forme d'un histogramme de distribution. Arbitrairement, la majeure partie des utilisateurs a décidé de représenter l'intervalle de référence comme celui comprenant 95% des valeurs observables dans la population de référence, et donc d'utiliser les quantiles 0,025 et 0,975 comme limites de l'intervalle de référence. Une conséquence délicate à gérer en pratique clinique quotidienne est que 5% des sujets en bonne santé ont des valeurs en dehors de l'intervalle de référence. Pour définir des intervalles de référence, il apparaît donc que l'étape la plus difficile, la plus longue et la plus coûteuse est la sélection des

individus de référence. Les étapes ultérieures (prélèvements, mesures et calculs) sont beaucoup plus simples, même si souvent l'abondance des publications a pu laisser penser que les calculs statistiques étaient la difficulté majeure dans la détermination des intervalles de référence. La base des recommandations internationales depuis de nombreuses années a été d'utiliser un minimum de 120 sujets de référence pour chaque groupe de partition (lié à l'âge, au sexe, aux conditions de vie, etc.) puis d'appliquer des méthodes non paramétriques de détermination des limites de référence et de leurs intervalles de confiance. Plus récemment, des approches alternatives moins lourdes ont été proposées et font l'objet de la troisième partie de ce mémoire :

- le transfert et/ou la validation dans un laboratoire utilisateur d'intervalles de référence correctement déterminés préalablement,

- le calcul d'intervalles de référence à partir de populations de plus petite taille ou bien des déterminations *a posteriori* avec des sélections d'individus de référence éventuellement polluées par des sujets malades ; ces approches se fondent sur des méthodes statistiques plus complexes que les simples déterminations non paramétriques.

Article n°1

Geffré A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP.

Reference values: a review.

Veterinary Clinical Pathology 2009;38:288-98.

Veterinary Clinical Pathology ISSN 0275-6382

INVITED REVIEW

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

Correspondence
A. Geffré, Département des Sciences Cliniques, Ecole Nationale Vétérinaire de Toulouse, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France
E-mail: a.geffre@envt.fr

DOI:10.1111/j.1939-165X.2009.00179.x

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.

This review article was peer-reviewed. Dr. Braun, an editor of the journal, was not involved in the peer review process or the decision to publish this article.

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 intervals* (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.

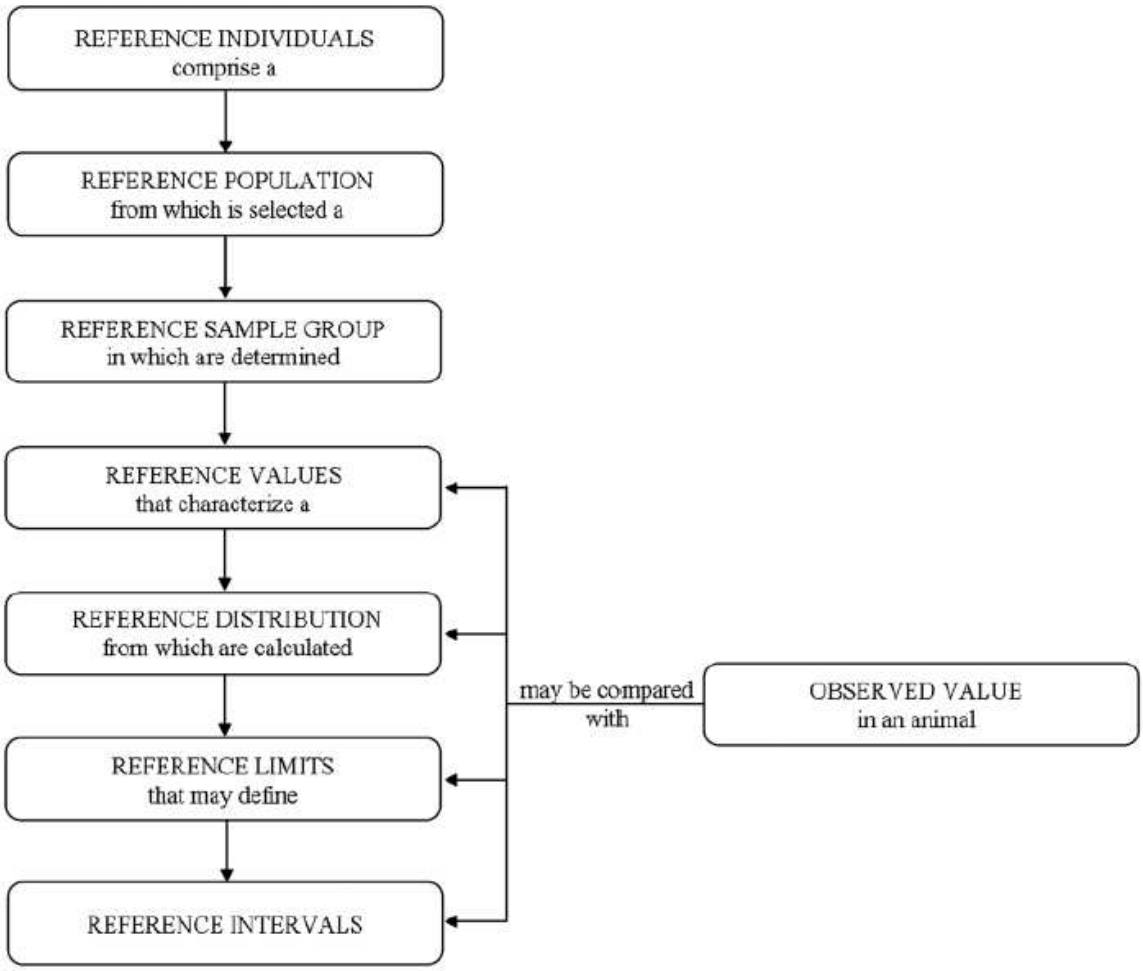


Figure 1. Relationships between the terms related to reference values according to the Clinical Laboratory and Standards Institute (CLSI) and International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine document C28-A3.[5]

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]

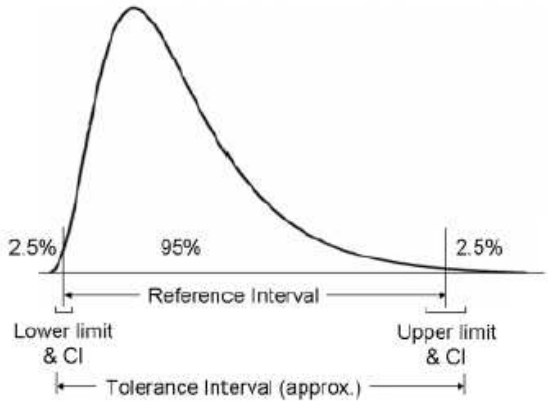


Figure 2. Schematic representation of a reference interval, reference limits, confidence intervals (CI) of the limits, and tolerance interval.

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

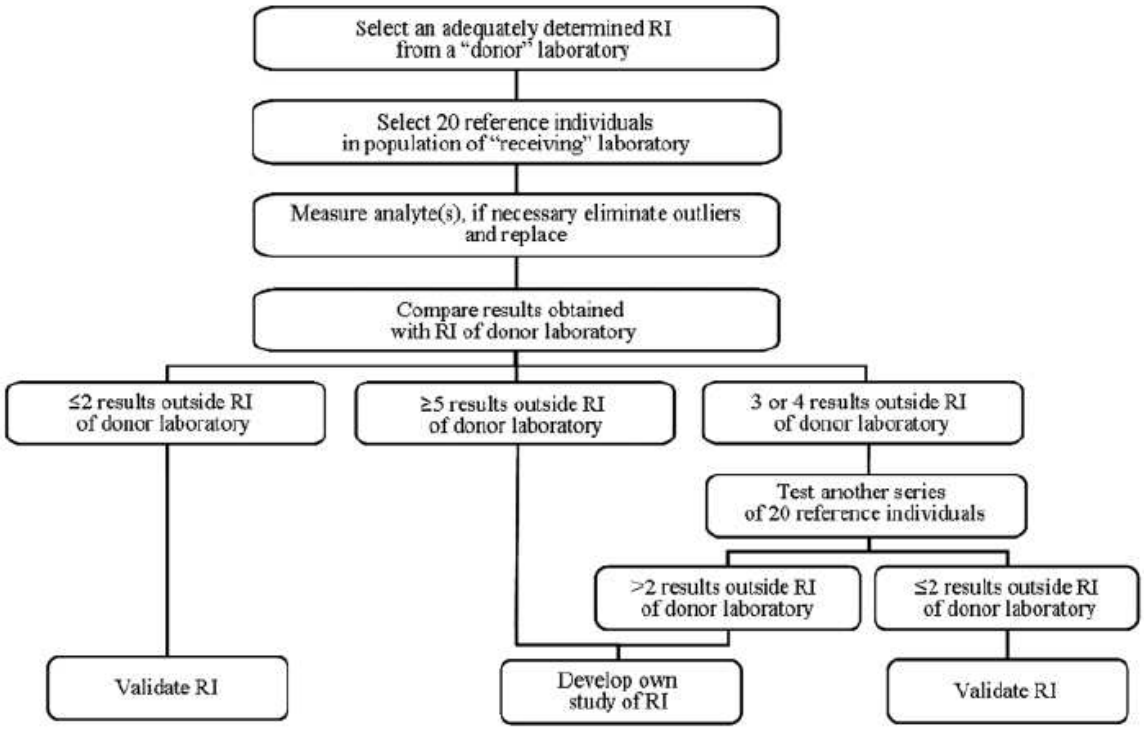


Figure 3. Algorithm of actions to validate a pre-existing reference interval according to the Clinical Laboratory and Standards Institute (CLSI) and International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine document C28-A3.[5]

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.

Partie 2
Tester l'application des recommandations internationales dans des conditions optimales

La détermination d'un intervalle de référence, telle que recommandée par les groupes d'experts comme l'International Federation of Clinical Chemistry (IFCC) [16], [17], [18], [19], [20], [21] et le Clinical Laboratory and Standards Institute (CLSI) [13], reste parfois difficile à mettre en œuvre. La dernière mise à jour des recommandations du CLSI confirme la préconisation d'utiliser au moins 120 individus de référence. Ainsi, une méthode non paramétrique peut être appliquée pour la détermination des limites de référence et de leurs intervalles de confiance à 90%. Cette procédure, censée être appliquée pour chaque analyse biologique proposée dans chaque laboratoire de biologie médicale, devient vite extrêmement lourde, tant sur le plan matériel que financier. Si la détermination *de novo* des intervalles de référence est un défi à relever pour les grands laboratoires de biologie médicale humaine, elle reste illusoire pour la plupart des laboratoires vétérinaires.

Nous avons cependant voulu l'appliquer pour en analyser les difficultés et les limites ainsi que pour en utiliser les données dans d'autres études, notamment de validation et d'emploi de petits échantillons (Article n°4). Deux applications ont été entreprises, conformément aux recommandations internationales, pour la biologie médicale du chien, l'une dans le domaine de l'hémostase (Article n°2), l'autre dans celui de l'hématologie (Article n°3). Elles reposent sur :

- une sélection difficile et parfois discutable des individus de référence, qui doivent refléter la population des sujets rencontrés ultérieurement par l'utilisateur,

- une description détaillée des conditions métrologiques et de leur contrôle, essentielle pour les éventuels transferts ultérieurs,

- l'application de la méthode non paramétrique de détermination des limites et des intervalles de confiance de ces limites,

- une description détaillée des résultats, et notamment la représentation graphique de la distribution des valeurs. Ce dernier point n'est pas explicitement prévu par les recommandations (excepté comme outil de travail). Il est cependant très utile pour les utilisateurs et permet une meilleure visualisation de la dispersion des valeurs que les simples valeurs chiffrées des limites de référence (même lorsque ces dernières sont accompagnées de leurs intervalles de confiance, ce qui est rarement le cas dans la littérature).

Ces intervalles ayant été déterminés conformément aux recommandations, ils doivent pouvoir servir de base à des laboratoires extérieurs pour transfert ou validation sans avoir besoin de sélectionner à nouveau un échantillon de référence (cf. partie 3).

Article n°2

Geffré A, Grollier S, Hanot C, Vergez F, Trumel C, Braun JP.

Canine reference intervals for coagulation markers using the STA Satellite and the STA-R Evolution analyzers.

Journal of Veterinary Diagnostic Investigation 2010;22:690-5.

Article n°3

Bourgès-Abella N, Geffré A, Concordet D, Braun JP, Trumel C.

Canine hematology reference intervals for the XT-2000iV analyser.

Veterinary Clinical Pathology, 2011;40:in press.

J Vet Diagn Invest 22:690–695 (2010)

Canine reference intervals for coagulation markers using the STA Satellite® and the STA-R Evolution® analyzers

Anne Geffré,[1] Sophie Grollier, Camille Hanot, François Vergez, Catherine Trumel, Jean-Pierre Braun

Abstract. The aim of the current study was to determine canine reference intervals for prothrombin time (PT), activated partial thromboplastin time (APTT), fibrinogen, and antithrombin (AT) according to international recommendations. The STA Satellite® coefficients of variation of within-laboratory imprecision were 3.9%, 1.3%, 6.9%, and 5.1% for PT, APTT, fibrinogen, and AT, respectively. At 4°C, citrated specimens were stable up to 8 hr for whole blood and 36 hr for plasma, except for APTT, which increased slightly (<1 sec). Nonparametric reference intervals determined in citrated plasma from 139 healthy fasting purebred dogs were 6.9–8.8 sec, 13.1–17.2 sec, 1.24–4.30 g/l, and 104–188% for PT, APTT, fibrinogen, and AT, respectively. Based on Passing–Bablok comparison between STA Satellite and STA-R Evolution® using 60 frozen specimens from a canine plasma bank, the corresponding reference intervals were transferred to the STA-R Evolution: 7.1–9.2 sec, 12.9–17.3 sec, 1.20–4.43 g/l, and 94–159% for PT, APTT, fibrinogen, and AT, respectively.

Key words: Activated partial thromboplastin time; antithrombin; dogs; fibrinogen; prothrombin time; reference intervals.

Introduction

There are few studies in the veterinary literature on reference intervals for coagulation markers of dogs, and most of the studies do not meet international recommendations such as defined in the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and the Clinical and Laboratory Standards Institute (CLSI) guidelines.[2] Introduction of a new analyzer into the laboratory was the opportunity to make a proper a priori determination of coagulation test reference values in dogs, which may be further used for transference in other sites after proper verification using the new recommendations.[2] The current study was based on the 13-step procedure recommended by IFCC-CLSI and involved a reference sample representative of the canine population in France, according to the breed distribution reported by the Société Centrale Canine (http://www.scc.asso.fr/mediatheque/statistiques, accessed January 2008). The 4 analytes studied were chosen because they concerned the coagulation tests most frequently requested at the Central Laboratory of the clinics of the Ecole Nationale Vétérinaire de Toulouse (Toulouse, France).

The present study therefore consisted of determining the reference interval for prothrombin time (PT), activated partial thromboplastin time (APTT), fibrinogen, and antithrombin (AT) in purebred healthy dogs with the STA Satellite® analyzer[a] according to international recommendations. The reference intervals thus determined were transferred to another laboratory in the same city, using the same reagents but a different analyzer, namely, the STA-R Evolution® analyzer.[a]

Material and methods

The current study was performed between October 2007 and March 2008. Before each run of analyses, the STA Satellite analyzer was checked according to the manufacturer's recommendations.[a] Analyses were performed by coagulometric method for PT, APTT, and fibrinogen, and by a colorimetric test for AT. Reagents and characteristics of the techniques are summarized in Table 1. Where necessary, a veterinary configuration of the analyzer's software performed automatic dilutions, thereby increasing the standard human analytical ranges for the lowest and highest values.

As canine control solutions are not available for coagulation testing, the accuracy and imprecision of the STA Satellite analyzer were assessed with 2 human control solutions (STA Coag Control N + P)[a] according to the

From the Département des Sciences Cliniques (Geffré, Grollier, Hanot, Trumel, Braun), and UMR 181 Physiopathologie et Toxicologie Expérimentales, ENVT, INRA (Braun), Ecole Nationale Vétérinaire de Toulouse, Toulouse, France; and Groupe hospitalier de Rangueil-Larrey, Laboratoire d'Hématologie, Toulouse, France (Vergez).

[1]Corresponding Author: Anne Geffré, Département des Sciences Cliniques, Ecole Nationale Vétérinaire de Toulouse, 23 Chemin des Capelles, 31076 Toulouse cedex 3, France. a.geffre@envt.fr

Table 1. Reagents and characteristics of methods based on manufacturer's data and laboratory testing according to the Clinical and Laboratory Standards Institute.*†

			Manufacturer data			Laboratory data			
							Within-laboratory imprecision CV (n = 20)		
Analyte	Method	Reagent‡	Analytical range§	Repeatability CV (n = 21)	Within-laboratory imprecision CV (n = 10)	Trueness (n = 20)	Human control solutions	Repeatability CV (n = 139) Canine specimens
PT	Coagulation	STA Néoplastine CI Plus	3.0–120.0 sec	0.7% at 13.6 sec 0.5% at 22.7 sec	1.5% at 15.1 sec 1.6% at 29.4 sec	−23.0% at 13.8 sec −10.2% at 25.4 sec	1.9% at 13.8 sec 3.9% at 25.4 sec	0.6% at mean = 7.7 sec
APTT	Coagulation	STA Cephascreen	10.0–180.0 sec	0.6% at 29.8 sec 0.8% at 47.2 sec	1.4% at 29.8 sec 0.9% at 48.0 sec	+0.6% at 28.8 sec −2.1% at 52.1 sec	1.3% at 28.8 sec 1.0% at 52.1 sec	1.0% at mean = 15.0 sec
Fibrinogen	Coagulation	STA Fib2	1.5–9.0 g/l 0.6–12 g/l§	2.9% at 3.07 g/l 2.5% at 1.24 g/l	2.0% at 2.72 g/l 3.7% at 1.36 g/l	+8.7% at 3.0 g/l +10.8% at 1.0 g/l	5.7% at 3.0 g/l 6.9% at 1.0 g/l	2.3% at mean = 2.3 g/l
AT	Colorimetry	STA Stachrom AT	9–140%	2.6% at 105% 5.0% at 46%	4.8% at 105% 5.0% at 46%	+2.9% at 106% −3.2% at 42%	3.7% at 106% 5.1% at 42%	1.3% at mean = 143%

* CV = coefficient of variation; PT = prothrombin time; APTT = activated partial thromboplastin time; AT = antithrombin.
† Data from National Committee for Clinical Laboratory Standards.[11]
‡ All reagents were obtained from Diagnostica Stago, Asnières-sur-Seine, France.
§ With veterinary configuration of the software.

CLSI (formerly NCCLS).[11] Duplicate measurements of each control solution were done in the morning and afternoon for 5 consecutive days. Repeatability in canine plasma was evaluated from the duplicate results obtained for each analysis.

All canine specimens were collected from the jugular vein of overnight-fasted dogs into a 1.8-ml 3.8% Na_3 citrate vacuum tube with a 0.8 mm × 40 mm needle.[b] The specimens were gently inverted 10 times, identified, and then stored at 4°C until centrifugation for 15 min at 1,300 × g.[c] Because of the distance of certain breeding locations, the maximum time before centrifugation and subsequent analyses was 7.5 hr, with most analyses being performed within 4 hr of sampling. Reference intervals were determined in a reference sample group composed of 139 dogs selected according to the following inclusion criteria: 1) purebred dogs (registered in the Livre des Origines Français) directly sampled at the breeder's kennel, 2) healthy animals (questionnaire and physical examination), and 3) dogs fasted overnight.

The proportions of the different breeds in the sample group were as representative as possible of the proportions of the breeds in the French canine population. When more than 5 dogs had to be sampled in a single breed, they were taken from at least 2 different owners. Exclusion criteria were incorrect sampling or filling of the tube, abnormal color of plasma, any reported illness in the month preceding sampling, abnormal bleeding, any treatment (except external antiparasitic), and nonfasted dogs.

Normality of distributions was tested by Anderson–Darling test, and reference intervals were determined by the nonparametric method according to IFCC-CLSI recommendations,[2] which is based on the ranking of the 2.5 and 97.5 percentiles. Because part of the present study required sample transport from the dog breeding establishments to the laboratory, stability of the analytes had to be tested in a preliminary study. Aliquots of 6 citrated canine whole blood specimens were analyzed by taking duplicate measurements within 1 hr of sampling, and after storage of whole blood for 4 hr and 8 hr at 4°C. Centrifugation was performed just before each run of analyses. Aliquots of 6 citrated canine plasma specimens were also analyzed by taking duplicate measurements within 1 hr of sampling, and then after storage for 12, 24, and 36 hr at 4°C. Transference of the reference intervals was based on a procedure of method comparison[5,10] using 60 canine specimens obtained from the plasma bank of the Central Laboratory. All calculations were performed with an Excel spreadsheet, a statistical analysis software,[d] and the freeware Reference Value Advisor (http://www.biostat.envt.fr/spip/spip.php?article63) for the determination of reference intervals.

Results

The within-laboratory imprecision and repeatability coefficients of variation (CVs) of the STA Satellite in the laboratory were similar or slightly higher than the manufacturer's indications. Bias was low except for PT in the "P" control solution and for fibrinogen

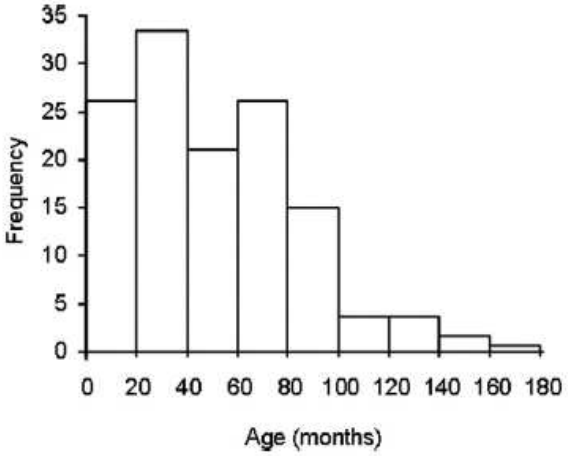


Figure 1. Distribution of ages of the 139 dogs used in the study of reference values of coagulation tests.

(up to 10.8% bias; Table 1). There was no statistically significant change of the analytes in whole blood stored up to 8 hr and in plasma stored up to 36 hr at 4°C (Wilcoxon test, with Bonferroni correction for multiple comparisons, $P > 0.05$), except for APPT, which was slightly (<1 sec) but significantly increased in whole blood at the eighth hour and in plasma at the 36th hour (Wilcoxon test, $P = 0.03$).

The main breeds of dog were Brittany (14), German Shepherd Dog (12), Labrador Retriever (10), Belgian Malinois (10), Yorkshire Terrier (9), Rottweiler (9), French Bulldog (9), and Australian Shepherd (9). There were 88 females, 2 spayed females, and 49 males. Ages ranged from 1 to 14 years (Fig. 1). The distributions of the 4 analytes in the reference sample group are presented in Figure 2. There was no outlier according to the Dixon–Reed criterion.[14] The AT distribution was not significantly different from Gaussian. All other distributions could be transformed into Gaussian, except for PT, which remained significantly different after all transformations tested. The best fit for PT was obtained by Box–Cox transformation (Anderson–Darling test, $P = 0.014$). Descriptive statistics with the corresponding reference intervals are summarized in Table 2. The nonparametric reference intervals thus determined were 6.9–9.0 sec, 13.1–17.7 sec, 1.24–4.62 g/l, and 104–188% for PT, APTT, fibrinogen, and AT, respectively. Partitioning according to sex was not relevant according to Harris–Boyd criterion,[4] even though the AT values were significantly lower in males than in females (Student's *t*-test after checking for homogeneity of variances, $P = 0.002$) with means of 136% and 147%, respectively. There were not enough values in each subgroup to make partitions according to age or breed.

For transference of the reference intervals to the STA-R Evolution analyzer, the results of some comparison analyses could not be obtained for technical reasons, so comparison was based on at least 53 paired analyses (Table 3). For AT, there was good correlation but a combined bias, with the results by STA-R Evolution being approximately 12% lower than with STA Satellite. Passing–Bablok equations were applied to the previously determined reference intervals to transfer reference intervals for the STA-R Evolution (i.e., 7.1–9.2 sec for PT, 12.9–17.3 sec for

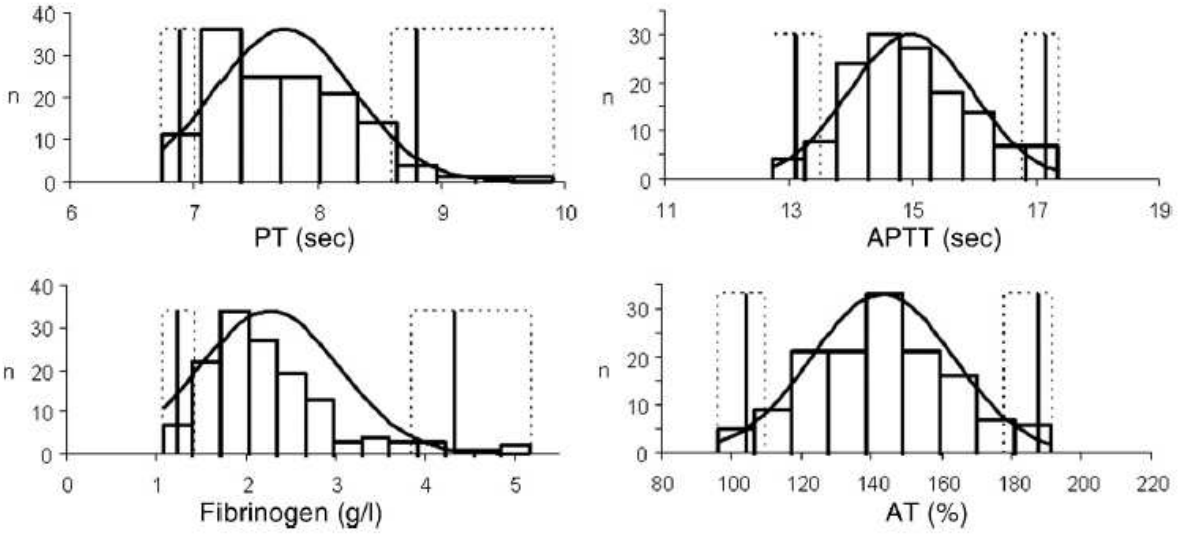


Figure 2. Histograms of prothrombin time (PT), activated partial thromboplastin time (APTT), fibrinogen, and antithrombin (AT) in 139 healthy dogs (black curve: Gaussian curve; vertical lines: reference limits with 90% confidence intervals).

Table 2. Descriptive statistics and reference limits of prothrombin time (PT), activated partial thromboplastin time (APTT), fibrinogen, and antithrombin (AT) in 139 healthy dogs, and characteristics of the distributions.*

	PT (sec)	APTT (sec)	Fibrinogen (g/l)	AT (%)
Mean	7.7	15.0	2.26	143
Minimum–Maximum	6.8–9.9	12.8–17.4	1.09–5.18	96–192
Normality, *P*	<0.0001	= 0.0406	<0.0001	= 0.58
Best Gaussian fit	—	Box–Cox > Ln	Box–Cox > 1/n	Box–Cox
Lower limit of reference interval	6.9 (6.8–7.0)	13.1 (12.8–13.5)	1.24 (1.09–1.43)	104 (96–110)
Upper limit of reference interval	8.8 (8.6–9.9)	17.2 (16.8–17.4)	4.30 (3.85–5.18)	188 (178–192)

* Numbers in parentheses are 90% confidence intervals.

APTT, 1.20–4.43 g/l for fibrinogen, and 94–159% for AT).

Discussion

The pre-analytical conditions were chosen according to recommendations for coagulation testing in human clinical pathology.[1,13] These conditions are routinely used in veterinary clinical pathology; although, to the authors' knowledge, they have never been validated except for absence of a significant effect of citrate concentration between 3.2% and 3.8%,[8] lack of effect of sampling by direct venipuncture or through a catheter,[7] significant effect of abnormal colors of plasmas,[9] and of dehydration.[12]

Prothrombin time, APTT, and fibrinogen are almost unchanged in canine plasma stored for up to 2 days at room temperature or at 4°C,[3] which is longer than the delays recommended for human specimens.[1] The durations of storage tested in the current study were chosen according to the actual conditions of the experiment where blood specimens collected at the individual breeders' locations usually arrived at the laboratory after less than 4 hr and, in a very few cases, after up to 7.5 hr. According to the results of the stability study, PT, fibrinogen, and AT would not have been altered even after 7.5 hr, but a moderate increase of APTT might have occurred. The latter change did not exceed the 90% confidence intervals of the reference limits, thus it should have no clinical relevance, which is the reason why the results were included in the study. Finally, the testing of plasma storage up to 36 hr was unnecessary: this had been done in case analyses needed to be postponed because of a technical difficulty.

Selecting the reference sample group according to "well-defined criteria" is the most difficult task when establishing reference intervals. The aim of selection is to make the reference individuals "as comparable to the patients as possible, except for the disease."[6] In the current study, an attempt was made to match the canine population in France by selecting dogs according to prevalence of the main breeds and by not taking more than 5 animals from the same breeder to avoid a "family" effect. There remained 3 biases. The first one is the arbitrary choice to sample only purebred animals. This was based on the fact that, to the authors' knowledge, the proportions and types of crossbreeds are not reported in the veterinary literature, thus determining the corresponding reference sample group was hazardous. Moreover, crossbreeds reflect the mixing of different breeds, and it can be expected, but not proven, that values in crossbreeds fall somewhere within values obtained in purebred dogs. The second bias is the sex ratio, which is close to 2 females for 1 male in this sample group. This results from the authors' choice to sample dogs from breeders who have more females than males. This ratio is much higher than usually reported in the general canine population, but this should have had very little influence on the reference intervals as Harris–Boyd test showed that no partitioning according to sex should be applied. The third bias is

Table 3. Comparison of results obtained with the STA Satellite and the STA-R Evolution analyzers.*

	n	*r*	a†	b†	Mean bias
PT	57	0.95	1.000 (0.980; 1.064)	0.200 (−0.345; 0.378)	0.2 sec
APTT	57	0.84	1.076 (0.969; 1.250)	−1.181 (−3.875; 0.477)	−0.1 sec
Fibrinogen	58	0.93	0.954 (0.918; 1.007)	0.024 (−0.125; 0.146)	0.0 g/l
AT	53	0.92	0.780 (0.725; 0.852)	12.540 (5.167; 19.477)	−11.8%

* *n* = number of values compared; *r* = Pearson correlation coefficient; PT = prothrombin time; APTT = activated partial thromboplastin time; AT = antithrombin.

† Parameters of the following Passing–Bablok equation: STA-R Evolution = a × STA Satellite + b. Numbers in parentheses indicate 95% confidence intervals.

because all specimens were collected from breeders within approximately 200 km of the authors' laboratory, both to ensure the stability of the specimens and allow rapid analyses on the day of sampling. It is likely that this "geographic" sampling bias had little effect on the reference interval results as populations of dogs have not been reported to differ notably between different regions of France.

The reference intervals were determined as recommended[2] by the nonparametric method. Except for AT, the distributions were statistically different from Gaussian, which might explain some of the differences with values previously reported for coagulation tests in dogs, which were mostly based on parametric approaches.

The transfer of the reference intervals from STA Satellite to STA-R Evolution is only a palliative method. It is recommended by IFCC-CLSI when reference populations can be expected to be identical, which was the case in the present study. However, there were some discrepancies in the results obtained by the 2 analyzers especially for AT, which may lead to a greater imprecision of the limits thus calculated for STA-R Evolution.

It has been reported that significant differences in coagulation test results could be due to the use of different reagents, especially for APTT. This supports the claim that each laboratory should determine its own reference interval for its specific conditions. However, such determinations are mostly performed from small samples with unverifiable assumptions about the underlying distributions. According to IFCC-CLSI, "individual laboratories should focus more on verifying reference interval established elsewhere."[2] In the present case, the reference limits determined in a sufficient number of dogs with the nonparametric method can be used for validation studies involving a limited number of specimens, as indicated by IFCC-CLSI.[2]

The reference intervals determined in the current study are similar to previously reported values, except for AT. This is due to the calibration of the techniques. In some studies, control dogs were used to determine the "normal" (i.e., 100%), after which the measurements were compared with this calibration, resulting in values centered on 100%. In other studies, including the present one, calibration was based on the use of a human calibration solution. The use of human calibrators has 2 major consequences: 1) it produces a bias, and 2) it probably results in a better transferability of results, as no canine calibrator is available to the authors' knowledge, and human calibrators are readily available and comparable in all settings.

According to IFCC-CLSI guidelines, the canine reference intervals determined in the present study can be used without prior validation by users of these 2 analyzers with the same manufacturer's reagents, for animals that fall into the same age range, and where the proportions of the canine breeds are almost identical to those in France. In other countries or with other analyzers or reagents, validation should be performed.

Acknowledgements

The authors are grateful to J. M. Palis and H. Doury (Diagnostica Stago Company) for technical help. Results partially presented at European Society of Veterinary Clinical Pathology Congress, Barcelona, Spain, September–October 2008. Reagents were offered by the manufacturer, who was not involved in study design, analyses, data interpretation, writing the manuscript, or the decision to submit for publication.

Sources and manufacturers

a. Diagnostica Stago, Asnières-sur-Seine, France.
b. Venoject®, Terumo Europe N.V., Leuven, Belgium.
c. Rotofix 32A, Andreas Hettich GmbH & Co. KG, Tuttlingen, Germany.
d. Analyse-it for Microsoft Excel (version 2.10), Analyse-it Software Ltd., Leeds, United Kingdom.

References

1. Clinical and Laboratory Standards Institute (CLSI): 2003, Collection, transport, and processing of blood specimens for testing plasma-based coagulation assays and molecular hemostasis assays, 5th ed., approved guideline H21-A5. CLSI, Wayne, PA.
2. Clinical and Laboratory Standards Institute (CLSI): 2008, Defining, establishing, and verifying reference intervals in the clinical laboratory, 3rd ed., approved guideline C28-A3. CLSI, Wayne, PA.
3. Furlanello T, Caldin M, Stocco A, et al.: 2006, Stability of stored canine plasma for hemostasis testing. Vet Clin Pathol 35:204–207.
4. Harris EK, Boyd JC: 1990, On dividing reference data into subgroups to produce separate reference ranges. Clin Chem 36:265–270.
5. Jensen AL, Kjelgaard-Hansen M: 2006, Method comparison in the clinical laboratory. Vet Clin Pathol 35:276–286.
6. Kairisto V, Hanninen KP, Leino A, et al.: 1994, Generation of reference values for cardiac enzymes from hospital admission laboratory data. Eur J Clin Chem Clin Biochem 32:789–796.
7. Maeckelbergh VA, Acierno MJ: 2008, Comparison of prothrombin time, activated partial thromboplastin time, and fibrinogen concentration in blood samples collected via an intravenous catheter versus direct venipuncture in dogs. Am J Vet Res 69:868–873.
8. Morales F, Couto CG, Iazbik MC: 2007, Effects of 2 concentrations of sodium citrate on coagulation test results, von Willebrand factor concentration, and platelet function in dogs. J Vet Intern Med 21:472–475.
9. Moreno P, Ginel PJ: 1999, Effects of haemolysis, lipaemia and bilirubinaemia on prothrombin time, activated partial thromboplastin time and thrombin time in plasma samples from healthy dogs. Res Vet Sci 67:273–276.

10. National Committee for Clinical Laboratory Standards (NCCLS): 2002, Method comparison and bias estimation using patient samples, 2nd ed., approved guideline EP9-A2. NCCLS, Wayne, PA.
11. National Committee for Clinical Laboratory Standards (NCCLS): 2005, User verification of performance for precision and trueness, 2nd ed., approved guideline EP15-A2. NCCLS, Wayne, PA.
12. O'Brien SR, Sellers TS, Meyer DJ: 1995, Artifactual prolongation of the activated partial thromboplastin time associated with hemoconcentration in dogs. J Vet Intern Med 9:169–170.
13. Polack B, Schved JF, Boneu B: 2001, Preanalytical recommendations of the 'Groupe d'Etude sur l'Hemostase et la Thrombose' (GEHT) for venous blood testing in hemostasis laboratories. Haemostasis 31:61–68.
14. Reed AH, Henry RJ, Mason WB: 1971, Influence of statistical method used on the resulting estimate of normal range. Clin Chem 17:275–284.

Canine hematology reference intervals for the XT-2000iV analyser
(in press in Veterinary Clinical Pathology)

Nathalie Bourgès-Abella, Anne Geffré, Didier Concordet, Jean-Pierre Braun, Catherine Trumel.

From the Small Animal Clinical Sciences Department (Bourgès-Abella, Braun, Geffré, Trumel) and the UMR181 Physiopathologie et Toxicologie Expérimentales,INRA,ENVT (Concordet, Braun), Ecole Nationale Vétérinaire de Toulouse, France.

This work was presented as a poster at the Congress of the European College of Veterinary Clinical Pathology, Barcelona, Spain, September 30 - October 3, 2008. Corresponding author: Nathalie Bourgès-Abella, Ecole Nationale Vétérinaire de Toulouse, 23 Chemin des Capelles, BP 87614, 31076 Toulouse Cedex 03, France (n.bourges-abella@envt.fr).

Key Words: Reference interval, hematology, dog, XT-2000iV, platelet

Abstract

Background: Laser-based hematology analyzer Sysmex XT-2000iV is increasingly used in large veterinary clinical pathology laboratories, and corresponding reference intervals are not available.

Objective: The purpose of this study was to establish canine hematology reference intervals for XT-2000iV according to IFCC-CLSI procedures.

Methods: Blood samples from 132 healthy pure breed dogs selected among the most prevalent canine breeds in France were analyzed. Blood smears were scored for PLT aggregates. Reference intervals were determined by the nonparametric method. PLT and RBC counts obtained by impedance and optical methods were compared. Effects of sex and age on reference intervals were determined.

Results: Means, distributions and reference intervals are presented. An excellent correlation was obtained between impedance and optical measurements of RBC (Pearson $r = 0.99$) and PLT counts ($r = 0.98$) but a significant difference between methods was apparent (Student's paired t-test $P<0.0001$ for both analytes). Differences between the sexes were not significant, except for HCT, PLT-I and PLT-O. WBC, lymphocyte and neutrophil counts decreased significantly with age (ANOVA $P<0.05$). Median eosinophil count was higher in Brittany spaniels (1.87 10^9/L), Rottweilers (1.41 10^9/L), German Shepherds (1.38 10^9/L) than in the overall population (0.9 10^9/L). PLT aggregates were responsible for lower PLT counts by impedance but not by the optical method.

Conclusion: Reference intervals for hematology analytes and calculated indices were determined according to international recommendations from a well characterized sample of dogs, under controlled preanalytical and analytical conditions.

Introduction

Reference intervals provide the best aids for interpreting clinical data in animal patients. For each new instrument, these reference intervals must either be transferred from a previous instrument or validated from pre-existing reference intervals. When transfer or validation are not possible, reference intervals must be determined *de novo* according to international recommendations by the International Federation of Clinical Chemistry (IFCC) and Clinical and Laboratory Standards Institute (CLSI), which have been recently updated.[1]

To our knowledge there is only one report describing canine hematology reference values by flow cytometry analysis. This was based on 46 dogs[2], i.e. much fewer than the minimum of 120 specimens recommended for use of the nonparametric method.[1,3] Moreover, reference intervals have not been studied with the Sysmex XT-2000iV which uses both impedance and flow cytometry and has been recently validated for the main hematological analytes of dogs, cats, horses, rats and mice.[4-6]

The major difficulty when determining a reference interval is the selection of a well characterized reference population. Pure breed dogs could provide a good basis if the number of selected animals is based on the prevalence of canine breeds in France.[7] The objective of this study was thus to establish canine haematological reference intervals for the Sysmex XT-2000iV according to the IFCC-CLSI recommended procedures.

Materials and methods

The experimental protocol was designed according to the "outline for obtaining reference values and establishing reference intervals" for a new analyte or analytical method[1], and was performed over one month between February and March 2008.

1/ Preanalytical factors were defined according to recommendations for blood collection and processing in veterinary clinical pathology.[8] Venipunctures were performed after clinical examination by an experienced phlebotomist on fasted dogs at rest in their kennels at early afternoon to limit possible variability due to transport of the animals. A 5-ml K_3-EDTA sample (VenoJect EDTA (K_3) K3E, Terumo Europe- Belgium) was taken from the jugular vein with a 0.8x40 mm needle (Venoject, Terumo Europe N.V., Leuven, Belgium) then gently homogenized by 10 inversions, identified and stored at +4°C prior to laboratory analyses. Most analyses were performed within less than 4 hours and the maximum time before analysis was 6 hours.

2/ Inclusion criteria were as follows: pure breed animals; proportions of the different breeds as representative as possible of the proportions of the same breeds in France; age $\geq$ 6 months; healthy animals based on a questionnaire and a clinical examination performed by an experienced veterinarian. This examination included measures of cardiac and respiratory frequencies, capillary refill time and rectal temperature, examination of the mucocutaneous, cardio-respiratory, digestive and musculoskeletal systems and palpation of the abdomen and mammary glands.

The animal *a priori* exclusion criteria were: lactation, heats, antecedents of disease within the last month; history of unusual bleeding; any treatment except external antiparasitic; any abnormality found during clinical examination; non-fasted dogs. Samples were excluded in case of incorrect filling of the tube or macroscopically visible clot before analysis.

A written informed consent form was completed by the owners.

3/ According to the recommendation,[1] it was decided that a minimum of 120 results needed to be obtained to allow use of the nonparametric method. Given that some specimens might not be usable or some animals might have to be excluded *a posteriori*, 137 dogs were sampled.

5/ Analyses were performed with a Sysmex XT-2000iV analyzer using the settings for canine samples (software version 00-09; Sysmex, Kobe, Japan) after testing for trueness and precision, although this had already been done for the most frequently measured analytes[4-6].

As canine reference blood samples are not commercially available, these analyses were carried out using the manufacturer's blood control samples at low, medium and high levels (Sysmex e-check L1, L2 and L3 levels) according to EP 15-A2,[9] by obtaining duplicate measurements in the morning and afternoon of five consecutive days. Repeatability in canine blood was determined from the duplicates measured in the subsequent study.

Duplicates of the whole panel of analyses available were obtained after daily checking with the 3-level control samples : RBC count by optical (RBC-O) and impedance (RBC-I) measurements, hemoglobin (Hb), HCT, MCV, MCH, MCHC, reticulocytes (RET), low, medium and high fluorescence ratios (LFR, MFR, and HFR) as grades of reticulocyte maturation, RBC distribution width (RDW) expressed as RDW-SD (standard deviation) and RDW-CV (coefficient of variation), WBC count, neutrophil, lymphocyte, monocyte, and eosinophil counts, platelet count by optical (PLT-O) and impedance (PLT-I) measurements, mean platelet volume (MPV), platelet distribution width (PDW), platelet large cell ratio (P-LCR) as indice of platelet activity and plateletcrit (PCT). MPV, PCT, PDW and P-LCR were analyzed using impedance methodology. The basophil counts are not reported as they have been shown to be unreliable in canine samples.[5,6]

A blood smear was made on each sample and stained with May-Grünwald Giemsa to assess possible PLT aggregates as described by Norman for cats.[10] A first examination of the smear was made under low power (x100 and x200 magnifications) to check for platelet's aggregates, particularly at the feathered edge. If platelet clumps were observed, a random observation was done under high magnification (x1000) on 10 fields selected at the feathered edge and the aggregate size was evaluated. The following scores were assigned: 5 for n≥ 30 platelets; 4 for n=20-29 platelets; 3 for n=10-19 platelets; 2 for n=5-9 platelets; 1 for n=2-4 platelets and 0 for no aggregate. The mean score was calculated by averaging the scores obtained in 10 high magnification fields (Eclipse 50i, x100 oil-immersion objective- ocular field number x22, Nikon, France).

6/ The histograms of all the results were visually inspected to detect possible outliers. As the recommendation[1] states that "emphasis should be on retaining rather than deleting" outliers, only values outside the median ± 3 interquartile range were deleted. The normality of the distributions of native or of transformed values was tested by Anderson-Darling test. Finally, the reference intervals and 90% confidence intervals of the limits were determined by nonparametric method using the Reference Value Advisor macroinstructions[11,12] for Excel®. Partitioning of the reference values according to sex was based on Harris & Boyd's z test.[13] Reference intervals for the subgroups were estimated by parametric and robust methods from the distribution giving the best fit with a gaussian distribution.[14] Partitioning according to age was impossible due to the low number of reference individuals in the subgroups. Possible effects of age were therefore estimated by determining regression-based reference limits based on the 95% prediction intervals of the polynomial regression of reference values vs. age.[15] Comparisons of impedance and optical PLT and RBC counts were based on EP15-A2[9] and the general recommendations for method comparison[16,17] using Passing-Bablok's regression analysis and Bland-Altman diagrams of difference with an Excel spreadsheet and the Analyse-It set of macroinstructions (Analyse-it software, Ltd., Leeds, United Kingdom).

Results

Analytical characteristics

Results of trueness and precision testing of all analytes in control solutions at three levels are given in Table 1. The coefficients of variation (CV) of within-laboratory imprecision were below the manufacturer's indications. Whatever the blood control level, low CVs (less than or equal to 1.3%) were observed for RBC-O, RBC-I, Hb, HCT, MCV, MCH, MCHC and high

CVs (between 9.8% and 41.3%) for MFR, HFR and IRF. Repeatability CVs estimated from the duplicates of this study were also lower than the manufacturer's indications (Table 1).

Score of platelet aggregates

Platelet aggregates were observed in 73 out of 132 dogs (55.3%), but in most cases the mean score of platelet aggregation was low (score<1: 6.1% (n=8); score≥1 and <2: 28.8% (n=38)). High scores (≥2) were observed in 27 samples (20.5%), which were considered as potential sources of error for the platelet count. It was thus decided to consider these samples as possible outliers and to evaluate PLT-I and PLT-O counts in the whole reference population and also in the population from which animals with aggregates score≥2 had been removed.

Comparison of results obtained by impedance and optical methods for RBC and PLT

Comparison of the RBC and PLT counts are summarized in Table 2.

RBC counts showed a high correlation with a Pearson r coefficient of 0.99 but the differences were highly significant (Student's paired t-test, P<0.0001). Bias was proportional. Differences (O-I) were significantly different in the 25% highest and 25% lowest values and the mean differences were -0.18 and 0.33 10^{12}/L (Student's t-test after checking homogeneity of variances, P<0.001).

PLT counts (all PLT counts (n=132) and PLT count in specimens with no or moderate aggregates (score<2) (n=105)) showed a high correlation with Pearson r coefficients of 0.98 and 0.99 respectively. However, the differences were highly significant (Student's paired t-test, P<0.0001). Biases were proportional. Differences (O-I) were significantly different in the 25% highest and 25% lowest values. The mean differences (O-I) were -42.0 and 18.2 10^9/L for whole PLT count and -43.9 and 7.3 10^9/L for the PLT count in specimens with no or moderate aggregates (Mann-Whitney's test after checking heterogeneity of variances, P<0.001).

Characteristics of the reference population

Five specimens were excluded *a posteriori* on account of visible clots or insufficient blood volume at inspection before counting and 132 specimens were finally analyzed. The population distribution according to breed was representative of the most prevalent breeds in France, except for Australian Shepherd dogs, Brittany spaniels and English pointers which were over represented and English setters which were absent (Table 3).

There were 83 females (breeding bitches in *anoestrus)*, 2 neutered females, and 47 males. Dog age ranged from 6 months to 14 years (9 dogs were less than one year-old and 4 dogs older than 11 years) with a median of 43 months. 85.6 % of the dogs were between 1 and 8 years old (Figure 1). There was no effect of sex on age, the means being 51.8 and 50.4 months in males and females respectively (Student's t-test after checking the homogeneity of variances, P = 0.999).

Reference intervals

Results are presented separately in Table 4 and Figure 2 for analytes which had been previously validated and in Table 5 for other results given by the analyzer. No outlier was visible, on visual inspection of the histograms or according to Tukey's criterion, for most of the analytes. All outliers deleted from further analysis are indicated in the footnotes of Tables 4 and 5. Results could not be obtained for MPV, P-LCR, PCT and PDW in 6 specimens due to poor separation of RBC and PLT by impedance in these specimens.

Many distributions were significantly different from Gaussian (Anderson-Darling test, P <0.05) but not after Box-Cox transformation. However, the distribution of HCT could not be transformed into a normal one whatever the transformation tested.

In non- and moderately-aggregated specimens (n =105, score <2), the lower limit of impedance PLT was notably higher than in the whole reference sample (n=132), and the upper limit was unchanged. The corresponding PLT reference intervals (90% confidence interval in parentheses) were: 115. 9 (62.5-151.6) to 559.4 (517.0-633.0) 10^9/L for PLT-O and 125.9 (47.0-163.5) to 608.4 (535.0-661.1) 10^9/L for PLT-I..

Effects of sex, age and breed

As there were only 2 spayed females in the reference group, these were not included in the study of sex effects on the measured variables. According to Harris & Boyd's z factor, sex was a partitioning factor only for HCT and PLT. Reference intervals evaluated for these analytes in the sex subgroups by the robust method on Box-Cox transformed data are given in Table 6. There was very little difference between the sexes.

Significant decreases of WBC, lymphocyte and neutrophil counts were observed with age (ANOVA $P < 0.05$) (Figure 3) with more marked decreases for lymphocytes which reduced by about 50% between 1 and 9-10 years. There was no significant changes for other analytes (ANOVA $P > 0.05$).

A systematic review of the results, breed by breed, showed that the median eosinophil count was higher in Brittany spaniels (1.87 10^9/L), Rottweilers (1.41 10^9/L) and German Shepherd dogs (1.38 10^9/L) than in the overall population (0.9 10^9/L). When these 3 breeds were eliminated from the whole set of values, the estimated reference interval for the 100 remaining dogs, based on the robust method and Box-Cox transformed data, was 0-1.5 10^9/L.

Discussion

Preanalytical conditions

The control of preanalytical factors is essential to minimize possible effects on clinical decision making. The preanalytical conditions were therefore chosen according to recommendations for collecting and handling a blood sample in veterinary clinical pathology.[8] Venipunctures were performed after clinical examination by an experienced phlebotomist on resting fasted dogs. Blood collection was performed from the jugular vein because it is easy to access in dogs and results obtained are not different from those of blood taken from the cephalic vein.[8,18] Blood collection tubes were stored in the cold and rapidly transported to the laboratory to limit possible preanalytical alterations of specimens.

Analytical performances of the analyzer

The reliability of reference intervals is partly based on the reliability of the analytical methods used and according to international recommendations analytical performances must be tested and reported[1]. The XT-2000iV had already been validated with canine blood for most routinely used analytes and indices except basophils.[4-6] However, it has not been validated for the following analytes: LFR, MFR, HFR, IRF, MPV, P-LCR, PCT and PDW. These latter results are therefore reported in a separate table. Such information might be useful in further studies although the clinical value of these new indices has been little investigated in human [19-21] or veterinary hematology.[22,23]

In our hands, the within-laboratory precision of the analyzer with human blood and its repeatability with canine blood was satisfactory, as shown by the low CVs measured. Trueness could only be tested with human control samples; in each case, the results obtained with the analyzer were within the manufacturer's range of acceptability.

This analyzer provides both optical and impedance measurements of RBC and PLT counts. The correlation between the two sets of measurements was excellent, but there was a significant difference between the two methods, which would indicate that two different reference intervals should be considered according to the methodology adopted. However, the

differences between the two methods were so slight that very few misclassifications were possible, except for two platelet counts which showed thrombopenia only with the impedance measurement. In one case, no reason could be detected from the impedance or optical histograms but numerous PLT aggregates were observed on the blood smear (score of 3.5). In the second case, a Cavalier King Charles (CKC) spaniel, the separation between PLT and RBC by impedance was not satisfactory. The optical PLT dot plot was close to the RBC cloud showing the presence of macro-platelets. This was confirmed by blood smear examination which also revealed PLT aggregates (score of 2). Such idiopathic thrombocytopenia is a common laboratory finding in CKC spaniels, resulting in discordant measurements between impedance and optical methods.[24-26] Nevertheless, as previously reported in samples with numerous large platelets, the PLT-O count appeared to be more accurate and the best reference interval to use for PLT would be the one from the PLT-O method even though there was good agreement between both methods.[4]

Selection of the reference sample

The selection of a reference group is the most difficult task when establishing "health-associated" reference intervals. In this study, an attempt was made to match the canine population according to the prevalence of the main breeds in France. However, we were limited to south-west France and to within 200 kilometres from Toulouse to ensure specimen stability and analyses within a few hours of sampling. This local sampling led to a biased selection of dogs, e.g. English setters, one of the top 10 breeds registered in France in 2006, were absent from the reference sample analyzed. However the reference sample in this study closely resembled the distribution of canine breeds in France. Another bias resulted from the choice to use only pure breed dogs. As the proportions and types of cross breed dogs have not been documented to our knowledge, it was impossible to have a representative selection of these animals. Moreover as they result from the mixing of different breeds and no major difference was observed according to breed, it is reasonable to consider that cross breed dogs would fit within the reference intervals determined for pure breeds, even if no definitive proof could be found. Finally, the collection of samples in kennels may have limited the intraindividual factors of variation as compared to animals presented in veterinary practices, which may thus be expected to show a higher variability of blood analytes.

The distribution of breeds may be different in other countries. As long as a possible breed effect has not been more extensively investigated, the reference intervals established here will have to be validated in other settings if demographic conditions are different. This can be done with a limited number of reference samples under local laboratory conditions, as recommended by IFCC-CLSI.[1]

As advised in the CLSI guidelines the reference individuals for the determination of a health-associated reference interval do not necessarily have to be only young adults but should resemble as much as possible the patient population.[1] This is why the animals sampled in this study covered a large range of ages. However, most animals in kennels are young animals kept for reproductive capacity, which explains why most of the animals were 1 to 8 years old. It was also decided to exclude *a priori* dogs younger than 6 months because previous studies had demonstrated major changes in young dogs.[27-29] The use of dogs in breeding units would also explain the overrepresentations of females (which constituted approximately 2/3 of the reference sample) compared to males.

Methods of determination of reference interval

The reference intervals were determined, as recommended, by the nonparametric method.[1] This method could be used on the whole set of animals as the total number of reference individuals was above 120 and very few outliers were finally identified in the native or

transformed distributions. As recommended, "the emphasis [has been] on retaining rather than deleting them [the outliers]".[1] A complete determination of reference intervals in partitioning groups such as age, sex or breed would have required at least 120 reference individuals to be sampled in each subgroup i.e., an enormous and expensive task. For the sex effect, sufficient animals were available to allow the use of parametric procedures[1] although large biases have been reported in small samples.[30] A regression analysis was used to determine the age effect, although it was known *a priori* that imprecision would be high due to the low number of values available.[15]

Comparison of reference intervals determined in this study with data from literature

The most commonly cited canine haematology reference intervals[31,32] originate from textbooks published in 1961 and 1965 for the main analytes and in 1975 for platelets.[33] These reference intervals have been re-cited in most textbooks, including recent ones.[34,35] However, they do not meet the currently accepted international recommendations[1,3] as preanalytical and analytical conditions, population characteristics and statistics procedures are not reported. Moreover analytical procedures have been greatly improved and updating of the determination of reference intervals is warranted. A first attempt with the ADVIA[2], based on a limited reference sample which was too small for the recommended nonparametric method to be used, was published recently. However, as shown in Table 4, most intervals were similar to those of Schalm's textbooks[31-33] and ADVIA's study[2]. The three major differences were observed for reticulocytes, HCT and platelets. Results from such analytes as MCHC, MCV and HCT are very dependent on adjustments in the particular instrument, species settings and software version, so they should not be expected to be the same even between two instruments from the same producer.

As previously reported,[6] the reticulocyte counts obtained with XT-2000iV was higher than with ADVIA 120. These analysers use different methodology for reticulocyte counting. Thus, different reference values should be expected and used.

The higher reticulocyte count in this study could be correlated with an active erythropoïesis related to blood loss occurring during annual to biennial whelping in breeding bitches. However there was no difference between males and females.

The reference interval for HCT with XT-2000iV was lower than with ADVIA 120 but similar to that of Schalm[31]. The latter was obtained by centrifugation[36] whereas XT-2000iV sums up all individual volumes of the red blood cells counted and ADVIA 120 calculates HCT from RBC count and MCV. The observed difference could thus result from method differences.

The platelet reference intervals observed in this study with impedance and optical methods were wider than the commonly used 200-500 10^9/L[33] although a recent textbook has indicated 166-575 10^9/L[37]. The lower limit obtained by impedance in this study was particularly low (63.7 10^9/L) but twice higher (125.9 10^9/L) when established from non- and moderately-aggregated specimens (n=105). Moreover, the PLT-I reference interval determined on totally non-aggregated specimens (n=59) was very close to that determined from 105 samples (data not shown). On the other hand, whatever the reference sample or subgroup, the PLT-O reference intervals were nearly the same, indicating that platelet aggregation had little effect on this method. This confirms that examination of possible platelet aggregation is critical to the interpretation of low PLT counts, especially if measurements are made with an impedance analyzer.

Basophil counts have not been reported in this study because previous studies demonstrated very poor agreement between XT-2000iV and manual differential counts.[5,6] A similar discrepancy has already been observed with XT-2000iV and CELL-DYN analyzers[5] and high imprecision has been reported for low basophil counts with Advia 2120 and XT 2000iV.[38]

Effects of age, sex and breed

The reference animals were mostly young adults as about 85% were 1 to 8 years old. Thus partitioning according to age groups was almost impossible due to the low numbers of younger and older individuals. Moreover, young or old age is often considered to be breed-dependent: adulthood in larger dogs being attained at 15 months of age whereas small-breed dogs are considered adult at 9 months of age.[39] In contrast, larger breeds become senior and geriatric earlier than smaller ones.[40] Thus, except for the 3 significant effects observed in WBC, lymphocytes and neutrophils, no valid conclusions about the effect of age could be drawn from this study.

A decrease in WBC with age has already been reported in laboratory Beagles[28,41,42] and in Labrador retrievers.[43] In the latter, a fifty percent decrease of lymphocyte count has also been observed, as in this present study. This should be further investigated and likely taken into account when interpreting results in diseased dogs.

Partitioning according to sex was only relevant for hematocrit and platelet count, according to Harris & Boyd's z criterion. This clinically relevant criterion was chosen because "any observed difference, no matter how small or how questionable its clinical significance, can be statistically significant if the sample sizes are large enough".[13] Higher HCT values in males than in females are in agreement with the higher HGB concentration previously reported in laboratory Beagles[27,44]. This could result from sex differences but could also be related to blood loss in breeding bitches which were the major part of our female reference population sample. As HCT is tightly related to RBC, MCV and Hb, it was expected to observe the same sex-related pattern in at least one of these analytes but these did not fit Harris and Boyd's partition criteria.

The present sample could not be used to investigate a possible breed effect on hematology variables due to the low numbers of animals in each breed. However, a high eosinophil count has already been reported in German Shepherd dogs and Rottweilers[45], but not in Brittany spaniels. The higher RBC, HGB and HCT[46] values reported in German Shepherd dogs, Boxers and Dachshunds were not observed in this study.

Conclusion

Reference intervals for hematology analytes and calculated indices were determined according to international recommendations from a well characterized sample of dogs, under controlled preanalytical and analytical conditions. These reference intervals can be directly applied by laboratories using the same equipment, analogous analytical performances with populations of dogs analogous to the one studied here. For laboratories in which the patient demographics may differ, further validation of reference intervals should be done prior to use by the receiving laboratory

Acknowledgments

The Sysmex analyzer used in this study was made available as a free loan from Sysmex Europe GmbH.

Table 1: Trueness, within-laboratory imprecision and repeatability of blood analytes with the Sysmex XT-2000iV analyzer. Repeatability was estimated from duplicates of measurements in canine samples used for the determination of reference intervals. Trueness and imprecision were estimated according to EP15-A2 with the manufacturer's three levels of human control solutions L1, L2 and L3. For comparison, imprecision is indicated in last colomn according to manufacturer's information (no manufacturer's precision data were available for RBC-O)

		Within laboratory imprecision									Repeatability	Manufacturer's imprecision
		Level L1			Level L2			Level L3				
Analyte	Unit	Expected range	Mean measured	CV (%)	Expected range	Mean measured	CV (%)	Expected range	Mean measured	CV (%)	CV (%)	CV (%)
RBC-O*	10^{12}/L	2.16-2.64	2.33	1.3	3.87-4.73	4.16	0.8	4.69-5.73	5.05	1.0	1.00	-
RBC-I*	10^{12}/L	2.24-2.48	2.36	0.6	4.27-4.53	4.38	0.5	5.04-5.58	5.29	0.5	0.58	1.5
Hb*	g/L	56-60	58.4	0.9	117-125	120.8	0.8	156-166	161.5	0.6	0.53	1.5
HCT*	L/L	0.169-0.191	0.179	0.7	0.341-0.377	0.359	0.4	0.447-0.495	0.472	0.5	0.52	1.5
MCV*	fL	72.5-80.1	76.07	0.5	77.5-85.7	82.14	0.2	84.3-93.1	89.22	0.2	0.19	1.5
MCH*	pg	23.4-25.8	24.70	1.0	26.1-28.9	27.58	0.8	28.8-31.8	30.56	0.7	0.58	1.5
MCHC*	g/L	303-341	324.4	1.1	317-357	335.6	0.9	321-363	342.4	0.7	0.56	2.0
RDW-SD*	fL	37.8-51.2	42.73	0.4	39.0-52.8	45.20	0.5	38.2-51.8	44.65	1.0	0.50	3.0
RDW-CV*	%	13.9-18.7	15.78	0.9	13.4-18.2	15.48	0.7	12.3-16.7	14.45	0.6	1.20	3.0
RET*	10^9/L	0.12-0.24	0.16	4.6	0.07-0.14	0.10	3.9	0.039-0.080	0.045	4.7	7.68	15.0
RET*	%	4.97-10.33	6.94	4.6	1.55-3.23	2.29	4.1	0.73-1.51	0.92	4.7	7.78	15.0
WBC*	10^9/L	2.65-3.25	3.02	2.5	6.45-7.27	6.89	1.6	15.78-17.80	17.17	1.2	1.32	3.0
Neutrophils*	10^9/L	0.88-1.64	1.35	3.7	2.29-4.25	3.31	2.4	6.34-11.78	9.28	1.7	1.80	8.0
Lymphocytes*	10^9/L	0.61-1.43	1.05	4.0	1.46-2.72	2.10	6.2	2.90-5.38	4.31	1.4	2.93	8.0
Monocytes*	10^9/L	0.04-0.74	0.34	8.7	0.08-1.54	0.75	7.2	0.17-3.33	1.68	4.5	5.46	20.0
Eosinophils*	10^9/L	0.14-0.42	0.29	7.1	0.34-1.04	0.73	7.3	0.92-2.76	1.90	8.6	4.98	25.0
PLT-O*	10^9/L	29-75	54.4	5.3	172-232	201.0	3.2	431-583	473.2	2.5	3.38	6
PLT-I*	10^9/L	33-77	52.8	5.1	180-244	206.0	2.3	432-560	474.3	1.6	2.16	4.0
LFR†	%	58.8-88.8	73.60	3.7	59.4-89.4	71.55	3.9	68.5-98.5	79.88	3.9	3.64	30.0
MFR†	%	8.4-34.4	21.78	9.9	7.9-33.9	23.44	10.3	1.3-27.3	16.67	12.2	22.19	50.0
HFR†	%	0.0-9.8	4.63	21.7	0.0-9.7	5.02	15.1	0.0-7.2	3.46	41.3	35.03	100.0
IRF†	%	8.2-44.2	26.40	10.2	7.6-43.6	28.46	9.8	0.0-34.5	20.12	15.5	18.14	30.0
MPV†	fL	7.3-9.9	9.00	2.7	7.8-10.6	9.46	1.3	7.8-10.6	9.43	0.6	1.45	4.0
P-LCR†	%	1.6-20.0	12.65	10.0	6.7-20.3	14.98	4.9	6.5-19.7	14.21	3.2	2.82	18.0
PCT†	L/L	0.0002-0.0008	0.0005	9.1	0.0014-0.0026	0.0019	3.2	0.0034-0.0056	0.005	1.7	2.72	6.0
PDW†	fL	6.1-9.1	7.78	4.7	6.8-10.2	8.66	3.1	6.9-10.3	8.59	1.7	3.19	10.0

* analytes previously validated for canine blood[4-6] ; †: analytes not validated.

Table 2: Passing-Bablok regression equations of impedance and optical RBC and PLT measurements in canine blood with Sysmex XT-2000iV analyzer.

y= a * x + b	95%CI		
	a	b	
RBC-I = 1.10 * RBC-O – 0.36	1.07/1.12	-0.51/-0.23	10^{12}/L
PLT-I = 1.10 * PLT-O – 11.75†	1.06/1.13	-21.79/-1.62	10^9/L
PLT-I = 1.07 * PLT-O – 3.42‡	1.03/1.11	-13.51/9.31	10^9/L

†: all specimens (n= 132) included; ‡: specimens with no or moderate aggregates (n = 105)

Table 3: Comparison of the distribution of canine breeds in the reference sample group with major breeds in France.

	Canine breeds		
	Breeds in France' (%)	Reference sample	
		N	%
English setter	8.4	-	-
German Shepherd	7.6	11	8.3
Cavalier King Charles	6.7	6	4.5
Brittany spaniels	6.6	14	10.6
Golden Retriever	6.5	6	4.5
American Staffordshire Terrier	5.5	6	4.5
Yorkshire	5.4	9	6.8
Labrador	5.2	10	7.6
French Bulldog	4.9	8	6.1
Cocker Spaniel	4.6	6	4.5
Rottweiler	4.1	7	5.3
Belgian Shepherd (Malinois)	3.7	8	6.1
Beagle	3.3	-	-
Dachshunds (wired hair)	3.1	5	3.8
Boxer	3.0	4	3.0
Beauceron	2.9	2	1.5
Bernese Mountain dog	2.9	3	2.3
English Pointer	2.5	8	6.1
West Highland White Terrier	2.4	7	5.3
Pointers (German Shorthaired)	2.3	-	-
English springer spaniel	2.1	-	-
Shih Tzu	2.1	3	2.3
Australian Shepherd	2.0	9	6.8
Chihuahua	2.0	-	-
Total		132	100

Table 4: Mean and reference intervals (RI) of blood analytes previously validated and calculated indices with the Sysmex XT-2000iV hematology analyzer (n = 132, except when outliers were deleted, see footnotes). Normality testing was done by Anderson-Darling test on untransformed (N) and Box-Cox transformed values. Previously reported RI are provided for comparisons.

		Sysmex RI				Previously reported RI	
Analyte	Unit	Mean*	2.5th centile (90% CI)	97.5th centile (90% CI)	Normality P	Unreported equipment[31-33]	ADVIA 120 [2]
RBC-O	10^{12}/L	6.3	5.1 (4.73-5.26)	7.6 (7.32-8.02)	N : 0.7153 Box-Cox : 0.719	-	5.68-9.08
RBC-I	10^{12}/L	6.6	5.2 (4.89-5.39)	7.9 (7.67-8.52)	N : 0.7103 Box-Cox : 0.705	5.5-8.5[31]	-
Hb	g/L	158	124.0 (120-129)	191.5 (183-200)	N : 0.38 Box-Cox : 0.413	120-180[31]	137.7-203.8
HCT[b]	L/L	0.43	0.35 (0.33-0.36)	0.52 (0.50-0.54)	N : 0.0135 Box-Cox : 0.016	0.37-0.55[31]	0.42-0.62
MCV	fL	66	60 (56.4-61.0)	71 (70.3-73.0)	N : 0.8253 Box-Cox : 0.884	60.0-77.0[31]	62.7-74.56
MCH	pg	24	22 (20.5-22.6)	26 (25.8-26.9)	N : 0.5921 Box-Cox : 0.516	19.5-24.5[31]	20.46-24.81
MCHC	g/L	366	344 (328-353)	381 (379-383)	N : 0.0110 Box-Cox : 0.183	320-360[31]	316.1-343.5
RDW-SD	fL	35.1	31.1 (3.5-32.3)	38.9 (38.3-41.7)	N : 0.4923 Box-Cox : 0.791	-	-
RDW-CV[c]	%	16.2	13.2 (12.5-13.5)	19.1 (18.9-19.4)	N : 0.3276 Box-Cox : 0.328	-	12.00-13.15
Reticulocytes	10^9/L	58.2	19.4 (12.5-20.9)	150.1 (120.1-168.3)	N : <0.0001 Box-Cox : 0.150	-	10.92-110.96
Reticulocytes	%	0.89	0.30 (0.22-0.32)	2.3 7(1.99-2.56)	N : <0.0001 Box-Cox : 0.140	0.0-1.5[31]	0.14-1.48
WBC*	10^9/L	11.0	5.6 (4.89-5.93)	20.4 (19.43-21.66)	N : <0.0001 Box-Cox : 0.927	6.0-17.0[32]	5.84-20.26
Neutrophils*	10^9/L	6.6	2.9 (2.54-3.52)	13.6 (12.32-15.46)	N : <0.0001 Box-Cox : 0.106	3.0-11.5[32]	4.27-9.06
Lymphocytes*	10^9/L	2.6	1.14 (0.66-1.37)	5.28 (4.74-5.81)	N : <0.0001 Box-Cox: 0.461	1.0-4.8[32]	2.04-4.66
Monocytes	10^9/L	0.7	0.35 (0.30-0.39)	1.56 (1.43-1.72)	N : <0.0001 Box-Cox : 0.522	0.15-1.35[32]	0.24-2.04
Eosinophils	10^9/L	0.9	0.13 (0.02-0.16)	3.05 (2.66-3.41)	N : <0.0001 Box-Cox : 0.154	0.10-1.25[32]	0.10-1.20
PLT-O[b]	10^9/L	316	108 (62.5-136.5)	562 (525.5-720.5)	N : 0.4250 Box-Cox : 0.791	-	173.1-486.5
PLT-I[b]	10^9/L	330	63.7 (15.5-137.0)	613 (547.5-772.0)	N : 0.5017 Box-Cox : 0.887	200-500[33]	-

*: calculated on untransformed data; [a] and [b]: see differences according to age (Figure 3) and to sex (Table 6) respectively; [c]: technics of RDW determination are different in Sysmex XT2000iV and ADVIA120.

Table 5: Mean and reference intervals (RI) of unvalidated blood analytes and calculated indices with the Sysmex XT-2000iV hematology analyzer (n = 132; for MPV, P-LCR, PCT and PDW, n = 126 owing to insufficient PLT/RBC impedance discrimination in 6 samples). Normality testing was done by Anderson-Darling test on untransformed (N) and Box-Cox transformed values.

		Sysmex RI			
Analyte	Unit	Mean	2.5th centile (90% CI)	97.5th centile (90% CI)	Normality P
LFR	%	83.3	63.7 (58.1-68.5)	93.8 (93.3-94.8)	N : 0.0001 Box-Cox : 0.837
MFR	%	11.6	4.1 (2.6-4.5)	23.6 (21.9-25.8)	N : 0.0029 Box-Cox : 0.491
HFR[f]	%	5.1	1.2 (0.9-1.5)	14.3 (11.1-16.9)	N : <0.0001 Box-Cox : 0.808
IRF	%	16.7	6.2 (5.3-6.8)	36.3 (31.5-42.0)	N : 0.0001 Box-Cox : 0.986
MPV*	fL	10.62	9.05 (8.85-9.30)	12.68 (11.95-13.00)	N : 0.079 Box-Cox : 0.283
P-LCR	%	30.21	16.13 (12.90-18.30)	49.16 (44.85-51.50)	N : 0.54 Box-Cox : 0.249
PCT	L/L	0.0035	0.0014 (0.0005-0.002)	0.006 (0.005-0.007)	N : 0.044 Box-Cox : 0.260
PDW	fL	12.45	9.30 (8.85-9.85)	18.95 (17.20-20.60)	N : <0.0001 Box-Cox : 0.841

Outliers : f : HFR (f= 20.05)

* Previously reported RI for ADVIA 120 (8.56-14.41) is reported for comparison.

Table 6: Reference intervals for hematocrit and platelet counts partitioned by sex using the robust method on Box-Cox transformed values (n=47 males and 85 females). Partitioning criteria was according to Harris and Boyd[13] (z* = 2.2079). Normality was tested by Anderson-Darling test from untransformed (N) and Box-Cox transformed values.

		Normality	Reference interval	
		P	2.5th centile (90% CI)	97.5th centile (90% CI)
HCT (L/L) z=2.2296	Males	N: 0.2415 Box-Cox: 0.4555	0.37 (0.36-0.38)	0.52 (0.50-0.54)
	Females	N: 0.0293 Box-Cox: 0.0871	0.34 (0.33-0.36)	0.50 (0.49-0.51)
PLT-I (10^9/L) z=3.763	Males	N: 0.6257 Box-Cox: 0.8681	77.1 (47.3-109.5)	517.9 460.7-573.1)
	Females	N: 0.6095 Box-Cox: 0.6063	100.7 (56.0-146.0)	619.4 (576.4-664.4)
PLT-O (10^9/L) z=3.659	Males	N: 0.4218 Box-Cox: 0.7873	85.4 (60.6-112.6)	506.5 (449.6-563.9)
	Females	N: 0.7007 Box-Cox: 0.8641	133.7 (105.6-167.1)	595.3 (552.0-639.4)

Figure 1: Distribution of ages and sexes in 132 dogs (47 males and 85 females) used to determine canine hematology reference intervals with the XT-2000iV (horizontal bars show the medians).

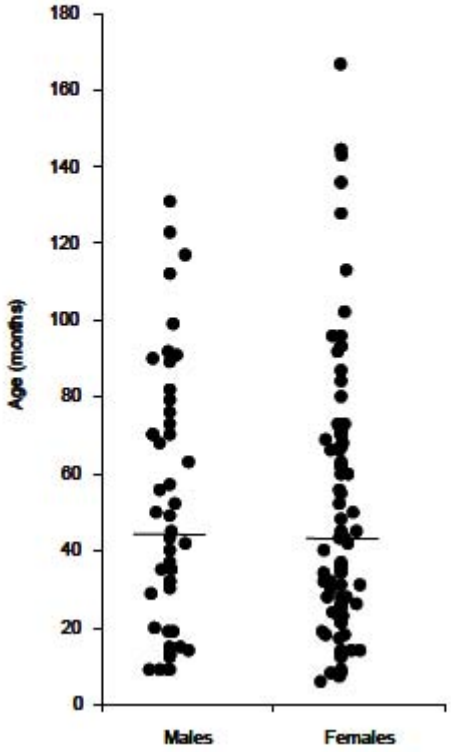


Figure 2: Observed (blue boxes) and fitted (purple line) distributions of blood hematology analytes in 132 healthy dogs. Blue vertical lines are the limits of the reference interval (with corresponding 90% confidence interval as dotted lines).

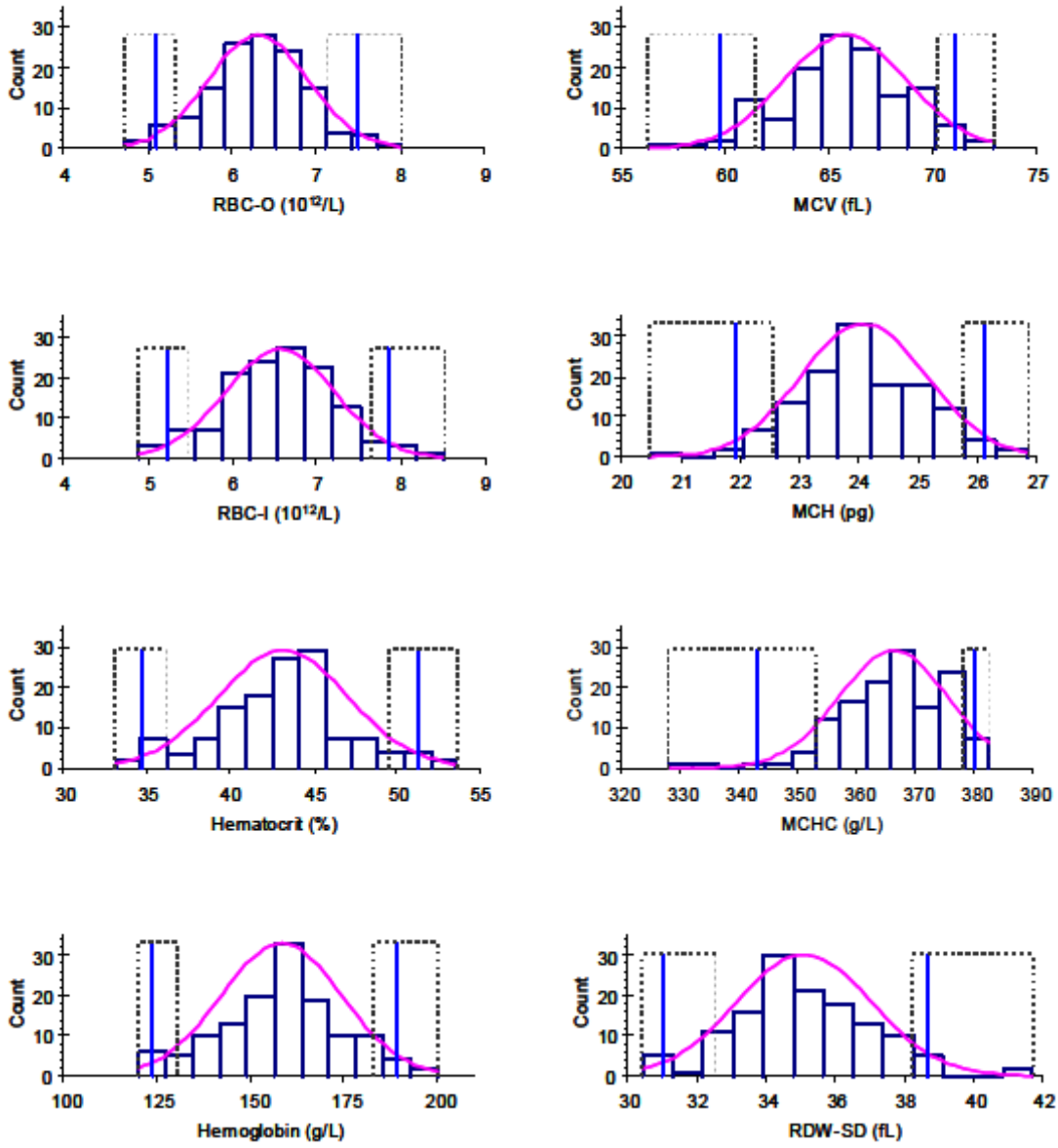


Figure 2 (continued-1): Observed (blue boxes) and fitted (purple line) distributions of blood hematology analytes in 132 healthy dogs. Blue vertical lines are the limits of the reference interval (with corresponding 90% confidence interval as dotted lines). Outliers were deleted for WBC (1 = 27.68); neutrophils (2 = 17.34 & 21.2); monocytes (1 = 2.38) and eosinophils (1 = 6.34) counts.

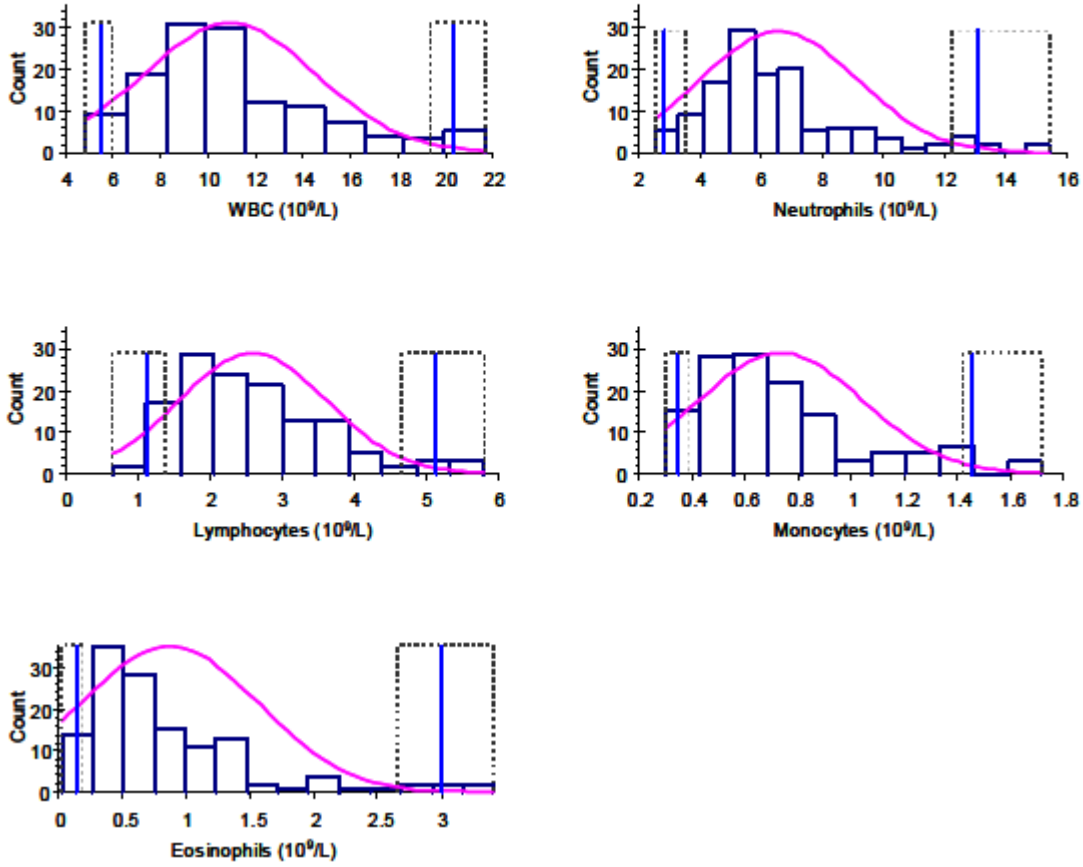


Figure 2 (continued-2): Observed (blue boxes) and fitted (purple line) distributions of blood hematology analytes in 132 healthy dogs. Blue vertical lines are the limits of the reference interval (with corresponding 90% confidence interval as dotted lines).

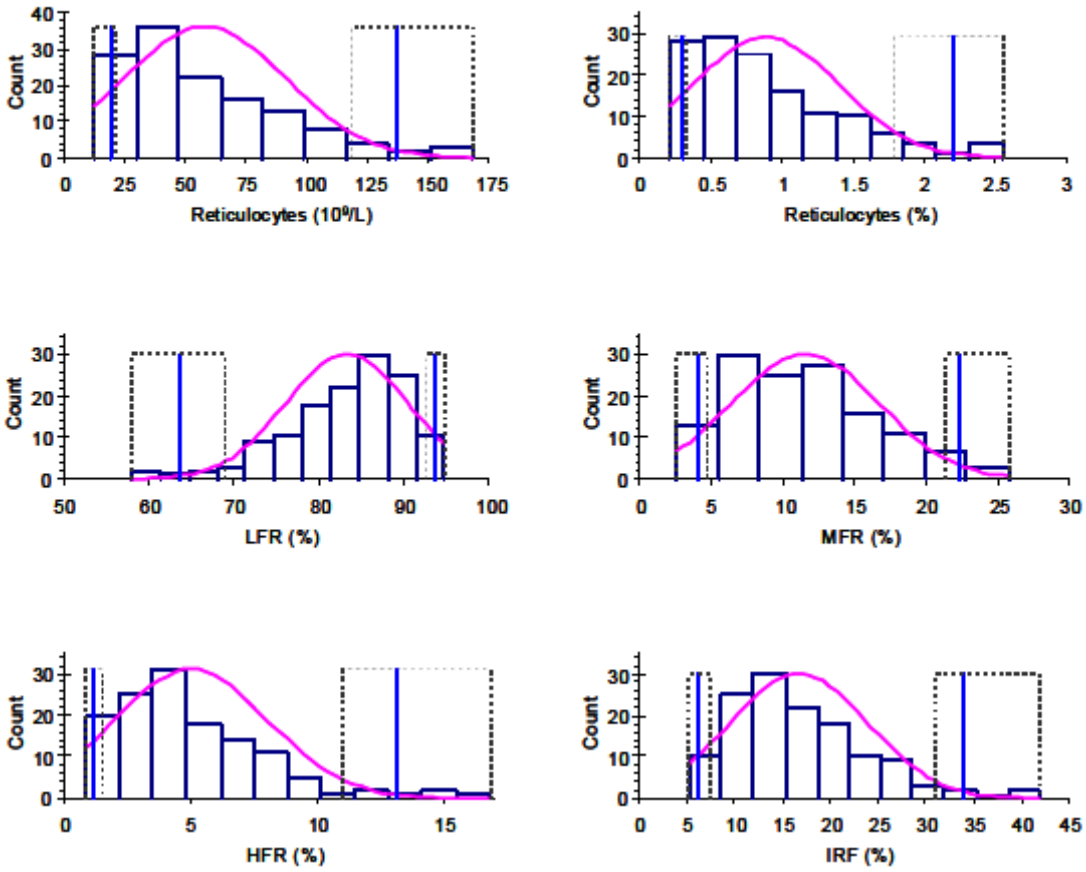
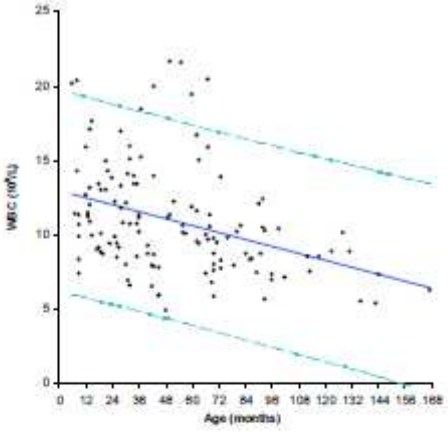
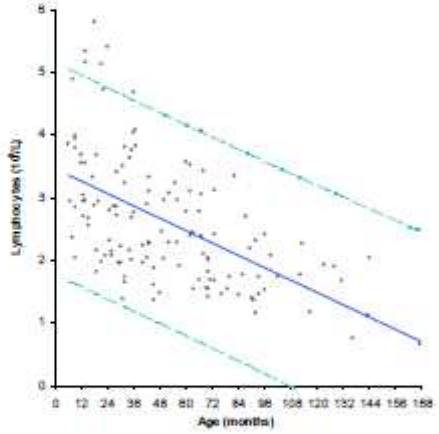
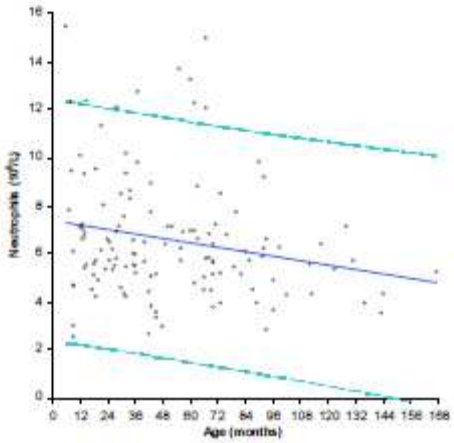


Figure 3: Effect of age on canine reference values for WBC (n=131), lymphocytes (n=130) and neutrophils (n=132). Scatter plots with polynomial fits (solid line) and 95% prediction interval (dotted line); same outliers as in Table 4.







References

1. CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; Approved guideline. Third edition. Clinical and Laboratory Standards Institute (CLSI) document C28-A3. Wayne, Pennsylvania, 2008.
2. Moritz A, Fickenscher Y, Meyer K, et al. Canine and feline hematology reference values for the ADVIA 120 hematology system. *Vet Clin Pathol*. 2004;33:32-38.
3. Geffré A, Friedrichs K, Harr K, et al. Reference values: a review. *Vet Clin Pathol*. 2009;38:288-298.
4. Lilliehook I, Tvedten H. Validation of the Sysmex XT-2000iV hematology system for dogs, cats, and horses. I. Erythrocytes, platelets, and total leukocyte counts. *Vet Clin Pathol*. 2009;38:163-174.
5. Lilliehook I, Tvedten H. Validation of the Sysmex XT-2000iV hematology system for dogs, cats, and horses. II. Differential leukocyte counts. *Vet Clin Pathol*. 2009;38:175-182.
6. Mathers RA, Evans GO, Bleby J, et al. Evaluation of the Sysmex XT2000iV haematology analyser for rat, dog and mouse whole blood samples. *Comp Clin Pathol*. 2008;17:137-144.
7. Société Centrale Canine. Confirmations des inscriptions au Livre des Origines Français. Available at:http://www.scc.asso.fr/mediatheque/statistiques/2006/Conft.pdf. Accessed December 2007.
8. Jain NC. Examination of the blood and bone marrow. In: Jain NC, ed. *Essentials of Veterinary Hematology*. Philadelphia: Lea & Febiger, 1993;1-18.
9. CLSI. User verification of performance for precision and trueness; Approved guide line. Clinical and Laboratory Standards Institute (CLSI) document EP15-A2. Second edition. ed. Wayne, Pennsylvania, 2005.
10. Norman EJ, Barron RC, Nash AS, et al. Evaluation of a citrate-based anticoagulant with platelet inhibitory activity for feline blood cell Counts. *Vet Clin Pathol*. 2001;30:124-132.
11. Geffré A, Concordet D, Braun JP, et al. Reference Value Advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel. *Vet Clin Pathol*. *[In press]*.
12. Reference Value Advisor macroinstructions. Available at: http://www.biostat.envt.fr/spip/spip.php?article63.
13. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem*. 1990;36:265-270.
14. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44:622-631.
15. Virtanen A, Kairisto V, Uusipaikka E. Regression-based reference limits: determination of sufficient sample size. *Clin Chem*. 1998;44:2353-2358.
16. Jensen AL, Kjelgaard-Hansen M. Method comparison in the clinical laboratory. *Vet Clin Pathol*. 2006;35:276-286.
17. NCCLS. Method comparison and bias estimation using patient samples; Approved guideline. NCCLS document EP9-A2. Wayne, Pennsylvania, 2002.
18. Jensen AL, Wenck A, Koch J, et al. Comparison of results of haematological and clinical chemical analyses of blood samples obtained from the cephalic and external jugular veins in dogs. *Res Vet Sci*. 1994;56:24-29.
19. Osselaer JC, Jamart J, Scheiff JM. Platelet distribution width for differential diagnosis of thrombocytosis. *Clin Chem*. 1997;43:1072-1076.
20. Singh H, Chaudhary R, Ray V. Platelet indices as quality markers of platelet concentrates during storage. *Clin Lab Haematol*. 2003;25:307-310.

21. Wiwanitkit V. Plateletcrit, mean platelet volume, platelet distribution width: its expected values and correlation with parallel red blood cell parameters. *Clin Appl Thromb Hemost.* 2004;10:175-178.
22. Pankraz A, Bauer N, Moritz A. Comparison of flow cytometry with the Sysmex XT2000iV automated analyzer for the detection of reticulated platelets in dogs. *Vet Clin Pathol.* 2009;38:30-38.
23. Yilmaz Z, Eralp O, Ilcol YO. Evaluation of platelet count and its association with plateletcrit, mean platelet volume, and platelet size distribution width in a canine model of endotoxemia. *Vet Clin Pathol.* 2008;37:159-163.
24. Bertazzolo W, Comazzi S, Sesso L, et al. Comparison of methods for determining platelet numbers and volume in Cavalier King Charles spaniels. *J Small Anim Pract.* 2007;48:556-561.
25. Stockham SL, Scott MA. Platelets. In: Stockham SL, Scott MA, eds. *Fundamentals of Veterinary Clinical Pathology*. Second ed. Blackwell publishing, Ames, Iowa, 2008: 223-257.
26. Tvedten H, Lilliehook I, Hillstrom A, et al. Plateletcrit is superior to platelet count for assessing platelet status in Cavalier King Charles Spaniels. *Vet Clin Pathol.* 2008;37:266-271.
27. Andersen AC, Gee W. Normal blood values in the beagle. *Vet Med.* 1958;53:135-156.
28. Bulgin MS, Munn SL, Gee W. Hematologic changes to 4 and one-half years of age in clinically normal beagles. *J Am Vet Med Assoc.* 1970;157:1064-1070.
29. Ewing GO, Schalm OW, Smith RS. Hematologic values of normal Basenji dogs. *J Am Vet Med Assoc.* 1972;161:1661-1664.
30. Geffré A, Braun JP, Trumel C, et al. Estimation of reference intervals from small samples: an example using canine plasma creatinine. *Vet Clin Pathol.* 2009; 38:477-484.
31. Schalm OW. Normal values in blood morphology. In: Schalm OW, ed. *Veterinary Hematology*. Philadelphia: Lea & Febiger, 1961:132-204.
32. Schalm OW. Normal Values in blood morphology. In: Schalm OW, ed. *Veterinary Hematology*. Second ed: Philadelphia: Lea & Febiger, 1965:190-292.
33. Schalm OW, Jain NC, Carroll EJ. Normal values in blood morphology with comments on species characteristics in response to disease. In: Schalm OW, Jain NC, Carroll EJ, eds. *Veterinary Hematology*. Third ed. Philadelphia: Lea & Febiger, 1975:82-218.
34. Jain NC. Comparative hematology of common domestic animals In: Jain NC, ed. *Essentials of Veterinary Hematology*. Philadelphia: Lea & Febiger, 1993:19 -53.
35. Rizzi TE, Meinkoth JH, Clinkenbeard KD Normal hematology of the dog. In: Weiss DJ, Wardrop KJ, eds. *Schalm's Veterinary Hematology*. Sixth ed. Ames, Iowa: Blackwell publishing Ltd, 2010:799-810.
36. Schalm OW. Blood volume and water balance In: Schalm OW, ed. *Veterinary Hematology*. Philadelphia: Lea & Febiger, 1961:15-28.
37. Tvedten H. Reference values In: Willard MD, Tvedten H, eds. *Small Animal Clinical Diagnosis by Laboratory Methods* Fourth ed. Elsevier, Missouri; 2004:417-419.
38. Tvedten H, Lilliehook I. Differential leukocyte counting: evaluation of four methods [abstract]. *Vet Pathol.* 2009;46:1026.
39. Hawthorne AJ, Booles D, Nugent PA, et al. Body-weight changes during growth in puppies of different breeds. *J Nutr.* 2004:134:2027S-2030S.
40. Metzger FL. Senior and geriatric care programs for veterinarians. *Vet Clin North Am Small Anim Prac.* 2005;35:743-753.
41. Shifrine M, Munn SL, Rosenblatt LS, et al. Hematologic changes to 60 days of age in clinically normal beagles. *Lab Anim Sci.* 1973;23:894-898.

42. Strasser A, Niedermuller H, Hofecker G, et al. The effect of aging on laboratory values in dogs. *Zentralbl Veterinärmed A.* 1993;40:720-730.
43. Harper EJ, Hackett RM, Wilkinson J, et al. Age-related variations in hematologic and plasma biochemical test results in Beagles and Labrador Retrievers. *J Am Vet Med Assoc.* 2003;223:1436-1442.
44. Michaelson SM, Scheer K, Gilt S. The blood of the normal beagle. *J Am Vet Med Assoc.* 1966;148:532-534.
45. Lilliehook I, Tvedten H. Investigation of hypereosinophilia and potential treatments. *Vet Clin North Am Small Anim Pract.* 2003;33:1359-1378.
46. Jain NC. The dog: normal hematology with comments on response to disease. In: Jain NC, ed. *Schalm's Veterinary Hematology*. Fourth ed. Philadelphia: Lea & Febiger; 1986:103-125.

Partie 3

Tester de nouvelles méthodes

Les difficultés rencontrées dans l'utilisation de la procédure de base ont été reconnues par les experts de l'IFCC et du CLSI qui ont indiqué que «les laboratoires individuels devraient s'attacher en priorité à vérifier des intervalles de référence déterminés ailleurs » [individual laboratories should focus more on verifying reference intervals established elsewhere], ce qui constitue les deux premières des quatre nouvelles pistes explorées :

1/- le transfert d'un intervalle préalablement déterminé, en se fondant sur une comparaison classique de techniques de mesure [22] ; cela n'est applicable que lorsque la démographie de la population du laboratoire « receveur » est similaire à celle du laboratoire « donneur ». Ces contraintes sont déjà couramment prises en compte par les laboratoires, notamment lors d'un changement de technique et/ou d'équipement. Nous avons appliqué cette procédure lors de détermination des intervalles de référence en hémostase canine avec deux analyseurs de deux laboratoires de la même zone géographique (Article n°2).

2/- la validation d'un intervalle de référence préalablement publié par un laboratoire « donneur » vers un laboratoire « receveur » en testant uniquement une série de 20 individus de référence. Si au moins 18 des valeurs de référence obtenues dans le laboratoire « receveur » concordent avec l'intervalle de référence du laboratoire « donneur », alors l'intervalle est validé. Cette procédure a été testée (Article n°4) pour des analyses d'hémostase, dans des conditions pré-analytiques et analytiques similaires mais non identiques au sein des deux laboratoires. La procédure, simulée à partir d'une banque de données (Article n°2), s'est avérée très utile pour rejeter rapidement l'intervalle de référence d'un laboratoire « donneur ». L'acceptation d'un intervalle de

référence de laboratoire « donneur » est en revanche plus délicate et parfois trompeuse, pouvant conduire à des erreurs diagnostiques (résultat du patient classé comme « normal » *vs* « anormal ») si les limites de référence sont utilisées comme seuils décisionnels.

3/- l'utilisation d'un nombre plus limité d'individus de référence, en compensant ce manque d'effectif par des outils statistiques puissants comme la méthode robuste. Cette autre piste est proposée par le CLSI, qui dans ses documents préliminaires ne donnait pas de limite inférieure pour l'effectif [23]. Une simulation de cette procédure a été testée pour des analyses biochimiques (Article n°5). Il est apparu que pour des échantillons de référence de très petite taille (27 individus), les intervalles de référence générés étaient très variables, correspondant parfois à celui obtenu à partir des 1439 individus de la base de données d'origine, mais étant aussi parfois totalement erronés. La sélection d'un faible nombre d'individus de référence peut créer une incertitude d'échantillonnage qu'il est malheureusement impossible de vérifier. Cependant, dans le cas particulier d'espèces sauvages ou rares [24], il peut arriver que l'on ne dispose que d'un nombre très faible de sujets ; l'application de la méthode robuste apparaît donc comme un palliatif, en espérant que la base de données pourra être ultérieurement complétée pour permettre une détermination plus précise des limites de référence.

4/- l'exploitation *a posteriori* des bases de données collectées dans les centres de médecine préventive, les hôpitaux ou les gros laboratoires de biologie médicale. Cette approche repose sur le fait que ces jeux de données contiennent à la fois les valeurs des individus considérés comme « sains », c'est-à-dire « non malades » pour le test en question, et les valeurs des individus « malades ». Il est possible, grâce à un modèle mathématique, de séparer les distributions des données des individus sains de celles des individus malades. Une fois extraite, la distribution des valeurs des individus sains est considérée comme la distribution de référence et les intervalles de référence en sont facilement déduits. Les simulations d'application à des exemples concrets de biologie humaine et animale ont donné de bonnes estimations des intervalles de référence déterminés selon les recommandations (Article n°6). Cependant, l'algorithme ayant été créé pour détecter des hétérogénéités de distribution, donc d'éventuelles sous-populations, il faut rester vigilant sur la nature de ces dernières : s'agit-il vraiment d'individus sains *vs* malades ou bien jeunes *vs* âgés ou encore femmes *vs* hommes ? De

plus, l'analyse de ces grosses bases de données peut parfois révéler plus de deux sous-populations. Ces perspectives de travail devront être explorées à l'avenir, mais il est clair que l'exploitation de ces séries de données nécessite de disposer d'un minimum d'informations démographiques pour éviter les éventuelles erreurs d'identification des sous-populations caractérisées par l'analyse mathématique.

Article n°4

Geffré A, Concordet D, Trumel C, Braun JP.

Validation of preexisting reference intervals: can the procedure be applied to canine hemostasis?

Journal of Veterinary Diagnostic Investigation, 2011;23:343-7.

Article n°5

Geffré A, Braun JP, Trumel C, Concordet D.

Estimation of reference intervals from small samples: an example using canine plasma creatinine.

Veterinary Clinical Pathology 2009;38:477-84.

Article n°6

Concordet D, Geffré A, Braun JP, Trumel C.

A new approach for the determination of reference intervals from hospital-based data.

Clinica Chimica Acta 2009;405:43-8.

J Vet Diagn Invest 23:343–347 (2011)

Validation of preexisting reference intervals: can the procedure be applied to canine hemostasis?

Anne Geffré,[1] Didier Concordet, Catherine Trumel, Jean-Pierre Braun

Abstract. The de novo establishment of reference intervals (RIs) for all variables is beyond the capabilities of many small laboratories. Thus, recent international recommendations propose procedures to adopt RIs established by "donor" laboratories after validation in "receiving" laboratories. The objective of the current study was to use recently published RIs of canine hemostasis tests as possible donor values and evaluate the validation procedure with randomized sets of values obtained in another study of canine RI determination of prothrombin time (PT), activated partial thromboplastin time (APTT), fibrinogen, and antithrombin (AT). The preanalytical, analytical, and demographic conditions of the donor and receiving laboratories were first compared. To represent new reference individuals, 25 validation sample sets of 20 results of the receiving laboratory were randomly selected for each variable and compared with the RI of the donor laboratory. Validation was rejected in all cases for APTT and AT. Donor RI could be validated in 14 of 25 cases for fibrinogen and in 4 of 25 cases for PT. When preanalytical and analytical differences existed between donor and receiving laboratories, validation procedures consistently rejected preexisting RI. When the differences are smaller, the variability of the results obtained in the validation sample sets tested may be responsible for validations or rejections, which can lead to further misinterpretations of results from patients. Validation of a preexisting reference interval is certainly an interesting option for small laboratories, but progressive determination of the laboratory's own reference interval is probably a better long-term solution.

Key words: Dogs; hemostasis; reference interval; validation procedure.

The International Federation of Clinical Chemistry and Laboratory Medicine–Clinical and Laboratory Standards Institute (IFCC-CLSI) recommends that each laboratory should establish its own reference intervals to ensure that the values correctly reflect the analytical characteristics of the methods used and the demographic characteristics of the individuals tested. However, the de novo determination of a reference interval is a long, difficult, and expensive process, and beyond the capabilities of most laboratories. Therefore, the latest issue[2] of the IFCC-CLSI recommendations proposes a procedure to validate preexisting, properly determined reference intervals: "every laboratory is more than capable of verifying the applicability of reference intervals in its own population."

The proposed validation procedure can be based on "subjective [...] judgment of the laboratorian" or, preferably, on a procedure "using small numbers of reference individuals."[2] This latter procedure consists of measuring the variable in specimens from 20 reference individuals from the receiving laboratory. After eliminating possible outliers, if all but 2 values are within the reference limits, the reference interval can be validated. If 3–4 values are outside the reference interval, another validation sample set of 20 reference individuals must be tested; if all but 2 values in this new validation sample set are within the reference limits, then the reference interval can be validated. When more than 4 values are outside the reference limits, the reference interval must be determined de novo according to the recommendations.[2,4]

To the authors' knowledge, the validation procedure has only been used once in veterinary clinical pathology for the determination and/or validation of reference intervals in Bernese Mountain dogs.[6] However, effects of interindividual variability within the small reference sample group used in the validation procedure have not been evaluated. The current study was therefore designed to test the soundness of transferring hemostasis reference intervals, and more specifically, the effects of the composition of the small reference sample group used by the receiving laboratory. It is based on previously published reference intervals of canine hemostasis in 56 dogs[1] (used herein as donor laboratory data) and individual results obtained in a previous study[5] (used herein to represent the procedure in the receiving laboratory). The variables studied were antithrombin (AT), prothrombin time (PT), activated partial thromboplastin time (APTT), and fibrinogen.

The subjective validation was based on a comparison of the preanalytical, analytical, and demographic conditions of the 2 laboratories as summarized from the 2 publications in Table 1. The 139 results obtained for each variable in the receiving laboratory were randomized 25 times using the RAND function of Excel.[a] The first 20 results obtained from each series were chosen to represent the small reference sample group hereafter called the validation sample set. The 25 validation sample sets were examined for possible outliers according to Tukey test at the >3

From the Département des Sciences Cliniques (Geffré, Trumel, Braun) and UMR 181 Physiopathologie et Toxicologie Expérimentales (Concordet, Braun), ENVT, INRA, Ecole Nationale Vétérinaire de Toulouse, Toulouse, France.

[1] Corresponding Author: Anne Geffré, Ecole Nationale Vétérinaire de Toulouse, Département des Sciences Cliniques, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France. a.geffre@envt.fr

Table 1. Comparison of the donor and receiving laboratory conditions for the test of validation of canine hemostasis reference intervals.

	Donor laboratory[1] (STA Compact analyzer)		Receiving laboratory[5] (STA Satellite analyzer)	
Prothrombin time				
Reagent	STA Neoplastin plus		STA-Neoplastine CI Plus	
Reference interval	5.7–8.1 sec		6.9–8.8 sec	
Confidence interval	Not reported		(6.8/7.0)	(8.6/9.9)
Distribution	Non-normal		Non-normal	
Activated partial thromboplastin time				
Reagent	STA APTT Kaolin		STA-Cephascreen	
Reference interval	10.0–14.3 sec		13.1–17.2 sec	
Confidence interval	(9.7/10.4)	(13.9/14.9)	(12.8–13.5)	(16.8–17.4)
Distribution	Log-normal		Normal after Box-Cox	
Fibrinogen				
Reagent	STA Thrombin		STA-Fibrinogen	
Reference interval	1.3–3.1 g/l		1.24–4.30 g/l	
Confidence interval	(1.2–1.4)	(2.9–3.4)	(1.09–1.43)	(3.85–5.18)
Distribution	Normal		Normal after Box-Cox	
Antithrombin				
Reagent	STA Antithrombin III		STA-Stachrom AT III	
Reference interval	107.9–128.0%		104–188%	
Confidence interval	Not reported		(96–110)	(178–192)
Distribution	Non-normal		Normal	
Demography	n = 56; 1–6 years old; ~55% males–45% females; a variety of breeds		Sets of 20 values (of 139); 0.5–14 years old; ~35% males–65% females; a variety of breeds	
Preanalytics	Fasted, resting dogs; Na_3-citrate, 3.18%; 2 × 10 min, 850 × g centrifugation; plasma stored –80°C ≤3 weeks		Fasted dogs; Na_3-citrate, 3.8%; 1 × 15 min, 1,300 × g centrifugation; no storage, analysis ≤7.5 hr	

interquartile range criterion. When an outlier was thus detected, the next value in the series was used as a substitute, as stated in the recommendations.[2] Results were then ordered and compared with the donor reference limits to be validated, based on the following 3 criteria: 1) if 2 or less values were outside the limits, the reference interval was considered validated; 2) if 5 or more values were outside the limits, the procedure was stopped and the validation was considered impossible; and 3) if 3 or 4 values were outside the limits, the next 20 results in the corresponding series of randomized values were tested as previously, to mimic a new validation sample set of 20 reference individuals.

The reference limits of the donor laboratory with their 90% confidence intervals (when available) are indicated on Table 1, which shows that the different equipment and reagents were from the same manufacturer. Preanalytical conditions also differed, but in both cases the stability of the specimens had been validated.[1,5] Comparison of demographic conditions showed that a variety of breeds was used in both studies, and that there was a larger age range and a higher proportion of females in the receiving laboratory.

When evaluating the validation sample sets, there was no significant effect between draws when evaluating values of APTT, fibrinogen, and AT (analysis of variance [ANOVA], $P > 0.05$), but a significant effect was observed for PT (ANOVA, $P = 0.035$). The dispersion of values in the validation sample sets also differed according to the draw with coefficients of variation (CV) of 4.5–9.9% for PT, 5.5–9.1% for APTT, 19.4–45.7% for fibrinogen, and 8.9–18.6% for AT.

Detailed results obtained for the first 20 fibrinogen values from the 25 validation sample sets are presented in Figure 1. An outlier was identified in 2 validation sample sets and was replaced by the next value in the randomized nonordered series of values (the outlier in validation sample set number 7 was replaced by 1.38 g/l, and the outlier in validation sample set number 17 by 1.42 g/l). According to the reference limits of the donor laboratory, 10 validation sample sets had values of 2 or less outside the limits tested; 6 validation sample sets had values of 5 or more outside the limits tested; and 9 validation sample sets had 3 or 4 values outside the limits tested.

In a second step, the following 20 nonordered results of the 9 latter corresponding series were examined (Fig. 2). After this replacement, there were 4 validation sample sets, in which values of 2 or less were within the donor laboratory reference interval and 5 in which values more than 2 were outside. Finally, in this test of the IFCC-CLSI validation procedure for the fibrinogen reference interval, there were 14 of 25 cases in which the criteria for validation were met and 11 cases in which the criteria were not met.

For APTT and AT (Fig. 3), there were 5 or more values outside the reference limits of the donor laboratory in each validation sample set. For PT, there were 4 validation sample sets, where 2 or less values were outside the limits, 19 where there were 5 or more values outside the limits, and 2 where there were only 4. For the latter, there were 3 or 4

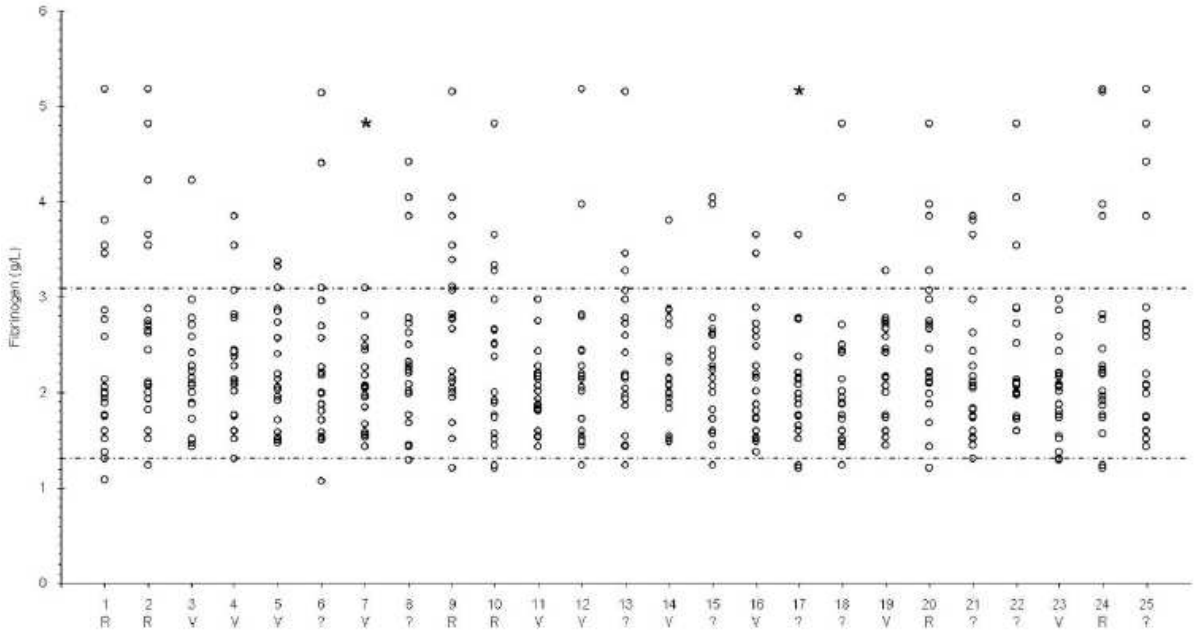


Figure 1. Test of the validation of a previously published reference interval for canine plasma fibrinogen concentration (dotted lines). Each vertical set of dots represents 20 randomly selected results obtained in a study of canine hemostasis reference intervals. * = outliers according to Tukey >3 IQR criterion; R = rejected; V = validated; ? = needs further investigation, according to International Federation of Clinical Chemistry and Laboratory Medicine–Clinical and Laboratory Standards Institute recommendations.

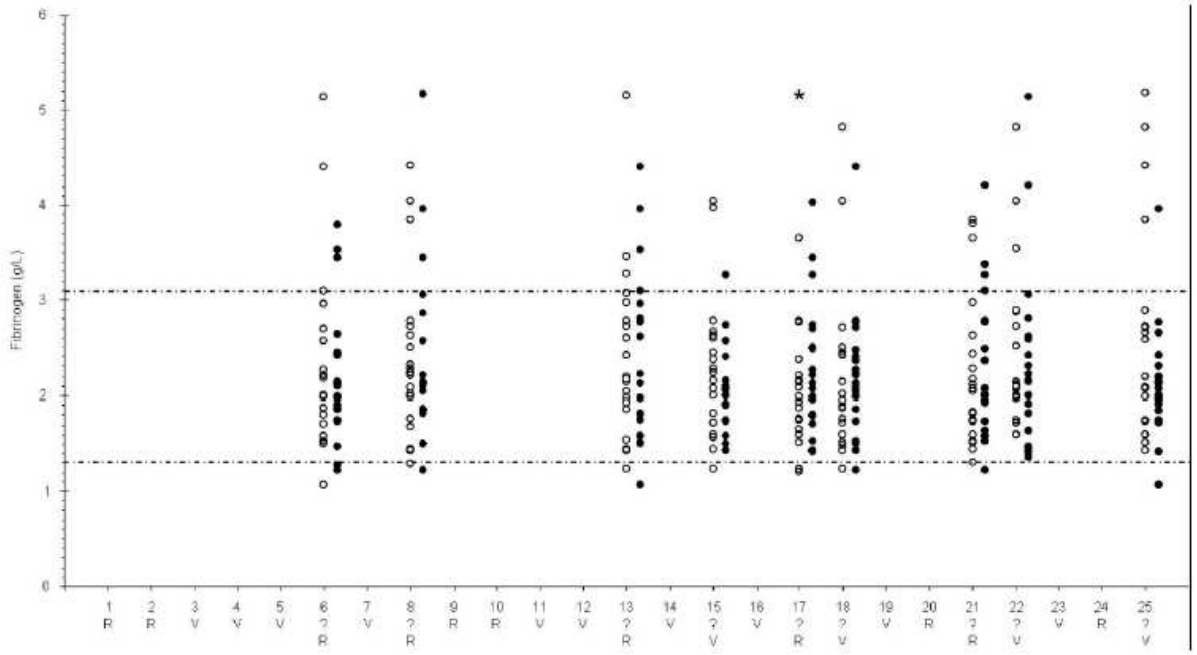


Figure 2. Results obtained in the second step of the test of validation of canine plasma fibrinogen reference intervals. Each black vertical set of dots represents the second series of 20 randomly selected results obtained in a study of canine hemostasis reference intervals. * = outliers according to Tukey >3 IQR criterion; R = rejected; V = validated; ? = needs further investigation, according to International Federation of Clinical Chemistry and Laboratory Medicine–Clinical and Laboratory Standards Institute recommendations.

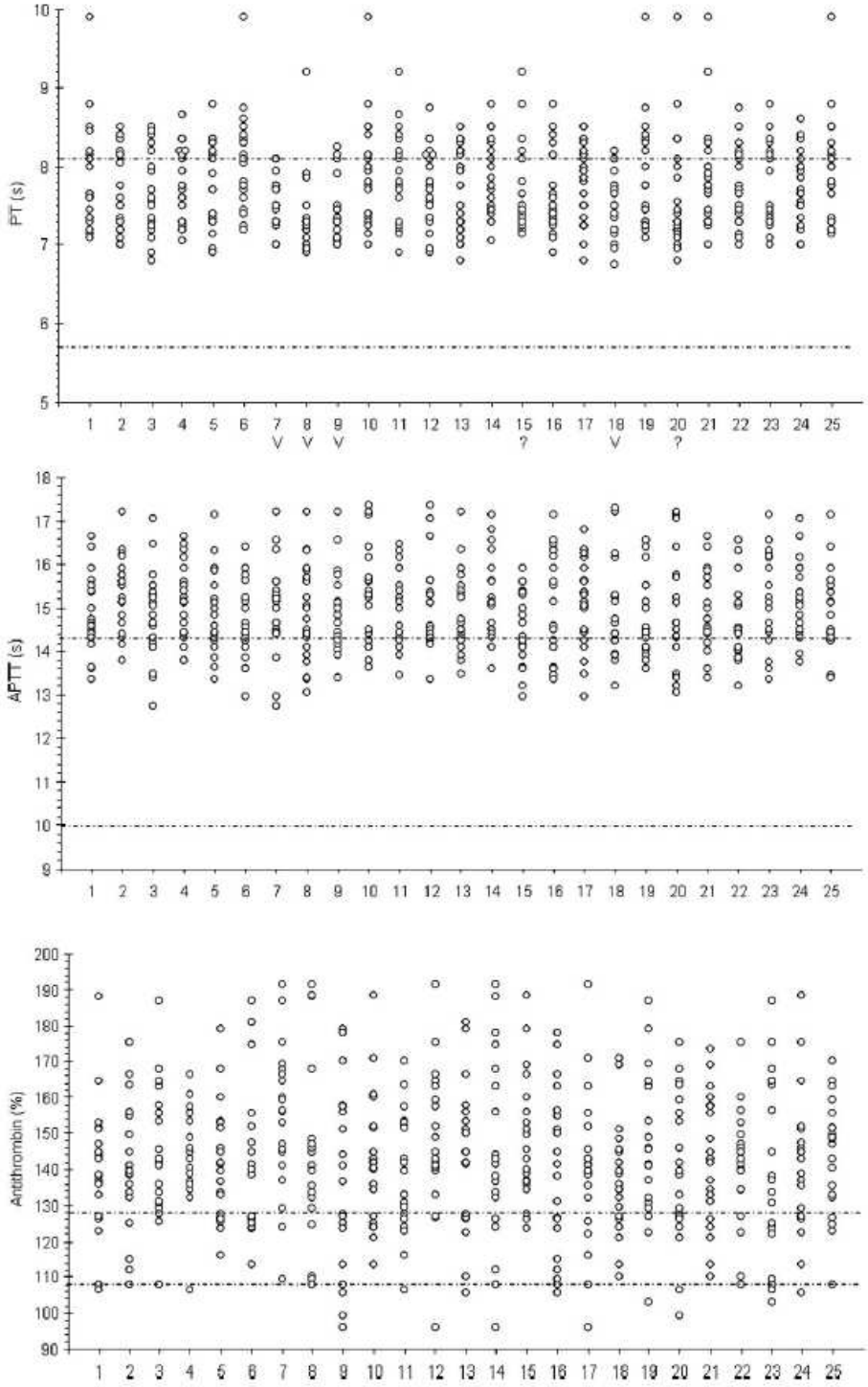


Figure 3. Test of the validation of previously published reference intervals for canine plasma prothrombin time (PT), activated partial thromboplastin time (APTT), and antithrombin (dotted lines). Each vertical set of dots represents 20 randomly selected results obtained in a study of canine hemostasis reference intervals. * = outliers according to Tukey >3 IQR criterion; V = validated; ? = needs further investigation, according to International Federation of Clinical Chemistry and Laboratory Medicine–Clinical and Laboratory Standards Institute recommendations.

values outside the limits in the next validation sample sets of 20 values (results not shown).

The validation of preexisting reference intervals by receiving laboratories is a very interesting option for all small- or medium-sized laboratories, as this may avoid the long, difficult, and expensive selection of a large number of well-characterized reference individuals for the establishment of de novo reference interval. In the current study, hemostasis tests were chosen because of the availability of the data and also because they are reported to be strongly instrument- and reagent-specific.[7] The first option for validation is "a subjective assessment" based on "careful inspection of the pertinent factors of the original appropriate reference value study."[2] This step is necessary but cannot be sufficient. Most often, the full details of the demographic variables are not extensively reported, and the complete set of reference values is not available. However, this first step should be used to exclude possible validation, when conditions are too different in the receiving laboratory or are not reported at all.

In the present case, the study used as basis for the donor laboratory was recent, and strictly based on IFCC-CLSI recommendations, except for the number of reference individuals ($n = 56$). Demographic conditions did not seem to differ notably. The analyzers used belonged to the same manufacturer and were based on the same technology, but the reagents were different. It was the subjective opinion of the investigators that the proofs of identity were not reliable enough to transfer the reference interval without a validation study.

In the test of validation, the random selection of 20-value validation sample sets represented the recommended procedure[2] to select 20 reference individuals representative of the demographics of the receiving laboratory. It was surprising to observe that there was such variability in the validation sample sets of randomly selected values. In the example of fibrinogen, the interindividual variability has recently been reported to have a CV of 28%.[8] In the present study, interindividual fibrinogen variability in the validation sample sets ranged from 19.4% to 45.7%, but this did not result in statistically significant differences between the validation sample sets. However, this variability increased the possibility of obtaining values outside a preexisting reference interval. Validation was estimated to be acceptable in slightly more than 1 of 2 cases (56%) by use of the small reference sample groups, which is not really better than coin tossing. The reference interval that had been established from the whole sample group of 139 dogs (Table 1) has a notably higher upper limit than the corresponding limit of the donor laboratory. As a consequence 8.6% of healthy dogs (12/139) would have been considered to have false-positive results by use of the upper limit of the donor laboratory (in healthy individuals, there are only 2.5% of false positives at this limit).

Less variability within the validation sample sets was observed for the other variables, as previously reported for PT and AT, whereas much higher between-dog variability (69.3%) had been reported for APTT.[8] Validation for APTT and AT was impossible in all cases. This is consistent with the notably different reference intervals of the 2 laboratories, and is likely due principally to the reagents used. For PT, validation would have been accepted in 4 of 25 cases, whereas the upper limit of the donor laboratory was 0.7 sec lower than in the current study,[5] which might cause clinical misinterpretations.

These results suggest that the validation of preexisting reference intervals can be an interesting option for a receiving laboratory, and that it seems especially valuable in demonstrating when a preexisting reference interval is inappropriate for a receiving laboratory. Results also suggest that there can be unexpected variability, and that users should be very cautious when granting validation with a first validation sample set of 20 values. This should entice laboratories wishing to use this procedure to progressively collect results obtained in reference animals to compute their own reference intervals, possibly using small reference sample groups with relevant statistical methods.[3] This is especially important when the analytical methods can notably impact the results, which is the case in hemostasis testing, or when reference intervals are only available from textbooks in which analytical and demographic details are rarely reported.

Sources and manufacturers

a. Microsoft Corp., Redmond, WA.

References

1. Bauer N, Eralp O, Moritz A: 2009, Reference intervals and method optimization for variables reflecting hypocoagulatory and hypercoagulatory states in dogs using the STA Compact automated analyzer. J Vet Diagn Invest 21:803–814.
2. Clinical and Laboratory Standards Institute (CLSI): 2008, Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline, 3rd ed. CLSI document C28-A3.. CLSI, Wayne, PA.
3. Geffré A, Concordet D, Braun JP, Trumel C: 2010, Reference Value Advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel. Vet Clin Pathol. In press.
4. Geffré A, Friedrichs K, Harr K, et al.: 2009, Reference values: a review. Vet Clin Pathol 38:288–298.
5. Geffré A, Grollier S, Hanot C, et al.: 2010, Canine reference intervals for coagulation markers using the STA Satellite and the STA-R analyzers. J Vet Diagn Invest 22:690–695.
6. Nielsen L, Kjelgaard-Hansen M, Jensen AL, Kristensen AT: 2009, Breed-specific variation of hematologic and biochemical analytes in healthy adult Bernese Mountain dogs. Vet Clin Pathol 39:20–28.
7. Stockham SS, Scott MA: 2008, Hemostasis. *In:* Fundamentals of veterinary clinical pathology, 2nd ed., pp. 259–321. Blackwell Publishing, Ames, IA.
8. Wiinberg B, Jensen AL, Kjelgaard-Hansen M, et al.: 2007, Study on biological variation of haemostatic parameters in clinically healthy dogs. Vet J 174:62–68.

Veterinary Clinical Pathology ISSN 0275-6382

ORIGINAL RESEARCH

Estimation of reference intervals from small samples: an example using canine plasma creatinine

A. Geffré[1], J.P. Braun[1,2], C. Trumel[1], D. Concordet[2,3]

[1]Department of Clinical Sciences, [2]UMR181 Physiopathologie and Toxicologie Expérimentales INRA, ENVT, and [3]Department of Biological Sciences, Ecole Nationale Vétérinaire, Toulouse, France

Key Words
Creatinine, dog, reference values, robust method, small sample

Correspondence
A. Geffré, Department of Clinical Sciences, Ecole Nationale Vétérinaire, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France
E-mail: a.geffre@envt.fr

DOI:10.1111/j.1939-165X.2009.00155.x

Background: According to international recommendations, reference intervals should be determined from at least 120 reference individuals, which often are impossible to achieve in veterinary clinical pathology, especially for wild animals. When only a small number of reference subjects is available, the possible bias cannot be known and the normality of the distribution cannot be evaluated. A comparison of reference intervals estimated by different methods could be helpful.
Objective: The purpose of this study was to compare reference limits determined from a large set of canine plasma creatinine reference values, and large subsets of this data, with estimates obtained from small samples selected randomly.
Methods: Twenty sets each of 120 and 27 samples were randomly selected from a set of 1439 plasma creatinine results obtained from healthy dogs in another study. Reference intervals for the whole sample and for the large samples were determined by a nonparametric method. The estimated reference limits for the small samples were minimum and maximum, mean ± 2 SD of native and Box–Cox-transformed values, 2.5th and 97.5th percentiles by a robust method on native and Box–Cox-transformed values, and estimates from diagrams of cumulative distribution functions.
Results: The whole sample had a heavily skewed distribution, which approached Gaussian after Box–Cox transformation. The reference limits estimated from small samples were highly variable. The closest estimates to the 1439-result reference interval for 27-result subsamples were obtained by both parametric and robust methods after Box–Cox transformation but were grossly erroneous in some cases.
Conclusion: For small samples, it is recommended that all values be reported graphically in a dot plot or histogram and that estimates of the reference limits be compared using different methods.

Introduction

The concept of using reference values for reporting the variability of analytes in healthy subjects is widely accepted as a basis for interpreting the individual values observed in patients, even though many medical classifications are based on decision limits or consensus values that differ from the reference limits.[1,2] The most recent international guidelines for the preparation of reference limits in human clinical pathology have been published by the International Federation of Clinical Chemistry (IFCC) and the Clinical and Laboratory Standards Institute (CLSI).[3] Most of these recommendations can be transposed to animal clinical pathology. However, as with some human subgroups (eg, newborns or elderly people), it is often impossible to obtain the minimum recommended number (120) of reference subjects. It is nevertheless recommended that in such cases "data should still be analyzed by the nonparametric method. As an alternative, the robust method may be used . . .."[3]

Robust methods are based on iterative processes that estimate the median and spread of the distribution.[4,5] In the IFCC-CLSI proposed guideline for robust

methods, they gave as an example the assessment of a reference interval for plasma calcium in women, in which 3 sets of 20 values each were randomly selected from a group of 120 values. The subsequent reference intervals calculated by the robust method were very close to the interval determined by a nonparametric method from the whole set of 120 values.[3] In that example, the distribution of 120 values was roughly Gaussian and the range was narrow (88–103 mg/L), with no outliers. This may explain why "the performance of the method (was) not dependent on getting a 'good' set of points; though of course, results vary depending on the specific values selected." Despite the good results obtained in that example, in the final approved revised guideline the working group was hesitant to recommend calculating reference intervals with sample numbers < 80 "except in the most extreme instances."[3]

In practice, when only a small number of reference subjects are available for selection, the possible bias resulting from this selection cannot be known, and the normality of the distribution cannot be evaluated. Thus some doubts will remain and a comparison of the reference intervals estimated by different methods could be helpful.

The aim of this study was to take a large set of canine plasma creatinine reference values obtained in a previous study[6] and to: (1) randomly select large samples (120-sample sets, which is the smallest number of subjects recommended for use of the nonparametric method) and determine reference intervals; (2) randomly select small samples (27-sample sets) and estimate the reference limits in each by different methods; and (3) compare the results obtained with the reference interval determined for the whole sample by the nonparametric method.

Materials and Methods

The whole sample consisted of 4097 dogs, of which 1439 were healthy animals of known body-weight class and plasma creatinine concentration. These healthy dogs consisted of 800 small (< 15 kg), 261 medium (15–35 kg), and 378 large (> 35 kg) dogs. No other possible factors of variation were taken into account in the study. Ten sets of 120 results each (large samples) were randomly selected from the whole sample using the ALEA function in Excel (Microsoft Corporation, Redmond, WA, USA). Ten additional large subsets consisting of 67 small, 22 medium, and 31 large dogs were selected to reflect the relative proportion of each body-weight class in the whole sample. Reference limits and their 90% confidence intervals (CIs) were determined for the whole population, for the random 120-sample subsets, and for the 120-sample subsets balanced for body weight, by the nonparametric method according to IFCC recommended procedure, which is based on the ranking of values and setting limits at the 0.025 and 0.975 fractiles (percentiles).[7]

Ten sets of 27 results each (small samples) were randomly obtained from the whole sample (R subgroup) by the same random selection method and 10 additional sets of 27 results were obtained by randomly selecting results from 15 small, 5 medium, and 7 large dogs, again to represent their relative proportion in the whole sample (S subgroup). In each of the 20 sets of 27 values, the following were calculated: (1) minimum–maximum interval (range); (2) mean $\pm$ 2 SD of native and Box–Cox-transformed values; (3) 2.5–97.5% intervals by a robust method with native and Box–Cox-transformed values; and (4) 2.5–97.5% intervals estimated graphically from the cumulative distribution functions (CDF) derived from the histograms.

Statistical analysis

Calculations were performed with an Excel spreadsheet (Microsoft Corporation) and the Analyse-It (Leeds, UK) set of macroinstructions. The Box–Cox λ coefficient was calculated with freeware R 2.7.0 (The R Foundation for Statistical Computing; http://stat.ethz.ch/CRAN/). The robust method used was the one recommended by Horn and colleagues.[3,4] Results were tested using ANOVA. Comparisons between groups was made using a Mann–Whitney *U*-test after testing the homogeneity of variances, and when necessary using Bonferroni's correction. Possible partitioning criteria were studied by Harris and Boyd's *z*-statistics,[8] and outliers were detected by visual inspection of distributions and confirmed by Tukey's criterion.[3]

Results

Reference limits of the whole reference sample ($n = 1439$)

The overall distribution was skewed and non-Gaussian (Figure 1) and could not be transformed into a Gaussian distribution by log or Box–Cox transformation (Anderson–Darling, $P = .0005$ and .001, respectively). The reference interval for plasma creatinine concentration determined nonparametrically (90% CI of limits in parentheses) was 53.1 (52.0–55.0) to 150.4 (148.0–159.0) µmol/L (Table 1). The effect of body weight was highly significant (ANOVA, $P < .001$) and

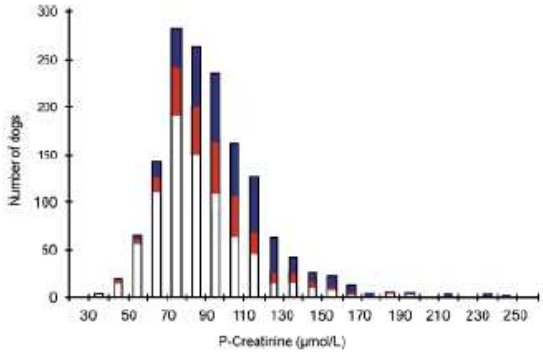


Figure 1. Distribution of plasma (P) creatinine concentration in 1439 clinically healthy dogs. White, body weight (BW) < 15 kg; red, BW 15–35 kg; blue, BW > 35 kg.

the 2 × 2 differences between the 3 body-weight classes were also significant (Mann–Whitney with Bonferroni correction, $P < .001$; Harris and Boyd $z \gg z^*$ after Box–Cox-transformation). The corresponding reference intervals for small, medium, and large dogs, respectively, were 51.0 (46.0–53.0) to 146.0 (140.0–153.0) µmol/L, 60.1 (53.1–62.0) to 143.9 (136.0–154.0) µmol/L, and 67.5 (61.9–70.0) to 168.6 (159.0–180.0) µmol/L.

Reference limits of the large samples

The distributions of all 120-sample subsets differed significantly from Gaussian (Anderson–Darling, $P < .05$) whereas none of the Box–Cox-transformed distributions, except 1 ($P = .014$), differed from Gaussian (Anderson–Darling, $P = .053$–$.977$). There was no difference between the limits determined by the non-parametric method in the randomly selected and body-weight class selected subgroups (Student's t-test after testing for homogeneity of variances, $P = .250$ and .829 for lower and upper limits, respectively). Lower and upper limits ranged from 44 to 57 (median 53.1) µmol/L and from 140 to 179 (median 150.7) µmol/L, respectively, and the 90% CIs ranged from 35 to 63 µmol/L and from 124 to 242 µmol/L for the lower and upper limits, respectively (Table 1 and Figure 2).

Reference limits of the small samples

The percentages of dogs in the 3 body-weight classes in the 10 randomly selected data subsets (R1–R10 subgroups) differed considerably from those in the whole sample (55.6%, 18.1%, 26.3%), whereas the percentages in the 10 S-subgroups, selected based on body-weight class, were almost the same (55.6%, 18.5%, 25.9%) as in the whole sample (Figure 3). Creatinine concentrations in the 20 subsets had similar ranges, with coefficients of variations (CVs) of 18.3–37.5% in the R subgroups and 18.2–39.4% in the S subgroups (Figure 4).

In most cases, the native values could not be used to estimate the reference limits by the robust method so the results are not reported. Extrapolation of the 2.5th and 97.5th percentiles from the histograms was imprecise due to the distribution of values (see 2 typical examples in Figure 5) and systematically gave values below the observed minimum and maximum (Wilcoxon's test, $P < .001$). Whatever the method used to calculate reference limits, no difference was found between the limits determined for the R and S subgroups, except for the mean ± 2 SD limit, which was

Table 1. Experimental approach and resulting reference limits for plasma creatinine concentration (µmol/L) in dogs.

				Reference Interval Results for Creatinine (µmol/L)	
Dataset	Method of Estimation	Data Type	Reference Interval Type	Lower Limit	Upper Limit
Whole sample (n = 1439)	Nonparametric (2.5–97.5%)	Native values	Limit (90% CI of limit)	53.1 (52–55)	150.4 (148–159)
Large subsets of 120 samples each	Nonparametric (2.5–97.5%)	Native values	Median (range) of limits	53.1 (44–57)	150.7 (140–179)
Small subsets of 27 samples each	Minimum–maximum	Native values	Median (range)	53.1 (40–75)	150.4 (125–239)
	Parametric (mean ± 2 SD)	Native values	Median (range)	42.0 (22–61)	146.1 (124–186)
		Box–Cox transformed	Median (range)	53.8 (40–74)	154.3 (126–283)*
	Robust	Native values	Median (range)	NC	NC
		Box–Cox transformed	Median (range)	52.4 (23–73)	165.8 (122–363)†
	Visual estimation	CDF	Median (range)	46.8 (30–70)	143.2 (120–223)

*Upper limit is an outlier; next nearest value is 203 µmol/L.
†Upper limit is an outlier; next nearest value is 215 µmol/L.
CDF, cumulative density distribution CI, confidence interval; NC, not calculated (not possible to calculate) for many subgroups.

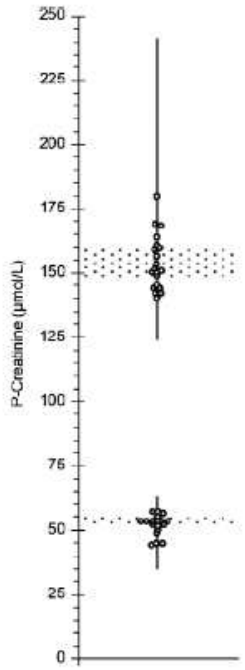


Figure 2. Upper and lower limits of the reference interval for plasma (P) creatinine determined by a nonparametric method in 20 randomly selected 120-sample subsets within the full dataset of 1439 healthy dogs. Dotted areas, 90% confidence intervals (CIs) of the limits determined for the whole sample; vertical bar, range of the 90% CI calculated from the 120-sample subsets.

lower in the S subgroups (Mann–Whitney, $P < .05$ after testing for homogeneity of variances).

Estimates of the lower limit of the reference interval obtained by the parametric and the robust method after Box–Cox transformation (Table 1 and Figure 6) were closer to the value determined in the whole dataset than those obtained by other methods. The range of upper limit estimates was wider than for the lower limit (Table 1 and Figure 7). Box–Cox-transformed results revealed an apparent outlier, which, according to Tukey's criterion, was eliminated. When this outlier was removed, the CV of the estimates was similar to that obtained for the lower limit. Closest estimates of the upper limit determined in the whole sample were obtained from the calculated mean ± 2 SD using native values and Box–Cox-transformed values.

The means of all estimates of the reference limits (50 and 158 µmol/L) were close to the values determined from the whole sample. The range of values was large, however, independent of the method used; the range was 39.8–73.6 µmol/L in the best case for the lower limit, and most values calculated from small

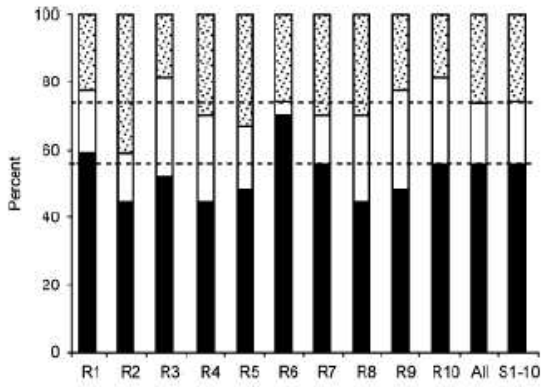


Figure 3. Percentage of small (black), medium (white), and large (speckled) dogs in the 10 sets of 27 results selected randomly (R1–R10) from the whole sample (All) and in the 10 sets of 27 results selected according to body-weight class (S1–10). Dotted lines are the percentages of dogs in each body-weight class in the whole sample.

samples were outside the CIs determined from the whole sample. Meaningful calculation of the CIs for the determined limits was precluded by the low number of individuals in the 20 small sets.

Discussion

The aim of this study was to determine whether reference limits calculated from small samples randomly selected from a large sample were identical to those calculated from the latter by nonparametric method. If so, more or less valid reference intervals could be estimated from small samples when large samples are not available. Our results, however, suggest the bias obtained from different random small samples resulted in

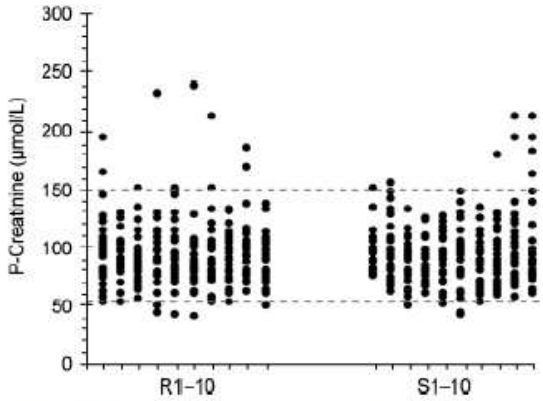


Figure 4. Distribution of plasma (P) creatinine values in the 10 sets of 27 results randomly selected (R1–R10) and in the 10 sets of 27 results selected according to body-weight class (S1 -10) from the whole sample. The dotted lines are the limits of the reference interval determined from the whole sample.

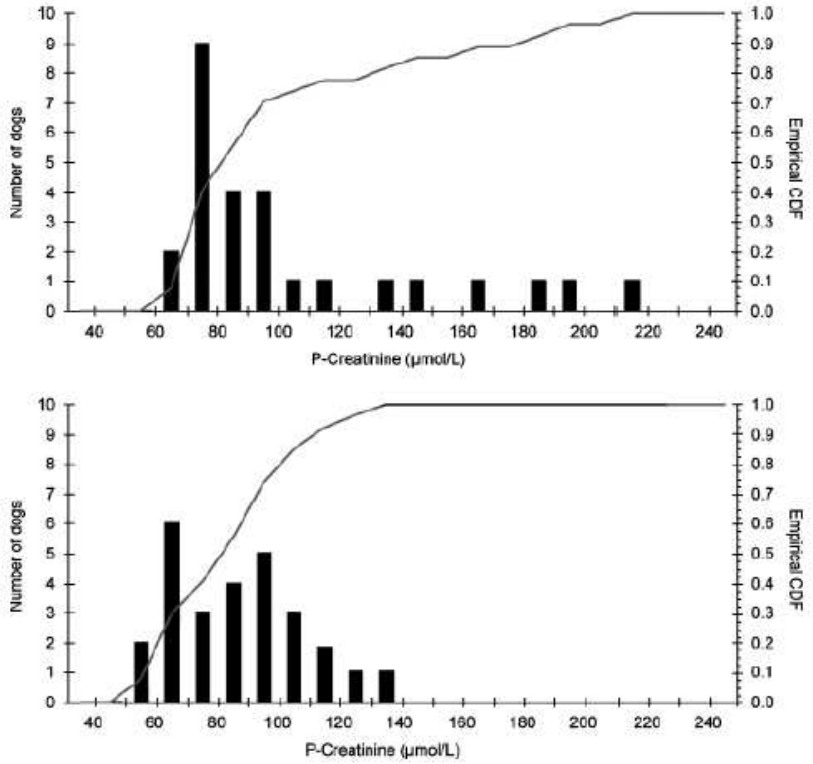


Figure 5. Examples of histograms (black bars) and cumulative distribution function (CDF; thin line) of plasma (P) creatinine values obtained from two 27-sample subsets of values.

very different estimates of reference limits compared with those obtained with large samples.

It was presumed that all recommended criteria concerning preanalytical and analytical criteria were respected.[3,9] These were not reported here as the aim was not to determine reference intervals for plasma creatinine concentration in dogs, but rather to compare different methods of estimating reference intervals from small samples. Even though some criteria may have been inadequate, their effect on the whole

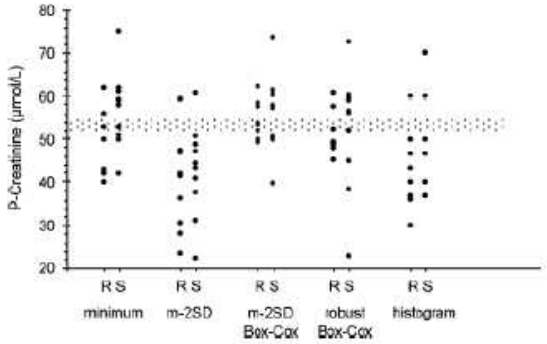


Figure 6. Estimates of the lower limits of the reference interval for plasma (P) creatinine as determined by different methods in the 10 sets of 27 results selected randomly (R) and in the 10 sets of 27 results selected according to body-weight class (S). The dotted area is the 90% confidence interval of the lower limit determined from the whole sample. m, mean; SD, standard deviation.

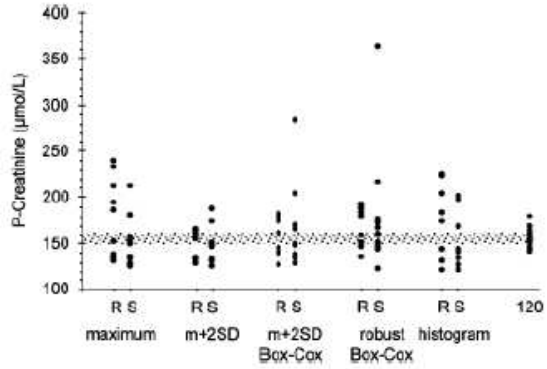


Figure 7. Estimates of the upper limit of the reference interval for plasma (P) creatinine as determined by different methods in the 10 sets of 27 results selected randomly (R) and in the 10 sets of 27 results selected according to body-weight class (S). The dotted area is the 90% confidence interval of the upper limit determined from the whole sample. m, mean; SD, standard deviation.

sample and the subgroups would have been identical and therefore validate the comparisons.

No apparent heterogeneity could be detected from the shape of the histogram of plasma creatinine concentration in the whole sample, except that it was heavily skewed toward high concentrations, which suggested that plasma creatinine concentrations might have been higher in 1 subgroup of dogs. Not all the possible partitioning factors evidenced in the preceding study were taken into account.[6] Body weight was the only partitioning factor used to investigate possible effects on results obtained from small samples, as this was reported previously to influence canine plasma creatinine concentration.[10,11] It is usually acknowledged that differences between subgroups may be statistically significant, when a large number of samples is compared, even though they may not be clinically relevant, which was the case in this study.[12] When applied to medical data, tests of partitioning such as that described by Harris and Boyd[8] are only suitable for comparisons of 2 sets of values. In the case of body weight, all 2×2 comparisons were highly significant. Thus, partitioning results according to the 3 body-weight classes was considered relevant, and validated the original selection of data subsets based on body-weight distribution.

The overall upper limit of the reference interval in this study was almost the same as the limit reported in the previous study in which all healthy animals were included (151 µmol/L; $n = 1516$ values).[6] In the present study, only those cases with known body-weight class were analyzed ($n = 1439$).

Although it was not the main aim, the variability of reference intervals determined using the recommended nonparametric method and the minimum number of reference individuals (120) was examined and compared with the variability of the estimates obtained from small numbers. The mean of the range of reference limits for the 20 random large samples selected from the whole sample and that of the reference limits determined from the whole sample, was almost identical. The variance of the limits thus determined was almost the same (they differed by ~7%) for the lower and upper limits. This was much lower than the variability of estimates from the small samples, but much broader than the 90% CIs determined for the whole sample. The narrowness of this latter was due to the large number of values used, uncommon in studies of reference values except those based on hospital data.[13] It was surprising, however, to see that in some cases the calculated limits derived from the minimum recommended number of 120 samples differed notably from those determined from the whole sample.

Twenty-seven was chosen as the number of samples for the small subsets in this study because many reports of reference values for nondomestic animal species include < 30 animals and because this number permitted a relative weighting of body-weight classes proportional to that of the whole sample.

In this study we confirmed that the bias obtained from different random samplings of 27 results resulted in very different estimates of reference limits but the results for the R and S subgroups, ie, with or without a partitioning factor (here, body-weight class) did not differ. Partitioning factors, based on a priori estimates of possible effects of sex, age, season, etc, on the results, are sometimes taken into account in studies with small numbers of animals. This approach may therefore not be appropriate if the number of animals in each category is too small to allow a proper study of differences between subgroups.

When only a small sample is available, recommendations state that all of the results may "serve a useful clinical purpose as a guide in the form of a list of all the values, [. . .] ordered according to increasing magnitude."[7] The major advantage of this type of data presentation is that no information is lost as all values are reported, but the list of numbers is not easy to evaluate. The values can also be reported in a dot plot, histogram, or diagram of CDF; these forms of data presentation may be more useful from a clinical standpoint. However, such lists or figures still are less easy to apply routinely than the upper and lower limits of a reference interval. When the number of samples is low, extrapolation of values from a CDF diagram resulted in 2.5% and 97.5% limits that were below the observed minimum and maximum of the dataset and therefore are not relevant. A nonparametric approach cannot be used as the number of reference samples is below 40.

Reference intervals from small samples are reported in the literature in many different ways, including median and maximum–minimum values, mean ± SD, and 2.5th and 97.5th percentiles estimated by parametric methods, with or without transformation of the data. However, it is impossible to correctly assess the type of distribution in small samples and, as shown in this study, the shapes of the distributions and the numbers of apparent outliers differed considerably between the 20 sets of values. Thus it is difficult to make relevant decisions on the best mathematical model to apply. However, as blood analyte distributions are often skewed, it would seem reasonable to include all the values obtained from small samples.

In this study, the lower reference limits were underestimated when the mean − 2 SD interval from

untransformed values was used on the small samples, which would be expected from a distribution skewed toward high values. It was surprising, however, that the mean+2 SD limits were close to the upper limit calculated from the whole sample. The parametric approach is more valid if a better fit to Gaussian distribution can be obtained by mathematical transformation. This can often be achieved by Box–Cox transformation, which is now readily available in the R 2.7.0 freeware program. The distribution of the Box–Cox-transformed values did not differ significantly from Gaussian except for 1 subset of data, and parametric estimation of the reference interval from the Box–Cox-transformed values was close to the CIs of the reference limits determined from the whole sample. In the one exception, the upper limits determined as the mean+2 SD from Box–Cox-transformed values or by the robust method were grossly erroneous (Figure 7) compared with other methods of estimation.

Robust methods have been recommended for determining the quantiles of a distribution of small samples. In this case, the method of Horn et al[4] could not be applied to untransformed canine plasma creatinine values. As already mentioned, the robust method should preferably be used when the data fit a Gaussian distribution, and in this study, the robust approach was efficient on Box–Cox-transformed values. In the example in which the robust method was used for human plasma calcium values, estimation of reference limits was more accurate than in this study. This may be due to the higher interindividual variance of canine plasma creatinine (~15%)[14,15] than of human plasma calcium (~3%),[16,17] such that randomly selected small subsets of canine creatinine data may not be representative of the whole set of values, resulting in the determination of erroneous limits.

In summary, none of the methods used in this study were very satisfactory for estimating reference intervals in small samples, and probably no method can be used that is generally applicable. Whatever calculations and mathematical models are applied, the limiting factor remains the a priori assumption that the small sample of values available is representative of samples to be tested in the future for diagnosis.[18] Reference intervals always should be estimated from the largest possible number of animals available so that nonparametric methods can be used to determine the reference limits and their CIs, and allow evaluation of possible partitioning factors. When only small samples are available, the estimation of reference intervals is biased by the sample, which may be more or less representative of the whole population, and this bias cannot be determined. In this case, as much information as possible should be reported, including lists of ordered values, dot plots, and/or histograms. A good approach when estimating reference intervals is to transform the data to obtain the best possible fit with Gaussian distribution and to compare the estimated limits obtained by parametric and robust methods. The real and estimated reference limits may differ considerably, demonstrating the bias inherent in reporting a single estimate for small samples.

References

1. Grasbeck R, Saris NE. Establishment and use of normal values. *Scand J Clin Lab Invest*. 1969;26(Suppl 110):62–63.
2. Petitclerc C. Normality: the unreachable star? *Clin Chem Lab Med*. 2004;42:698–701.
3. CLSI. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline*. 3rd ed. Wayne, PA: CLSI; 2008.
4. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44:622–631.
5. Horn PS, Pesce AJ, Copeland BE. Reference interval computation using robust vs parametric vs nonparametric analyses. *Clin Chem*. 1999;45: 2284–2285.
6. Concordet D, Vergez F, Trumel C, et al. A multicentric retrospective study of serum/plasma urea and creatinine concentrations in dogs using univariate and multivariate decision rules to evaluate diagnostic efficiency. *Vet Clin Pathol*. 2008;37:96–103.
7. 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.
8. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem*. 1990;36:265–270.
9. Solberg HE, Stamm D. International Federation of Clinical Chemistry (IFCC) IFCC recommendation. The theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. *Ann Biol Clin (Paris)*. 1991;49: 487–490.
10. Medaille C, Trumel C, Concordet D, et al. Comparison of plasma/serum urea and creatinine concentrations in the dog: a 5-year retrospective study in a commercial veterinary clinical pathology laboratory. *J Vet Med A Physiol Pathol Clin Med*. 2004;51:119–123.

11. Feeman WE III, Couto CG, Gray TL. Serum creatinine concentrations in retired racing Greyhounds. *Vet Clin Pathol*. 2003;32:40–42.
12. Lahti A. Partitioning biochemical reference data into subgroups: comparison of existing methods. *Clin Chem Lab Med*. 2004;42:725–733.
13. Solberg HE. Using a hospitalized population to establish reference intervals: pros and cons. *Clin Chem*. 1994;40:2205–2206.
14. Pagitz M, Frommlet F, Schwendenwein I. Evaluation of biological variance of cystatin C in comparison with other endogenous markers of glomerular filtration rate in healthy dogs. *J Vet Intern Med*. 2007;21:936–942.
15. Jensen AL, Aaes H. Critical differences of clinical chemical parameters in blood from dogs. *Res Vet Sci*. 1993;54:10–14.
16. Harris EK, DeMets DL. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. V. Estimated biological variations in ionized calcium. *Clin Chem*. 1971;17:983–987.
17. Lacher DA, Hughes JP, Carroll MD. Estimate of biological variation of laboratory analytes based on the third national health and nutrition examination survey. *Clin Chem*. 2005;51:450–452.
18. Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta*. 2003;334:5–23.

Clinica Chimica Acta 405 (2009) 43–48

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

A new approach for the determination of reference intervals from hospital-based data

D. Concordet [a,*], A. Geffré [b], J.P. Braun [a,b], C. Trumel [b]

[a] UMR181 Physiopathologie & Toxicologie Expérimentales INRA, ENVT, Ecole Nationale Vétérinaire, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France
[b] Department of Clinical Sciences, Ecole Nationale Vétérinaire, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France

ARTICLE INFO

Article history:
Received 23 February 2009
Received in revised form 27 March 2009
Accepted 31 March 2009
Available online 5 April 2009

Keywords:
Hospital-based reference interval
Indirect method
E.M. algorithm

ABSTRACT

Background: Reference limits are some of the most widely used tools in the medical decision process. Their determination is long, difficult, and expensive, mainly because of the need to select sufficient numbers of reference individuals according to well-defined criteria. Data from hospitalized patients are, in contrast, numerous and easily available. Even if all the information required for a direct reference interval computation is usually not available, these data contain information that can be exploited to derive at least rough reference intervals.
Methods: In this article, we propose a method for the indirect estimation of reference intervals. It relies on a statistical method which has become a gold-standard in other sciences to separate components of mixtures. It relies on some distributional assumptions that can be checked graphically. For the determination of reference intervals, this new method is intended to separate the healthy and diseased distributions of the measured analyte. We assessed its performance by using simulated data drawn from known distributions and two previously published datasets (from human and veterinary clinical chemistry).
Results and discussion: The comparison of results obtained by the new method with the theoretical data of the simulation and determination of the reference interval for the datasets was good, thus supporting the application of this method for a rough estimation of reference intervals when the recommended procedure cannot be used.

1. Introduction

Reference values are some of the most powerful tools in medical decision-making both in human and veterinary medicine, even though decision limits are being introduced in an increasing number of cases. Recently, the recommendations for "Determining, establishing, and verifying reference intervals in the clinical laboratory" have been updated by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and the Clinical and Laboratory Standards Institute (CLSI) [1].

This working group acknowledges that the *de novo* determination of reference limits, and their regular updating for the main partitioning groups is an enormous, time-consuming and very expensive undertaking, even in human clinical pathology, for which very large data bases are available, *e.g.* through Preventive Medicine Centers [2]. This task is even more difficult in animal clinical pathology, due to the various species, breeds, breeding conditions, productions, etc., and often the small number of animals available, *e.g.* in wild species. Moreover, the absence of well documented reference intervals (RI) for many analytes sometimes makes it impossible to transfer or verify previously established reference intervals in veterinary clinical pathology.

When direct sampling is not possible, the recommendation states that indirect sampling techniques may be used "based on the assumption, confirmed by observation, that most results, even on hospital and clinic patients appear "normal"" [1]. Data thus obtained, preferably from relatively healthy individuals, can be used for calculations of reference intervals. RI determined in this manner "should be considered rough estimates at best", as the datasets are contaminated by an unknown number of values obtained from individuals that are not healthy.

Very large clinical pathology databases exist in human hospitals, and also in veterinary hospitals and animal research centers. Several attempts have been made [3,4] to use them in the estimation of human reference intervals and some are still being pursued, often with very large numbers of data [5]. To our knowledge, few attempts have been made in animal clinical pathology [6].

Several statistical methods have been proposed to estimate the 2.5 and 97.5 centiles from such "polluted" datasets, the simplest ones being based on a cut-off value below or above which the observed data are discarded [7,8]. This approach is easy to implement but can

* Corresponding author. Tel.: +33 561 193 929.
E-mail address: d.concordet@envt.fr (D. Concordet).

0009-8981/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cca.2009.03.057

severely bias results. After recursively removing the values of individuals considered as diseased outside, Kairisto and Poola [8] modelled the distribution of the remaining values as two half Gaussian distributions. More sophisticated methods have been proposed (see [6] for a brief comprehensible review). Bhattacharya [9] proposed an indirect method based on the assumption that the distributions of healthy and diseased are Gaussian. As advocated by Baadenhuijsen and Smit [10] this method does not properly describe the distributions of most analytes which present skewed distributions. To solve this problem, Baadenhuijsen and Smit extended Bhattacharya's method to mixtures of log-normal and gamma distributions. These two methods are very practical since the estimations can be performed graphically without using a modern computer. Oosterhuis et al. [11] proposed a weighting scheme that decreased the influence of outliers in the Baadenhuijsen and Smit method. However, such approaches are specific to the chosen distributions and even for mixtures of Gaussian, log-Gaussian or gamma distributions, they do not give the maximum likelihood estimate which is the best estimate that it is possible to build when the number of data available (n) is large.

The aim of this paper is to propose a method that allows separation of the distributions of healthy and diseased individuals from observation of their mixture. This method generalizes and improves Bhattacharya-like methods.

As suggested in [1] when the actual patients status is known, we assume that there are two Box–Cox transformations that respectively make Gaussian the distributions of healthy and diseased individuals. We used this distributional assumption to estimate the percentage of healthy individuals and the distributions of healthy and diseased individuals, from which we derived a reference interval.

2. The proposed method

Let us first describe intuitively how the proposed method works. Assume that high values are rather observed on diseased individuals. A high value has thus a small probability to have been observed in a healthy individual and a high probability in a diseased individual. The model (Eq. (1)) described hereafter allows to formally quantify these probabilities. The distribution of healthy is computed as if all observed values came from healthy individuals. However, during this computation, each observed value is weighted by the probability that it has been obtained on a healthy individual. The same process is used for the diseased distribution. As an example, a very high value has a small probability (≈ 0) to come from a healthy individual. Its weight during the computation of healthy distribution is ≈ 0: this means that this value is discarded. One can see that proceeding so leads to a kind of circular reasoning: we need to know the distribution of healthy/diseased to compute the probability that a value has been obtained on a healthy/diseased individual while knowing these probabilities allows to compute these distributions. The method we propose iterates this idea until there is a strict accordance between the probabilities used to perform the computations and the distributions thus obtained. We now describe how to do it practically.

Let us assume that one observes Y_i on the ith individual of the sample

$$Y_i = U_i X_i^1 + (1 - U_i) X_i^2, i = 1, .., n \tag{1}$$

where U_i is the unobserved status of this individual equal to 1 when he is healthy and 0 otherwise. Consequently, when the individual is healthy, one observes X_i^1 while X_i^2 is observed when he is diseased. The distribution of Y_i is therefore a mixture of the distributions for healthy and diseased individuals.

We assume that the random variables X_i^j that appear in the model (Eq. (1)) are mutually independent, independent of U_i and respectively distributed according to a $N(m_j, \sigma_j^2)$ up to a Box–Cox transformation k_{λ_j} where

$$k_\lambda(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} \text{ if } \lambda \neq 0 \\ \log(x) \text{ if } \lambda = 0. \end{cases}$$

If the actual status U_i of the ith individual was known, the probability density function (pdf) of the observation Y_i would be

$$\gamma_{m,\sigma^2,\lambda}(y_i) = \frac{y_i^{\lambda - 1}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(k_\lambda(y_i) - m)^2\right]$$

with $(m,\sigma^2\lambda) = (m_1,\sigma_1^2\lambda_1)$ if $U_i = 1$ and $(m,\sigma^2\lambda) = (m_2,\sigma_2^2\lambda_2)$ otherwise.

In this case, one would estimate the reference interval

$$\left[\left(1 + \hat{\lambda}_1\left(\hat{m}_1 - 2\hat{\sigma}_1\right)\right)^{1/\hat{\lambda}_1}; \left(1 + \hat{\lambda}_1\left(\hat{m}_1 + 2\hat{\sigma}_1\right)\right)^{1/\hat{\lambda}_1}\right] \tag{2}$$

where $\hat{m}_1, \hat{\sigma}_1^2$, and $\hat{\lambda}_1$ are the parameter estimations computed from the data of healthy individuals only.

Since the U_i's are not observed, we need to estimate the percentage of healthy individuals p and the parameters of the "healthy" distribution $(m_1,\sigma_1^2\lambda_1)$. The healthy distribution can be well separated from the diseased distribution when the parameters from the "diseased" distribution $(m_2,\sigma_2^2\lambda_2)$ are also estimated. We thus need to estimate the parameter $\theta = (p, m_1,\sigma_1^2\lambda_1, m_2,\sigma_2^2\lambda_2)$ to get an estimation of the reference interval.

The best way to proceed is to use the maximum likelihood estimate of the parameter θ that can be obtained as the value of θ that maximizes

$$L(\theta) = \prod_{i=1}^{n} \left(p\gamma_{m_1,\sigma_1^2\lambda_1}(y_i) + (1 - p)\gamma_{m_2,\sigma_2^2\lambda_2}(y_i)\right).$$

Unfortunately, the direct optimisation of this function is often intractable. This is the reason why we suggest using the so-called EM algorithm [12] to solve it. The EM algorithm consists of iterations of an Expectation and a Maximization step. At the kth iteration, the E step computes the conditional expectation of the log-likelihood of the complete data (Y, U) with respect to the distribution of the missing, or non-observed data U given the observed data Y at the current estimated parameter value $\theta^{(k)}$:

$$Q\left(\theta, \theta^{(k)}\right) = E\left(\log P(Y, U) | Y, \theta^{(k)}\right).$$

The M step finds $\theta^{(k+1)}$ so that

$$\theta^{(k+1)} = \arg\sup_{\theta} Q\left(\theta, \theta^{(k)}\right).$$

These two-step iterations are repeated until convergence [13].

In our problem, the actual status U_i of each individual is not observed. Estimating the mixture of distributions of diseased and healthy individuals with the EM algorithm amounts to repeating the following iterations:

[1] For each individual compute the weights w_i and v_i that respectively depend on the likelihood that this individual belongs to the healthy and diseased group.
[2] Compute the mean, variance and λ parameter of each group by affecting to each individual the corresponding weight.
[3] Repeat steps 1 and 2 until the estimated parameters no longer change between two successive iterations.

This problem can be reduced to the following iterations. The algorithm starts with some initial value $\theta^{(0)}$ of $\theta = (p, m_1,\sigma_1^2\lambda_1, m_2,\sigma_2^2\lambda_2)$. At iteration k of the algorithm, $\theta^{(k)} = (p^{(k)}, m_1^{(k)},\sigma_1^{2(k)}\lambda_1^{(k)}, m_2^{(k)},\sigma_2^{2(k)}\lambda_2^{(k)})$ is known and $\theta^{(k+1)}$ is computed from $\theta^{(k)}$ by minimizing the following function with respect to θ:

$$\sum_{i=1}^{n} a_i^{(k)}\left[2(1 - \lambda_1)\log y_i + \log\sigma_1^2 + \left(k_{\lambda_1}(y_i) - m_1\right)^2/\sigma_1^2\right] + \left(1 - a_i^{(k)}\right)\left[2(1 - \lambda_2)\log y_i + \log\sigma_2^2 + \left(k_{\lambda_2}(y_i) - m_2\right)^2/\sigma_2^2\right]$$

where

$$a_i^{(k)} = \frac{p^{(k)}\gamma_{m_1^{(k)},\sigma_1^{2(k)}\lambda_1^{(k)}}(y_i)}{p^{(k)}\gamma_{m_1^{(k)},\sigma_1^{2(k)}\lambda_1^{(k)}}(y_i) + (1 - p^{(k)})\gamma_{m_2^{(k)},\sigma_2^{2(k)}\lambda_2^{(k)}}(y_i)}.$$

Solving this problem leads to the following iterations

$$\begin{cases} p^{(k+1)} = \dfrac{1}{n}\sum_{i=1}^{n} a_i^{(k)} \\ w_i^{(k)} = a_i^{(k)} / \sum_i a_i^{(k)} \text{ and } v_i^{(k)} = \left(1 - a_i^{(k)}\right) \Big/ \left(n - \sum_i a_i^{(k)}\right) \\ \lambda_1^{(k+1)} = \arg\inf_{\lambda} \log\sum_i w_i^{(k)}\left(k_\lambda(y_i) - m_1^{(k)}\right)^2 - 2\lambda\sum_i \log y_i \\ m_1^{(k+1)} = \sum_i w_i^{(k)} k_{\lambda_1^{(k+1)}}(y_i) \\ \sigma_1^{2(k+1)} = \sum_i w_i^{(k)}\left(k_{\lambda_1^{(k+1)}}(y_i) - m_1^{(k+1)}\right)^2 \\ \lambda_2^{(k+1)} = \arg\inf_{\lambda} \log\sum_i v_i^{(k)}\left(k_\lambda(y_i) - m_2^{(k)}\right)^2 - 2\lambda\sum_i \log y_i \\ m_2^{(k+1)} = \sum_i v_i^{(k)} k_{\lambda_2^{(k+1)}}(y_i) \\ \sigma_2^{2(k+1)} = \sum_i v_i^{(k)}\left(k_{\lambda_2^{(k+1)}}(y_i) - m_2^{(k+1)}\right)^2. \end{cases}$$

These iterations are continued until the difference between $\theta^{(k)}$ and $\theta^{(k+1)}$ becomes small. Let us denote $\tilde{\theta}=(\tilde{p},\tilde{m}_1,\tilde{\sigma}_1^2,\tilde{\lambda}_1,\tilde{m}_2,\tilde{\sigma}_2^2,\tilde{\lambda}_2)$ the value of $\theta^{(k)}$ after convergence of the preceding iterations. The reference interval is then obtained as described by equation (Eq. (2)).

As we use the maximum likelihood, this reference interval is the best estimate (with respect to the precision) that can be built when the distributional assumptions we made are reasonable. It is therefore important to check these distributional assumptions. We propose hereafter a test (*P*-value) and a graphical means of checking that fully exploits the statistical properties of $\tilde{\theta}$.

A *P*-value lower than 0.05 indicates that the assumptions about the distributions in the model (Eq. (1)) are not adequately chosen. Two main reasons lead to such *P*-values: 1) the choice of the distributions is not consistent with the data suggesting other distributions to be used (*e.g.* gamma, Weibull,...), 2) there exist "outliers". Their effect on the estimation of the upper limit of the reference interval seems to be limited as soon as a rigorous analysis is performed. When the test rejects the model, some extreme values should be removed from the analysis.

If our model (Eq. (1)) is correct, the actual distribution of the data, summarized by its cumulative distribution function (cdf) F, should be equal to the one implied by the model:

$$F_{\tilde{\theta}}(y)=p\Phi\left(\frac{k_{\lambda_1}(y)-m_1}{\sigma_1}\right)+(1-p)\Phi\left(\frac{k_{\lambda_2}(y)-m_2}{\sigma_2}\right) \quad (3)$$

where Φ is the cdf of the standard Gaussian distribution.

Thus, if this model is correct, F and $F_{\tilde{\theta}}$ should be equal. Since both F and $F_{\tilde{\theta}}$ cannot be observed, we propose to compare their estimates. An estimate of F is given by its empirical cumulative distribution function (cdf)

$$\hat{F}(y)=\frac{1}{n}\sum_{i=1}^{n}1_{[Y_i\le y]},$$

where $1_{[Y_i\le y]}=1$ if $Y_i\le y$ and 0 otherwise.

On the other hand, an estimate of $F_{\tilde{\theta}}$ is $F_{\tilde{\theta}}$. A quick way to proceed is to build a QQ-plot by representing $F_{\tilde{\theta}}^{-1}(\hat{F}(y_i))$ as a function of y_i. When the points obtained are about on a straight line, the model can be considered as reasonable. However and even if the chosen model is the good one, such graphics always exhibit points far away from the line because the sampling variation has not been taken into account to build it. Also, a confidence region/a *P*-value is helpful to compare these cdf. We propose to build a 95% uniform region $R_{0.95}$ for the QQ-Plot that can be used as follows: when at least one point $(y_i, F_{\tilde{\theta}}^{-1}(\hat{F}(y_i)))$ falls outside $R_{0.95}$ there is less than 5% risk that $F\neq F_{\tilde{\theta}}$. Equivalently, we propose to compute a *P*-value by simulating the distribution of $\sup_y|F_{\tilde{\theta}}(y)-\hat{F}(y)|$ under H_0: $F=F_{\tilde{\theta}}$.

To perform these simulations, we used the properties of the maximum likelihood estimator: the Central Limit Theorem guarantees that when n is large, $\sqrt{n}\left(\tilde{\theta}-\theta\right)$ is approximately distributed according to the following Gaussian distribution $N\left(0,I^{-1}(\theta)\right)$ where $I(\theta)$ is the Fisher information matrix. This matrix depends on the true value of θ which is unknown. Louis [14] proposed a method that gives a good approximation of $I(\theta)$ that we denote by $\tilde{I}$.

We therefore propose to build $R_{0.95}$ using Monte-Carlo simulations. First, draw a large number (K_1) of θ_j (say $K_1=10\,000$) from $N(\tilde{\theta},\tilde{I}^{-1}/n)$ and for each θ_j, we simulated $K_2=50$ samples of size n using (Eq. (1)). We thus obtained $K_1\times K_2$ samples of size n, each of them giving an empirical cdf $\hat{G}$ and the distance $\sup_y|F_{\tilde{\theta}}(y)-\hat{G}(y)|$. The *P*-value of the test is obtained as the percentage of $\sup_y|F_{\tilde{\theta}}(y)-\hat{G}(y)|$ greater than $\sup_y|F_{\tilde{\theta}}(y)-\hat{F}(y)|$. After discarding the 5% simulated samples with the highest distance $\sup_y|F_{\tilde{\theta}}(y)-\hat{G}(y)|$, we built the QQ-plot for each of the 95% remaining samples. The envelope of these QQ-plots gives $R_{0.95}$.

After the distributional assumptions of the model (Eq. (1)) have been checked, the standard error of the estimates of the reference limits are computed as the standard deviation of the K_1 reference limits that can be computed using equation (Eq. (2)) but with the Monte-Carlo sample θ_j instead of $\hat{\theta}$.

3. Performance assessment

The objective of this simulation was to create samples from a predefined model distribution and to see whether or not the proposed method was able to recover this known distribution from the samples thus created.

The following parameters were used for the model distribution $p=0.9$, $m_1=3.1$, $\sigma_1^2=0.014$, $\lambda_1=-0.18$, $m_2=2.90$, $\sigma_2^2=0.022$, $\lambda_2=-0.26$. These values are not purely fictive and, except for the value of p, come from example 2 given hereafter. In this distribution, the 2 subpopulations of healthy and diseased subjects are not clearly separated and greatly overlap.

The true reference limits of the healthy subpopulation are given in Table 1. As they have been determined from a distribution model and are purely theoretical, there is no confidence interval for these values.

For each sample, the method gives a parameter estimate, the mean of which is given in Table 1 for the simulations made as below. The actual imprecision of the calculation is measured from the sample-to-sample variation of this estimate. The standard deviation of this variation is the standard error of the estimate (fourth line of Table 1). In practice, only one single sample of size n is available for the calculations, which makes it impossible to evaluate the sample-to-sample variation. However, when n is large enough, it is possible to calculate an approximation of this sample-to-sample variation using the information matrix. The average of such approximations over samples is named the asymptotic standard error (a.s.e.) and is given in the second line of Table 1.

We simulated 100 samples of size $n=1000$ and $n=10{,}000$ and used our method to separate the distributions of healthy and diseased individuals. We estimated p and the limits of the reference interval of the "healthy" distribution thus separated. Estimates of p and of the reference limits from the 1000 and 10,000-samples are presented in Figs. 1 and 2, and summarized in Table 1. In both cases, the means were very close to the theoretical values, and the imprecision of the estimates was two-fold higher for the 1000-sample than for the 10,000-sample series. When 10,000-samples were used, the minimum and maximum estimations of p were 0.844 and 0.944, whereas the true value was 0.9. Estimates of the lower limit of the reference interval were very precise and accurate, whatever the number of values. The more precise estimation of the upper limit of the reference interval was obtained with the 10,000-samples, ranging from 145.3 to 167.4, whereas the true value was 155.76. This large imprecision can probably be explained by the position of the actual upper limit of the reference interval which is about at the mode of the distribution of diseased individuals. We can see in Fig. 2 that estimations of the upper limit are more often lower than the actual value of this limit. This suggests that the distribution of the corresponding estimator is as skewed as the distribution of the extreme values.

The estimates of the reference limits given by our method are unbiased since the mean of the 100 estimates is close to the true value for the 3 parameters. It can also be observed that the standard error of the estimator is very close to the a.s.e. But a standard statistical result shows that the maximum likelihood is the best estimate that it is possible to build when n is large. Consequently, no unbiased estimate with a smaller standard error exists, which implies that there is no better estimation than the one thus provided.

The parameters that define the distribution of diseased individuals are not well estimated in the above example. This can be explained the low number of data available for the diseased individuals. The population contains only 10% of diseased individuals therefore, the samples of size $n=1000$ and 10,000 contained about 100 and 1000 diseased individuals. These numbers of diseased individuals appear too small for a precise estimation.

4. Two practical "real" examples

In this section, we show the results obtained when this method was applied to two previously reported examples.

4.1. Example of plasma TSH concentration in human males

An indirect determination of the plasma TSH upper reference limit was carried out using more than 19,000 unselected hospital results [15]. A 2056-value sample was selected by the authors comprising males for which repeat analyses had been discarded and the corresponding histogram was published cf Fig. 1B of [15]. The estimated upper limit, using the method proposed by Kairisto and Poola [8] implemented in the program GraphROC, was 3.5 mIU/L. The

Table 1
Parameters of the simulated model distribution and estimations of the percentage of healthy individuals and of the reference interval using the proposed method.

	$n=1000$			$n=10{,}000$		
	p	Lower lim.	Upper lim.	p	Lower lim.	Upper lim.
True value	0.90	53.73	155.76	0.90	53.73	155.76
a.s.e.	0.094	1.01	10.78	0.034	0.33	3.43
Mean	0.899	53.53	155.8	0.886	53.34	154.49
s.e.	0.045	1.18	9.62	0.022	0.6	5.19

(p = percentage of healthy individuals; Lower and Upper lim. = the limits of the reference interval; a.s.e. = average of asymptotic standard errors; s.e. = standard deviation of estimates obtained with the 100 samples).

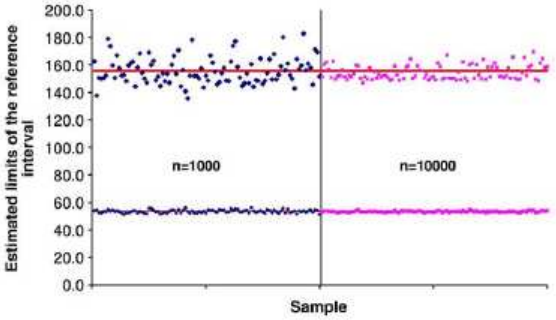


Fig. 2. Estimated reference limits obtained from 100 samples of size $n=1000$ and 100 samples of size $n=10{,}000$. The horizontal lines are the actual reference limits. The lower reference limit is well estimated while the estimated upper reference limit has quite a large imprecision probably because a large percentage of diseased individuals had concentrations close to this limit.

only information available being a histogram, we did not have individual data. We thus simulated individual data by assuming that they were uniformly distributed within each histogram class. We assumed that values included in negative classes corresponded to results below the limit of quantification (LOQ) and deleted them before subsequent analyses. The distribution of values above 0 was assumed to be a mixture of two Gaussian distributions after Box–Cox transformations. The observed and fitted distributions thus obtained as well as the components of the mixture are presented in Fig. 3. The overall shape of the distribution is very similar to the original one. The percentage of individuals with a TSH concentration below the limit of quantification was 4.21%. Using the proposed method, the estimated percentage of healthy individuals with a TSH concentration above the limit of quantification was 77.10%. The distributions of plasma concentration of healthy (with TSH > LOQ) and diseased patients were estimated to be Gaussian after a Box–Cox transformation with parameters 0.2835 and −1.433 respectively. Since 4.21% of individuals were assumed to be healthy because they had a value below the LOQ, the next step was to calculate the 97.50–4.21 = 93.29% quantile of the distribution of the healthy individuals to obtain the upper limit of the reference interval. This latter was estimated to be 3.770 ± 0.333 mIU/L which is slightly higher but very close to the estimate provided in the original study. The diagnostic plot represented in Fig. 4 shows that the proposed model is acceptable ($P>0.05$) as the observed QQ-Plot is entirely contained within the confidence region.

4.2. Example of creatinine concentration in dog

A study to determine the diagnostic efficiency of plasma creatinine and urea concentrations for the diagnosis of canine kidney diseases finally included 3822 cases, of which 37% were healthy [16]. The nonparametric reference interval was 53–151 µmol/L (90% confidence intervals were 52 to 55 µmol/L and 148 to 159 µmol/L). As discussed in this article, even if the health status of these animals was known, there was no way of ensuring the quality of the diagnoses. We therefore decided to ignore the available diagnoses in the present study. The observed distribution of plasma creatinine in the 3822 dogs is represented in Fig. 5. The diagnostic test described in the Proposed methods section gave a P-value = 0.580 thus showing that the proposed model is acceptable.

The estimated percentage of healthy individuals was $\hat{p}=79.93\%$, quite higher than the "true" percentage. The plasma concentration distribution of healthy patients was estimated to be $N(3.100;0.014)$ after a Box–Cox transformation with parameter −0.179. This led to an upper limit of the reference interval of 160.9 ± 8.7 µmol/L, which is close to the limit obtained in healthy dogs based on reported clinical status [16].

When the method proposed by Baadenhuijsen and Smit [10] was applied to the same data, the estimated upper limit was 135.1 µmol/L; for a mixture of gamma distributions. Using the same method but with the weighting scheme proposed by Oosterhuis et al. [11] we obtained 142.6 µmol/L. In contrast the method of Kairisto and Poola (with no use of external information) [8] gave 340.8 µmol/L.

These large differences between the estimations could be explained: 1/ by the high sensitivity of Baadenhuijsen's method to the choice of aligned points and 2/ by the threshold implied by the ±4 SD rule in the Kairisto and Poola method that keeps the data from diseased individuals

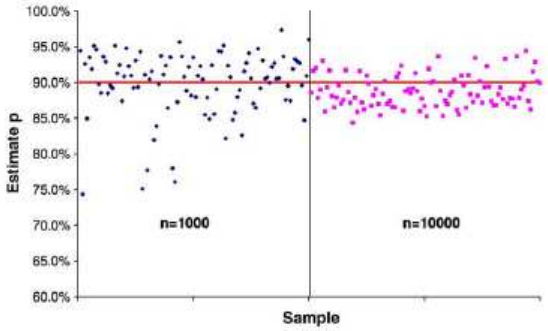


Fig. 1. Estimated percentage of healthy individuals obtained from 100 samples of size $n=1000$ and 100 samples of size $n=10{,}000$. The horizontal line is the value that was used for simulations. This percentage is well estimated; the imprecision decreases with n.

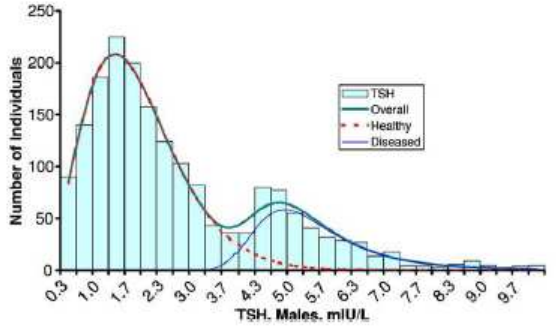


Fig. 3. The histogram represents the observed TSH concentrations in human males obtained from [16]. The concentrations below the LOQ have not been represented (but taken into account for the RI determination (see text). The dotted, thin and thick curves represent the TSH distribution obtained using the proposed method for the diseased, healthy individuals and for the entire population respectively.

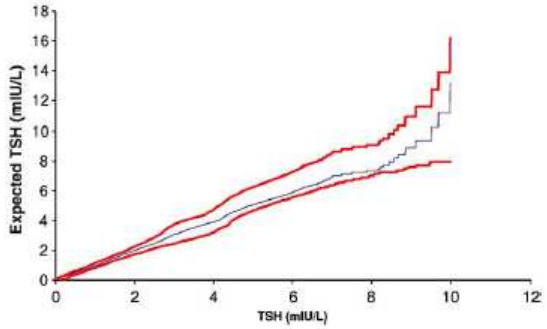


Fig. 4. Diagnostic plot proposed in Section 2 for the TSH data. The empirical cdf $\hat{F}$ (thin line) is totally included within the 95% confidence region (thick lines) or ($P > 0.05$). The assumptions about the model (shape of distributions) are thus in accordance with the data.

for computation when no manual exclusion of data is performed. On this example, the modification of the Bhattacharya method proposed by Oosterhuis, Modderman and Pronk produced a reasonable estimate but its imprecision cannot be computed.

5. Discussion

In this article we propose a method for estimating reference interval of an analyte from a large set of data when the actual status of the individuals in the analyzed sample is unknown. This method is not new; it is currently considered as the gold-standard method for mixtures analysis.

This method compensates some of the deficiencies of existing methods used to determine reference intervals. Bhattacharya and Baadenhuijsen's methods assume specific shapes, mainly Gaussian and log-Gaussian, for the distributions of healthy and diseased. These methods are not flexible enough to adapt to other distribution shapes, unlike the proposed method. With these methods the computation of reference intervals relies mainly on a graphical analysis that requires a subjective choice of aligned points in a scatter plot. This choice strongly influences the results obtained and the confidence intervals for the provided estimates cannot be computed. Actually, these relatively "old" methods were well adapted to a time when computations were not performed with computers.

When the actual distributions have a shape close to the assumed one, the method that we propose provides, for large n, the maximum likelihood estimation of the reference interval, *i.e.* the most precise unbiased estimate that can be computed.

We also propose a method that allows rapid (and visual) checking of whether or not these assumptions about shapes of distributions are reasonable.

Although this method is an improvement on existing ones, it has some weaknesses that restrict its use. First of all, it requires the distributions to be Gaussian up to a Box–Cox transformation. This assumption is essential for the method to give unbiased results.

Secondly, the method is designed to identify two subpopulations and it succeeds in doing this even if the subpopulations are not the ones which were expected. As an example, we used this method on another dataset and it separated young *versus* old individuals instead of healthy *versus* diseased ones.

Thirdly, this method correctly separates healthy individuals, but as previously stated, the limits thus obtained are "rough estimates at best" [1]. Even when this new method which minimizes the imprecision of such estimates was used, the s.e. of the upper limit was within the range of 5 to 10% in the simulations and in the examples studied, which means, that the 90% confidence interval of this limit was in the range of [140.0; 171.6] for $n = 1000$.

Fourthly, this method cannot be used to analyse serial measurements in the same individual. As the method assumes that the random variables are mutually independent, these data should be discarded from the analysis. In practice, omitting repeated measurements often reduces the number of available data.

Finally, this method estimates the distribution of the "minor" *i.e.* less represented subpopulation, with greater imprecision than that of the major. This algorithm converges (*i.e.* stabilizes after some iterations) only when the distributions of healthy and diseased individuals are different enough with regards to the sample size n. Even very close distributions can be distinguished when n is very large. On the contrary, this algorithm may fail to estimate quite different distributions when only a small sample size is available. This problem is connected with the general problem of identifiability of mixtures [17].

A large number of refinements and extensions can be considered. It is likely that the skewness we assumed (Gaussian after a Box–Cox

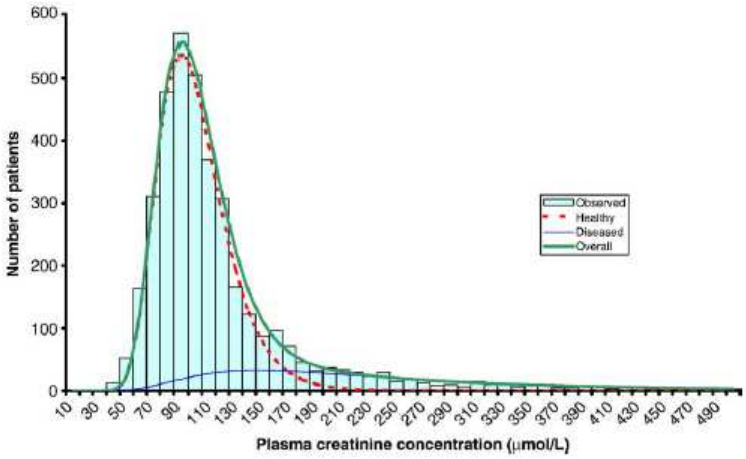


Fig. 5. Histogram of the plasma creatinine concentrations in dogs obtained from [17]. The dotted, thin and thick curves represent the creatinine distribution obtained for diseased animals, for healthy animals and for the entire population respectively, using the proposed method.

transformation) for the distributions cannot cover all the situations encountered in practice. However, following exactly the same EM steps, other distributions can be used, such as the exponential family (e.g. gamma, Weibull,...). An interesting study would be to try distributions with different shapes and to use the graphical diagnostic plot proposed in the paper to identify the shapes that are compatible with the data.

Examples composed of 2 subpopulations (healthy *versus* diseased individuals) are simplistic. Actual populations are more complex, e.g. subpopulations of individuals affected by a different disease than the one for which the analyte has been measured. Thus, a natural extension of this work will be to build an algorithm that can deal with an unknown number of subpopulations.

Finally, demographic variables such as age, gender, weight could be used as covariates as in [18]. Even if these variables are not known for each individual, they would help identify the populations separated, thus minimizing the risk of wrongly "labelling" the identified subpopulations as healthy or diseased.

References

[1] CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline. Third ed. Wayne, PA: CLSI; 2008.
[2] Siest G. Study of reference values and biological variation: a necessity and a model for Preventive Medicine Centers. Clin Chem Lab Med 2004;42:810–6.
[3] Neumann GJ. The determination of normal ranges from routine laboratory data. Clin Chem 1968;14:979–88.
[4] Glick JH. Statistics of patient test values: application to indirect normal range and to quality control. Clin Chem 1972;18:1504–13.
[5] Grossi E, Colombo R, Cavuto S, et al. The REALAB project: a new method for the formulation of reference intervals based on current data. Clin Chem 2005;51:1232–40.
[6] 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.
[7] Petitclerc C. Normality: the unreachable star? Clin Chem Lab Med 2004;42:698–701.
[8] Kairisto V, Poola A. Software for illustrative presentation of basic clinical characteristics of laboratory tests GraphROC for Windows. Scand J Clin Lab Invest Suppl 1995;222:43–60.
[9] Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. Biometrics 1967;23:115–35.
[10] Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. J Clin Chem Clin Biochem 1985;23:829–39.
[11] Oosterhuis WP, Modderman TA, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC? Ann Clin Biochem 1990;27:359–65.
[12] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 1977;39:1–38 (With discussion).
[13] Wu CF. On the convergence properties of the EM algorithm. Ann Stat 1983;11:95–103.
[14] Louis TA. Finding the observed information matrix when using the EM algorithm. J R Stat Soc B 1982;44:226–33.
[15] Giavarina D, Dorizzi RM, Soffiati G. Indirect methods for reference intervals based on current data. Clin Chem 2006;52:335–6.
[16] Concordet D, Vergez F, Trumel C, Diquélou A, Lanore D, et al. A multicentric retrospective study of serum/plasma urea and creatinine concentrations in dogs using univariate and multivariate decision rules to evaluate diagnostic efficiency. Vet Clin Pathol 2008;37:96–103.
[17] Holzmann H, Munk A, Gneiting T. Identifiability of finite mixtures of elliptical distributions. Scand J Statist 2006;33:753–63.
[18] Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. Stat Med 2000;19:2943–62.

Partie 4
Mettre au point un outil statistique d'utilisation simple et rapide

Nous avons vu que lorsque cela est possible, l'utilisation de la méthode non paramétrique est recommandée pour la détermination des limites de référence. Cela impose donc un nombre minimum de sujets égal à 40. De plus, avec un si petit effectif, il est impossible de faire une estimation non paramétrique des intervalles de confiance des limites.

Dans tous les cas où les effectifs sont faibles, il est donc nécessaire de faire appel à des calculs plus ou moins complexes, à des transformations mathématiques permettant de rendre les distributions voisines de distributions gaussiennes, etc. A l'heure actuelle, les calculs les plus simples sont disponibles dans des logiciels de statistiques classiques et dans des logiciels commerciaux plus ou moins dédiés (Analyse-It [25], RefVal [26]). Certaines des nouvelles méthodes statistiques recommandées par les groupes d'experts ne sont disponibles que dans des logiciels spécifiquement mathématiques, difficiles d'accès pour un utilisateur non averti, voire pas disponibles du tout ou peut-être dans un nouveau logiciel d'un coût prohibitif [27]. A l'opposé, le logiciel Excel est extrêmement répandu et relativement simple d'utilisation. Il a donc été décidé de mettre au point un jeu de macro-instructions pour Excel qui, d'un simple clic, produit pour chaque série de valeurs de référence, cinq types de calculs d'intervalles de référence, générés par les méthodes statistiques précédemment citées. Le logiciel réalise également automatiquement des figures permettant d'évaluer la distribution des valeurs de référence ainsi que quelques commentaires sur la pertinence d'utiliser tel ou tel intervalle de référence calculé. Ces suggestions d'utilisation sont fondées sur les

recommandations internationales appliquées à chaque jeu de données entré dans le logiciel. Elles peuvent guider un utilisateur néophyte pour choisir la procédure de détermination de l'intervalle de référence la plus appropriée. Cependant, elles ne pourront jamais compenser un mauvais échantillonnage des individus de référence, cette étape restant souvent le point critique de la détermination d'un intervalle de référence.

Article n°7

Geffré A, Concordet D, Braun JP, Trumel C.

Reference Value Advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel.

Veterinary Clinical Pathology, 2011;40:107-12.

Veterinary Clinical Pathology ISSN 0275-6382

TECHNICAL NOTE

Reference Value Advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel

Anne Geffré[1], Didier Concordet[2], Jean-Pierre Braun[1,2], Catherine Trumel[1]

[1]Department of Clinical Sciences and [2]UMR181 Physiopathologie and Toxicologie Expérimentales INRA, ENVT, Ecole Nationale Vétérinaire, Toulouse, France

Key Words
Box–Cox, confidence intervals, Gaussian, *Q–Q* plot, reference values, robust method

Correspondence
A. Geffré, Department of Clinical Sciences, Ecole Nationale Vétérinaire, 23 Chemin des Capelles, 31076 Toulouse Cedex 3, France
E-mail: a.geffre@envt.fr

DOI:10.1111/j.1939-165X.2011.00287.x

Abstract: International recommendations for determination of reference intervals have been recently updated, especially for small reference sample groups, and use of the robust method and Box–Cox transformation is now recommended. Unfortunately, these methods are not included in most software programs used for data analysis by clinical laboratories. We have created a set of macroinstructions, named Reference Value Advisor, for use in Microsoft Excel to calculate reference limits applying different methods. For any series of data, Reference Value Advisor calculates reference limits (with 90% confidence intervals [CI]) using a nonparametric method when $n \geq 40$ and by parametric and robust methods from native and Box–Cox transformed values; tests normality of distributions using the Anderson–Darling test and outliers using Tukey and Dixon–Reed tests; displays the distribution of values in dot plots and histograms and constructs *Q–Q* plots for visual inspection of normality; and provides minimal guidelines in the form of comments based on international recommendations. The critical steps in determination of reference intervals are correct selection of as many reference individuals as possible and analysis of specimens in controlled preanalytical and analytical conditions. Computing tools cannot compensate for flaws in selection and size of the reference sample group and handling and analysis of samples. However, if those steps are performed properly, Reference Value Advisor, available as freeware at http://www.biostat.envt.fr/spip/spip.php?article63, permits rapid assessment and comparison of results calculated using different methods, including currently unavailable methods. This allows for selection of the most appropriate method, especially as the program provides the CI of limits. It should be useful in veterinary clinical pathology when only small reference sample groups are available.

Determination of reference intervals is a long, difficult, and expensive task for all laboratories, mainly owing to the difficulty in selecting sufficient numbers of well-characterized healthy subjects and in ensuring perfectly controlled preanalytical and analytical conditions.[1–3] Computations comprise standard statistical procedures,[4–6] but are often perceived as the most difficult part of the task and are sometimes performed using simplistic assumptions about the data. International recommendations have been recently updated by the International Federation of Clinical Chemistry (IFCC) and the Clinical and Laboratory Standards Institute (CLSI).[7] These recommendations confirm that a nonparametric method is preferred when the number of reference individuals within 1 group is at least equal to 120. The revised CLSI guideline introduces determination of reference intervals from smaller reference samples based on a robust method,[8] preferably after transformation of the data to a distribution that is closer to Gaussian or normal. Box–Cox transformation is often used to transform data to normality, but is only available on statistical software, such as R (http://www.r-project.org), S-PLUS (Tibco Software Inc., Palo Alto, CA, USA), and SAS (SAS Institute Inc., Cary, NC, USA). Moreover, the robust method is not currently widely available. Thus, the aim of this study was to

create a set of macroinstructions for Microsoft Excel (Microsoft, Redmond, WA, USA) that would permit evaluation and transformation of data distributions and computation of reference intervals with the corresponding 90% confidence intervals (CI) with a single click. In addition, we aimed to provide minimal guidance in the form of comments based on international recommendations. Thus, the program was named Reference Value Advisor.

Reference Value Advisor

The software is a set of visual basic macros developed in Microsoft Excel for Windows v. 2003, chosen as the development tool because of its wide availability. The selection of computations performed was guided by the IFCC–CLSI recommendations and included:

- Common descriptive statistics: sample size, mean, median, standard deviation, minimum, and maximum values
- Test of normality according to Anderson–Darling, with histograms and *Q–Q* plots: regardless of the results of this test on raw data, the generalized Box–Cox transformation is performed. Its λ_1 and λ_2 parameters are determined by the maximum likelihood estimator that provides the highest precision.
- Tests of "outliers": Dixon–Reed and Tukey's tests are used.[7] The former detects a single potential outlier based on the ratio of its distance to the nearest value divided by the whole range of values. The latter is based on the median and interquartile range. The number of potential outliers is reported on the main results sheet with a list of outliers produced on a separate sheet. A box and whiskers diagram (median, 25th and 75th percentiles, 95% CI of mean), with a dot plot showing all values, can also be used to visually identify the outliers.
- Computations of reference intervals: for any series of data, Reference Value Advisor calculates and reports 5 reference intervals based on assumptions about data distribution. The first 4 intervals are obtained by using standard and robust methods on both nontransformed and transformed data, thus combining the need for data transformation with assumptions regarding Gaussian or symmetrical distribution of data (Table 1). The last (fifth) interval is obtained without any assumption about data distribution. Thus, it is distribution free or equivalently nonparametric, but is only computable when the sample size is large enough ($n \geq 40$). The standard intervals require data distribution to be Gaussian before or after Box–Cox transformation, whereas the robust method only requires symmetry of the data.

Table 1. Assumptions regarding data distribution.

	Native Data	Box–Cox Transformed Data
Standard interval	Gaussian	Gaussian after Box–Cox transformation
Robust Horn interval	Symmetrical	Symmetrical after Box–Cox transformation

- CIs of the reference limits computed for each method: for the standard method, the 90% CI is obtained using a parametric bootstrap when $n \leq 20$. In all other cases a nonparametric bootstrap method is used. For the nonparametric method, the CI is determined according to tables for $120 < n < 370$[7,9]; if $n < 120$, a bootstrap method is used.

It was decided to consistently provide all computation results to permit selection of the most relevant results by expert users. However, it was also decided that, for inexperienced users, guidance should be provided based on IFCC–CLSI recommendations. Thus, a color code is used when reporting reference limits: green indicates *in agreement with recommendations*, orange indicates *use with caution or avoid using because possible outliers are detected*, and red indicates that *the distribution is not Gaussian*. In a comments section, the following explanations are added and are activated according to the series of data analyzed, information in the IFCC–CLSI recommendations, and the calculation methods used:

- Possible outliers detected according to Tukey or Dixon: IFCC–CLSI C28-A3 recommends that unless outliers are known to be aberrant observations, emphasis should be on retaining rather than deleting them.
- Suspect data detected according to Tukey: IFCC–CLSI C28-A3 recommends that unless these data are known to be aberrant observations, emphasis should be on retaining rather than deleting them.
- The sample size is large enough to compute a nonparametric reference interval.
- The sample size is too small ($n < 40$) to compute a nonparametric reference interval.
- The sample size is too small ($n < 80$) to compute precise CIs for the limits of the nonparametric reference interval.
- The CIs of the limits of the nonparametric reference interval were determined using a bootstrap method.
- Data should be analyzed with a nonparametric method. As an alternative, the robust method with Box–Cox transformation may be used after checking the symmetry of the distribution.

Table 2. Example of 83 randomly selected plasma creatinine values (μmol/L) in healthy dogs.

50	53	58	62	62	62	62	62	64	64	64	67
67	68	69	71	71	71	72	72	75	77	79	80
80	80	80	80	81	81	82	82	82	83	84	84
85	87	88	88	89	90	91	91	92	93	94	96
96	97	97	97	97	99	99	100	100	101	106	106
106	107	107	107	109	112	114	115	115	115	117	119
120	120	120	124	127	127	141	146	147	168	177	

- The robust method with Box–Cox transformation of the data gives the following reference interval (x; y) (this message is provided when the nonparametric reference interval cannot be determined in accordance with the IFCC–CLSI C28-A3 recommendation and when the performance of the robust method+Box–Cox is satisfactory).
- The sample size is too small ($n < 20$) to compute CI for the limits of the reference interval obtained with the robust method with Box–Cox transformation.
- The 90% CI of 1 (or more) limit is wider than is recommended in IFCC–CLSI C28-A3.

Example of Use

Reference Value Advisor was utilized to evaluate a set of 83 canine plasma creatinine values randomly selected using the Excel RAND function from a large sample of values obtained from healthy dogs used in a previous study[10] (Table 2). Descriptive statistics, detection of possible outliers, and estimations of reference intervals were performed (Figure 1), and a histogram and Q–Q plot were generated (Figure 2). Results obtained using Analyse-It (Analyse-It, Leeds, UK) were identical to those obtained using Reference Value Advisor (data not shown).

Discussion

Reference Value Advisor met the established objectives as it automatically performs a series of computations on any set of data and provides more information than is currently available in other software programs. Provision of these calculations is also potentially risky if they are used in the absence of a certain degree of expertise in determining reference intervals. For this reason, we provided comments that may aid less experienced users and encourage them to refer to IFCC–CLSI recommendations.

The main strengths of the Reference Value Advisor are the:

- ease of use and availability of computations, such as the robust method and Box–Cox transformation, which are either rarely performed or not performed at all by commonly used mathematical programs;
- reporting of all values obtained in a graphical format, either as a dot plot or histogram;
- detection of "extreme" values, ie, possible outliers, by the 2 methods recommended by IFCC–CLSI and their clear identification on a separate report sheet. These outliers have little influence on determination of limits by nonparametric methods but greatly influence the width of the CI for these limits, especially for the upper limit. However, emphasis should be on retaining rather than rejecting outliers unless it is certain they represent aberrant observations.[7] In the case of aberrant values, computations must be performed again after the outliers are deleted;
- systematic determination of the 90% CI of the reference limits. It is recommended that the width of the 90% CI of reference limits should not exceed the width of the reference interval by 20% or more[6];
- possible comparison of reference intervals estimated by different methods in small samples. When different methods do not give similar results, the validity of the data may be questioned.

Reference Value Advisor also has weaknesses:

- In some cases, Gaussian distribution cannot be achieved by Box–Cox transformation, and other transformations, not included in this software, might be better.
- There is no test for distribution symmetry, which is required for the robust method. The current version includes only a test for normality, which requires more than symmetrical distribution.
- Reference Value Advisor currently lacks some procedures, such as Harris and Boyd's test for partitioning[11] and a method for regression-based determination of reference intervals, eg, for age-dependent variables.[12] However, these methods can be implemented in the future.
- The set of macroinstructions currently works on Microsoft Excel for Windows up to Excel XP and 2003; updated versions are currently being developed for Excel for Mac and Excel 2010.

From a practical point of view, when $n > 120$ the recommended nonparametric method can be

Reference Value Advisor v1.3

Results for Creatinine

Date 12/02/2010
Performed by A. GEFFRE

Method	Untransformed data Standard	Untransformed data Robust	Box-Cox transformed data Standard	Box-Cox transformed data Robust	Nonparametric
N	83	83	83	83	83
Mean	93,0		9,7		
Median	89,5	90,1	9,7	9,7	
SD	24,9	25,3	2,5	2,5	
Minimum	50,0	50,0	3,5	3,5	
Maximum	177,0	177,0	16,2	16,2	
λ 1 coefficient **Box-Cox**			-41,5	-41,5	
λ 2 coefficient **Box-Cox**			0,418	0,418	
p-value **Anderson-Darling**	0,025	0,025	0,862	0,862	
Outliers **Dixon**					
Outliers **Tukey**	0	0	0	0	
Suspect data **Tukey**	2	2	0	0	
Lower limit of reference interval	43,1	39,7	54,8	54,7	53,4
Upper limit of reference interval	142,9	140,4	152,1	151,8	146,9
90% CI for lower limit	37,1	29,4	51,6	51,3	50,0
	50,4	48,9	58,4	58,3	62,0
90% CI for upper limit	132,2	131,0	139,8	139,0	126,8
	153,6	151,0	164,7	164,3	176,3

Comments
The sample size is large enough to compute the nonparametric reference interval : [53.3 ; 146.9].
The confidence intervals of the limits of the nonparametric reference interval was determined using a bootstrap method.
The 90% CI of one (or more) limit is larger than recommended in IFCC-CLSI C28-A3.

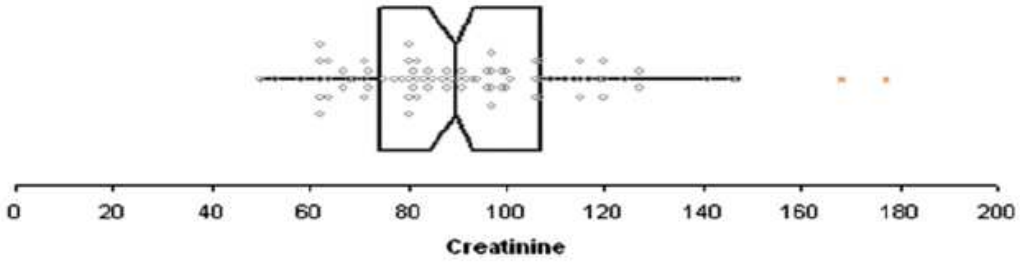


Figure 1. Printout of computations made with a series of 83 canine plasma creatinine concentrations in Table 2 using Reference Value Advisor. The untransformed distribution is not Gaussian (*P* [Anderson–Darling] = 0.025), but Box–Cox-transformed values fit a Gaussian distribution (*P* [Anderson–Darling] = 0.862). This can be visually inspected using a *Q–Q* plot and histogram (Figure 2). The 2 suspect outliers detected by Tukey's test appear as orange crosses on the box and whiskers diagram; they are close to the rest of the data and there is no valid reason to reject them. The color code indicates that the results calculated from the untransformed data (red and orange) should preferably not be used as the distribution is not Gaussian. The green color indicates that the reference limits are in agreement with recommendations.

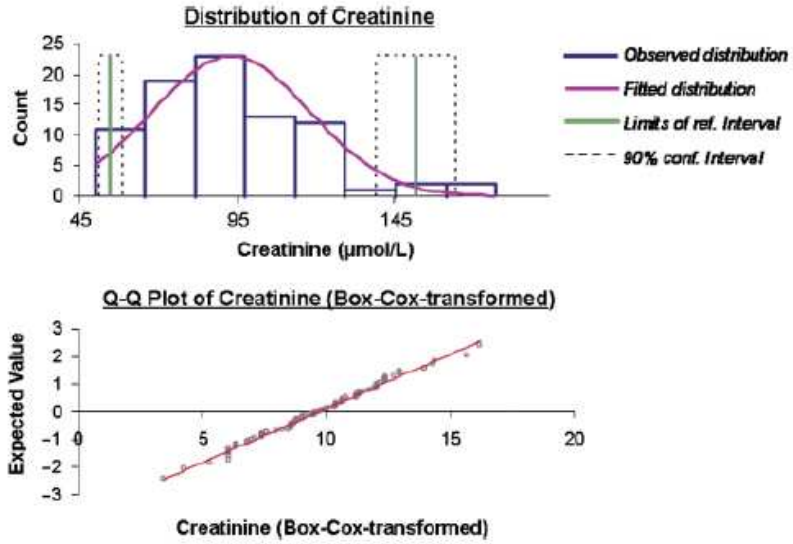


Figure 2. Histogram of untransformed and *Q–Q* plot of Box–Cox-transformed canine plasma creatinine results obtained from 83 healthy dogs. The histogram shows the asymmetric distribution of creatinine values; the linear arrangement of the *Q–Q* plot of Box–Cox-transformed values shows that the transformed distribution was very close to Gaussian.

satisfactorily used, except when visual inspection of data indicates that distribution is not unimodal. For smaller samples, the robust method on Box–Cox transformed data usually gives the narrowest CIs for the reference limits.

Conclusion

Reference Value Advisor computes reference intervals, regardless of data type or species. However, it must be emphasized that computations represent only the last step in determination of reference intervals and that the most critical steps in this process are selection of reference individuals and preanalytical and analytical conditions. Computations cannot compensate for lack of quality in any of these areas.

Availability of Reference Value Advisor

The freeware can be downloaded at http://www.biostat.envt.fr/spip/spip.php?article63

Acknowledgments

The authors greatly appreciate the input of Dr. Kristen Friedrichs, University of Wisconsin–Madison, who evaluated preliminary versions of this software and provided a series of detailed observations, comments, corrections, and suggestions.

Disclosure: The authors have indicated that they have no affiliations or financial involvement with any organization or entity with a financial interest in, or in financial competition with, the subject matter or materials discussed in this article.

References

1. Geffré A, Friedrichs K, Harr K, et al. Reference values: a review. *Vet Clin Pathol.* 2009;38:288–298.
2. 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.
3. Horn PS, Pesce AJ. *Reference Intervals. A User's Guide.* Washington, DC: AACC Press; 2005.
4. Solberg HE. The theory of reference values Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem.* 1983;21:749–760.
5. Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal program. *Clin Chem Lab Med.* 2004;42:710–714.
6. Harris EK, Boyd JC. *Statistical Basis of Reference Values in Laboratory Medicine.* New York, NY: Marcel Dekker; 1995.
7. CLSI. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline.* 3rd ed. Wayne, PA: CLSI; 2008.

8. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44:622–631.
9. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*. 1971;17:275–284.
10. Concordet D, Geffré A, Braun JP, et al. A new approach for the determination of reference intervals from hospital-based data. *Clin Chim Acta*. 2009;405:43–48.
11. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem*. 1990;36:265–270.
12. Virtanen A, Kairisto V, Uusipaikka E. Regression-based reference limits: determination of sufficient sample size. *Clin Chem*. 1998;44:2353–2358.

Comme nous venons de le voir, les questions posées par les valeurs de référence peuvent être abordées de la même manière en biologie animale qu'en biologie humaine et les recommandations qui ont été faites par l'IFCC et le CLSI sont très largement transposables, même si la validation d'intervalles préalablement établis et l'utilisation de petits échantillons amènent à de sérieuses réserves.

Cependant, en biologie animale, les intervalles de référence sont beaucoup plus difficiles à définir qu'en biologie humaine pour de multiples raisons qui conduisent à s'interroger sur la validité et sur les utilisations possibles des intervalles de référence publiés dans des revues scientifiques et davantage encore dans des traités généraux :

- dans les principales espèces domestiques, où les effectifs disponibles sont souvent suffisamment nombreux pour permettre de sélectionner des échantillons de référence de grande taille, les races ou les conditions d'élevage sont des facteurs de partition encore insuffisamment étudiés Par exemple, la variabilité liée à la taille chez le chien commence à être étudiée, surtout grâce aux lévriers « en retraite » qui sont fréquemment des donneurs de sang aux Etats-Unis et en Grande Bretagne [28], [29], [30], [31], [32]. Il est par contre plus difficile de disposer de nombreux individus de référence dans les races de petite taille. De même, les races de vaches ou de brebis, à viande ou laitières, diffèrent par leur patrimoine génétique mais aussi par leurs conditions d'élevage, leur stade de production et/ou leur alimentation. En toute rigueur, il faudrait connaître l'impact éventuel de ces facteurs sur les principales variables biologiques utilisées en pratique ;

- dans les espèces de compagnie facilement stressables, comme les chats ou les lapins, on peut s'interroger sur la représentativité d'intervalles de référence déterminés chez des sujets tranquillisés voire anesthésiés [33] ; il en est de même pour les espèces sauvages chez lesquelles les animaux peuvent être poursuivis avant la capture, éventuellement tranquillisés puis prélevés ;

- dans les espèces sauvages ou de zoo, les effectifs sont parfois très faibles, dépassant à peine une dizaine de sujets dont l'état de santé est souvent mal caractérisé. Cependant, les quelques valeurs collectées peuvent avoir un intérêt non négligeable pour évaluer l'état de santé de congénères.

Face à ces questions, la multiplication d'études de détermination *de novo* d'intervalles de référence constitue une impossibilité en raison du temps et du coût

impliqués, voire du nombre d'animaux existants au total [24]. Cela impose donc des palliatifs. Dans ce domaine, les pistes pourraient être les suivantes :

- pour les espèces majeures et les analytes les plus fréquents, suivre d'aussi près que possible les recommandations internationales ;

- pour les autres cas, faire ce qui est possible mais ne pas essayer de faire dire aux valeurs observées davantage que ce qu'elles signifient. Lorsque les effectifs sont faibles, l'information d'un clinicien est probablement aussi bonne avec une médiane et un intervalle minimum-maximum qu'avec l'illusion de précision donnée par une moyenne et un écart-type ;

- dans tous les cas, pour que le travail effectué puisse être exploité, donner un maximum d'information démographiques, indiquer précisément les critères d'inclusion et d'exclusion, même s'ils ont été établis *a posteriori*, et les caractéristiques des techniques analytiques.

Il n'existe pas actuellement en biologie animale de structure internationale similaire à l'IFCC ou au CLSI et le sujet des valeurs de référence est perçu à tort comme une simple question mathématique, *a priori* rébarbative. Il en résulte qu'un certain nombre de spécialistes de pathologie méconnaissent la question des valeurs de référence et omettent dans leurs publications les informations minimales qui rendraient leurs résultats plus largement utilisables. Un effort devrait être entrepris par les Collèges européen [3] et américain [34] de spécialistes de biologie médicale vétérinaire pour que des recommandations soient faites, spécialement dans le cas d'échantillons de petite taille.

Les intervalles de référence de population ne sont certes que l'un des auxiliaires des cliniciens pour contribuer au diagnostic, au pronostic ou au suivi des affections humaines ou animales. On assiste progressivement au développement d'autres approches, comme les valeurs de référence individuelles [35] et les seuils de décision [36] mais il est vraisemblable que l'étude des intervalles de référence de population va encore rester pendant de nombreuses années l'un des moyens privilégiés d'approche de la variabilité interindividuelle des sujets en bonne santé et donc l'une des bases du diagnostic médical et du suivi des patients.

Il semble que l'un des points importants soit aussi que les utilisateurs finaux, médecins ou vétérinaires, soient mieux informés des limites de ce qu'ils lisent. Un grand nombre d'entre eux sont avides d'informations simplifiées, sous forme de tableaux à multiples entrées avec les classes de partition dans un sens et les variables dans l'autre, ou bien des « valeurs normales » indiquées à côté des résultats sur une feuille de compte-rendu d'un laboratoire. Pour atteindre cet objectif, il faudrait tendre à donner des comptes-rendus contenant davantage d'informations : sur les incertitudes qui entourent les limites de référence mais aussi sur les résultats d'analyses. Il est vraisemblable que cela sera mal accepté dans un premier temps car cela devrait générer davantage de doutes. Cependant, c'est toute l'expertise des cliniciens de prendre des décisions en pesant des incertitudes, encore faudrait-il qu'ils soient informés de certaines d'entre elles.

Enfin, la question des valeurs de référence dépasse très probablement le domaine de la biologie médicale. Ces approches doivent être transposables à d'autres secteurs, par exemple à tout ce qui peut être mesuré en biologie [37] ou pourquoi pas à la variabilité d'un process industriel.

Bibliographie

1. Ministère de la Santé, R.F., *Arrêté du 26 novembre 1999 relatif à la bonne exécution des analyses de biolgie médicale (GBEA)* JO République Fr, 1999. **287(11Décembre)**: p. 18441.

2. N. *European Board of Veterinary Specialisation*. [cited; Available from: http://www.ebvs.org/.

3. N. *European College of Veterinary Clinical Pathology and European Society of Veterinary Clinical Pathology*. [cited; Available from: http://www.esvcp.org/.

4. Grasbeck, R. and N.E. Saris, *Establishment and use of normal values.* Scand J Clin Lab Invest, 1969. **26 (Supp 110)**: p. 62-63.

5. Siest, G. and M. Vernet-Nyssen, *Le concept de valeurs de référence en biologie clinique. Commission valeurs de référence. Document A.* Annales de Biologie Clinique, 1981. **39**(6): p. 381-384.

6. Brétaudière, J.P., et al., *Langage et principes statistiques pour les valeurs de référence. Commission valeurs de référence. Document B.* Annales de Biologie Clinique, 1979. **37**(2): p. 119-124.

7. Brétaudière, J.P. and A. Albert, *Influence des facteurs analytiques sur les valeurs de référence. Commission valeurs de référence. Document C.* Annales de Biologie Clinique, 1985. **43**: p. 306-309.

8. Siest, G. and M. Vernet-Nyssen, *Variations biologiques des examens de laboratoire. Commission valeurs de référence. Document D.* Annales de Biologie Clinique, 1979. **37**: p. 229-239.

9. Drosdowsky, M. and C. Sachs, *Facteurs à prendre en considération pour le prélèvement sanguin en vue de l'établissement des valeurs de référence. Commission valeurs de référence. Document E.* Annales de Biologie Clinique, 1980. **38**(4): p. 251-265.

10. Siest, G. and J. Henny, *Production de valeurs de référence de sujets sains. Commission valeurs de référence. Document G.* Annales de Biologie Clinique, 1981. **39**(4): p. 235-244.

11. Albert, A., R. Gueguen, and C. Sachs, *Traitement des valeurs de référence et détermination de l'intervalle de référence. Commission valeurs de référence. Document H.* Annales de Biologie Clinique, 1983. **41**: p. 63-69.

12. Albert, A., R. Guéguen, and C. Sachs, *Présentation des valeurs observées par rapport aux valeurs de référence. Commission valeurs de référence. Document I.* Annales de Biologie Clinique, 1983. **41**(3): p. 225-231.

13. CLSI, *Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline. Third edition.* Vol. C28-A3. 2008, Wayne, PA: CLSI.

14. Sunderman, F.W., Jr., *Current Concepts of "Normal Values," "Reference Values," and "Discrimination Values" in Clinical Chemistry.* Clin Chem, 1975. **21**(13): p. 1873-1877.

15. N., *Constitution of the World Health Organization*. Vol. http://www.searo.who.int/LinkFiles/About_SEARO_const.pdf, adressed 20 06 07. 1946, New York: United Nations. 19.

16. Solberg, H.E., *Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values.* J Clin Chem Clin Biochem, 1987. **25**: p. 337-342.

17. PetitClerc, C. and H.E. Solberg, *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**: p. 639-644.

18. Solberg, H.E. and C. Petitclerc, *Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of refence values.* J Clin Chem Clin Biochem, 1988. **26**: p. 593-598.

19. Solberg, H.E. and D. Stamm, *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**(8): p. 531-535.

20. Solberg, H.E., *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**: p. 645-656

21. Dybkaer, R. and H.E. Solberg, *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**: p. 657-662.

22. CLSI, *Method comparision and bias estimation using patient samples; approved guideline - second edition (interim revision) EP 09-A2-IR*. 2010, Wayne, PA: CLSI.

23. CLSI, *Defining, establishing, and verifying reference intervals in the clinical laboratory; provisional guideline. Third edition.* Vol. C28-P3. 2007, Wayne, PA: CLSI.

24. Fushuku, S., et al., *Reference values and limited serological survey for the Iriomote cat in Japan.* J Wildl Dis, 2001. **37**(3): p. 653-6.

25. *Analyse-it for Microsoft Excel*. 2009. p. Analyse-it Software, Ltd.

26. Solberg, H.E., *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**(3): p. 247-56.

27. *StatisPro*.

28. Feeman, W.E., 3rd, C.G. Couto, and T.L. Gray, *Serum creatinine concentrations in retired racing Greyhounds.* Vet Clin Pathol, 2003. **32**(1): p. 40-2.

29. Snow, D.H., R.C. Harris, and E. Stuttard, *Changes in haematology and plasma biochemistry during maximal exercise in Greyhounds.* Vet Record, 1988. **123**: p. 487-489.

30. Lassen, E.D., A.M. Craig, and L.L. Blythe, *Effects of racing on hematologic and serum biochemical values in greyhounds.* J Am Vet Med Assoc, 1986. **188**(11): p. 1299-303.

31. Ilkiw, J.E., P.E. Davis, and D.B. Church, *Hematologic, biochemical, blood-gas, and acid-base values in greyhounds before and after exercise.* Am J Vet Res, 1989. **50**(4): p. 583-6.

32. LaVecchio, D., et al., *Serum cardiac troponin I concentration in retired racing greyhounds.* J Vet Intern Med, 2009. **23**(1): p. 87-90.

33. Rand, J.S., et al., *Acute stress hyperglycemia in cats is associated with struggling and increased concentrations of lactate and norepinephrine.* J Vet Intern Med, 2002. **16**(2): p. 123-32.

34. N. *American Society for Veterinary Clinical Pathology*. [cited; Available from: http://www.asvcp.org/.

35. Hansen, A.M., A.H. Garde, and N.H. Eller, *Estimation of individual reference intervals in small sample sizes.* Int J Hyg Environ Health, 2007. **210**(3-4): p. 471-8.

36. Williams, D.A. and R.M. Batt, *Sensitivity and specificity of radioimmunoassay of serum trypsin-like immunoreactivity for the diagnosis of canine exocrine pancreatic insufficiency.* J Am Vet Med Assoc, 1988. **192**(2): p. 195-201.

37. Bavegems, V., et al., *Electrocardiographic reference values in whippets.* Vet J, 2009. **182**(1): p. 59-66.

Reference intervals describe the variations of biological markers and they are commonly used by clinicians to interpret test results. They are determined in healthy individuals and take into account variation factors such as age or sex. Groups of experts are regularly updating the recommendations on “how to determine reference intervals?” The aim of this work was to test the latest international recommendations using practical examples and to test new approaches by two different ways: firstly working on small reference sample groups (determining the reference interval *de novo* or transferring a pre-existing reference interval), and secondly working on reference sample groups that are “contaminated” by non healthy individuals’ test results. An easy-to-use freeware has also been created.

Keywords: reference interval, reference value, validation procedure, indirect method, nonparametric method, robust method.

AUTEUR : GEFFRÉ Anne

TITRE : NOUVELLES APPROCHES DE LA PRODUCTION D'INTERVALLES DE REFERENCE DE POPULATION

DIRECTEURS DE THESE : BRAUN Jean-Pierre & TRUMEL Catherine

LIEU ET DATE DE SOUTENANCE :
ECOLE NATIONALE VETERINAIRE DE TOULOUSE, 1er JUILLET 2011

Les intervalles de référence décrivent les variations des marqueurs biologiques et sont utilisés quotidiennement par les cliniciens pour interpréter les résultats d'analyses. Ils sont déterminés chez des sujets supposés en bonne santé en tenant compte de nombreux facteurs de variations comme l'âge ou le sexe. Leur détermination fait l'objet de recommandations internationales et l'objectif de ce travail a été d'en tester l'application dans des cas concrets puis d'évaluer de nouvelles méthodes en suivant principalement deux pistes : d'abord en travaillant sur des échantillons de référence de petite taille (soit en déterminant un intervalle de référence *de novo*, soit en transférant un intervalle de référence préexistant), ensuite en travaillant sur des échantillons de référence « pollués » par des valeurs issues d'individus non sains. Un outil informatique permettant de répondre rapidement et simplement aux questions précédentes a également été mis au point et est aujourd'hui disponible gratuitement.

Mots-clés : Intervalle de référence, valeur de référence, procédure de validation, méthode indirecte, méthode non paramétrique, méthode robuste.

DISCIPLINE ADMINISTRATIVE :
PATHOLOGIE, TOXICOLOGIE, GENETIQUE & NUTRITION

INTITULE ET ADRESSE DU LABORATOIRE :
UMR 1331 TOXALIM – INRA
Ecole Nationale Vétérinaire de Toulouse, Laboratoire de Physiologie, 23 Chemin des Capelles BP 87614, 31076 TOULOUSE cedex 3 - France