*J Vet Clin* 24(4) : 509-513 (2007)

Journal of Veterinary Clinics
www.ksvc.or.kr

# Evaluation of Reference Intervals of Some Selected Chemistry Parameters using Bootstrap Technique in Dogs

**Eu-Tteum Kim and Son-Il Pak[1]**

*School of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chunchon 200-701, Korea*

(Accepted: December 3, 2007)

**Abstract :** Parametric and nonparametric coupled with bootstrap simulation technique were used to reevaluate previously defined reference intervals of serum chemistry parameters. A population-based study was performed in 100 clinically healthy dogs that were retrieved from the medical records of Kangwon National University Animal Hospital during 2005-2006. Data were from 52 males and 48 females (1 to 8 years old, 2.2-5.8 kg of body weight). Chemistry parameters examined were blood urea nitrogen (BUN) (mg/dl), cholesterol (mg/dl), calcium (mg/dl), aspartate aminotransferase (AST) (U/L), alanine aminotransferase (ALT) (U/L), alkaline phosphatase (ALP) (U/L), and total protein (g/dl), and were measured by Ektachem DT 60 analyzer (Johnson & Johnson). All but calcium were highly skewed distributions. Outliers were commonly identified particularly in enzyme parameters, ranging 5-9% of the samples and the remaining were only 1-2%. Regardless of distribution type of each analyte, nonparametric methods showed better estimates for use in clinical chemistry compare to parametric methods. The mean and reference intervals estimated by nonparametric bootstrap methods of BUN, cholesterol, calcium, AST, ALT, ALP, and total protein were 14.7 (7.0-24.2), 227.3 (120.7-480.8), 10.9 (8.1-12.5), 25.4 (11.8-66.6), 25.5 (11.7-68.9), 87.7 (31.1-240.8), and 6.8 (5.6-8.2), respectively. This study indicates that bootstrap methods could be a useful statistical method to establish population-based reference intervals of serum chemistry parameters, as it is often the case that many laboratory values do not confirm to a normal distribution. In addition, the results emphasize on the confidence intervals of the analytical parameters showing distribution-related variations.

***Key words*** : reference interval, serum chemistry, simulation, bootstrap, dog

## Introduction

The reference intervals are prerequisite to practitioners in constructing a differential diagnosis, confirming a presumptive diagnosis or managing patients by proper interpreting clinical or laboratory observations such as clinical chemistry, hematology and physiology. To accomplish these purposes efficiently, high degree of confidence in them is required, particularly for life-or-death decision making. Some important criteria must be considered to establish reference intervals for any laboratory tests and their useful application in clinical medicine: true health condition of subjects, size of subjects, factors that might affect the analytes of interest, random and independent sampling of subjects from the population and suitable statistical methods (20,24,28,31,34). The National Committee for Clinical Laboratory Standards (NCCLS) (24) and the International Federation of Clinical Chemistry (IFCC) (34) provided a guideline to determine the components of variations in establishing the intervals.

From a statistical point of view, one of the most prevalent methods for establishing reference intervals involves the estimation of central 95% interval defined by the lower and upper reference limits for the subset of individuals sampled. The parametric intervals can be calculated based on the mean and $\pm 2$ standard deviation (SD) from the mean. This 2SD method assuming the distribution of values is Gaussian distribution can result in misclassification of data when the distribution is nongaussian (7,28). If the distribution is a skewed or asymmetrical form, the values are transformed to produce a more Gaussian distribution and the interval is determined (5,6,14,25,29,32,33,35). Alternatively, nonparametric percentile method which does not make any assumption of distribution type nor do they use estimates of distribution parameters have been proposed as a method of choice (28,34,35). However, practical problems related to the large variance when sufficient dataset is not available have provided motivation for developing better methods (13). In together with these methods, bootstrap method, weighted percentile method, smoothing technique and simulation method have been proposed (16,17,20,31). Although statistical methods for establishment of new reference intervals are well described the evaluation of clinical usefulness of established intervals has not received much attention in veterinary medicine. Therefore, the authors attempted to reevaluate the established reference intervals that should be relevant for some selected clinical chemistry components by using different statistical techniques.

[1]Corresponding author.
E-mail : paksi@kangwon.ac.kr

## Materials and Methods

### Data

From 2005 to 2006, a total of 100 clinically healthy dogs (52 males, 48 females) weighing 2.2-5.8 kg (9 breeds including mixed) without underlying illness were retrieved from the laboratory database of Kangwon National University Animal Hospital. Dogs with suspected illness that might affect chemistry components under study or < 1 year or > 8 years of age were excluded in the analysis. Records including patient's signalment, clinical signs on initial visit, historical and physical examination findings, information on concurrent diseases, and outcome were also reviewed for inclusion. In this study the following 7 analytes were examined: blood urea nitrogen (BUN) (mg/dl), cholesterol (mg/dl), calcium (mg/dl), aspartate aminotransferase (AST) (U/L), alanine aminotransferase (ALT) (U/L), alkaline phosphatase (ALP) (U/L), and total protein (g/dl). All analytes were measured using an automated analyzer (Ektachem DT 60, Johnson & Johnson) according to the manufacturer's recommendations.

### Parametric and nonparametric percentile methods

To estimate confidence intervals parametric 2SD method or two nonparametric percentile methods were used depending on the distribution of the values. Initially, goodness-of-fit for normality of a component was determined by Kolmogorov-Smirnov (KS) test (16). For nonparametric method, the lower and upper limits were computed according to the guidelines of NCCLS and IFCC (24,34). Briefly, the values were initially ranked in ascending order according to size. Then 2.5 and 97.5 percentile values were calculated 0.025 (n + 1) and 0.975(n + 1), respectively, where n is number of values. Alternatively, for the 2.5 and 97.5 percentiles this corresponds to 0.025n + 0.5 and 0.975n + 0.5 ordered values, respectively (9). The results of the first method only were presented here since the two nonparametric methods yield the same results. The analysis was performed using the SAS statistical package (SAS, Cary, NC) and Microsoft® Excel (Microsoft, Seattle, WA).

### Bootstrap method

The above-mentioned nonparametric calculations were also performed using parametric and nonparametric bootstrap procedures (8,11,12). Briefly, each of the values was assigned the same probability of being reselected and for each set of resampled values, and then percentile estimates were computed. This bootstrap estimate was obtained as the means of percentiles by 2,000 times iterations the above procedures. All negative estimates were assigned to zero for meaningful interpretation of the results. This was performed using Microsoft® Excel and its add-on software, @Risk (Palisade, Newfield, NY).

## Results

The descriptive statistics and tests for assumption of normality for each analyte are presented in Table 1. All analytes but calcium were highly skewed to the right by the KS goodness-of-fit normality test. Only 1-2% of the samples for BUN, cholesterol, calcium and total protein were identified as outliers, but for 5-9% of the remaining enzyme parameters. The simulated means and intervals (or percentiles) are shown in Table 2, and the computed intervals were generally close to the values recommended by manufacturer with some variations depending on the analytes. Regardless of the distribution type of each analyte, nonparametric bootstrap method revealed better estimates with regard to positive lower limits of confidence intervals.

## Discussion

Even the basic concepts of reference intervals are well described elsewhere (21,22,32,35), clinical interpretation of chemistry parameters in the practical world can be subjected to increased risk when there are potentially large variations in the intervals. In this study, the reference intervals estimated by nonparametric methods were rather close to those recommended by manufacturer but the lower or upper limits in some analytes need to be reconsidered for clinical application. Six components were rightward skewed, indicating that the blood levels of these components in healthy dogs have a tendency to the higher levels. Several reasons could be stated to explain the difference, such as the different inclusion criteria for healthy subjects, different population at different time, different sample size, and by different users even in the same analytical methods. At this time it is not clear that these differences may truly reflect the real changes in the general

**Table 1.** Descriptive statistics and testing for normality (Kolmogorov-Smirnov test) of the analytes in 100 clinically healthy dogs

| Analyte | Minimum | Maximum | Mean | Median | SD | IQR[a] | % outlier | Normality |
|---|---|---|---|---|---|---|---|---|
| BUN (mg/dl) | 3.4 | 29.4 | 14.7 | 14.9 | 4.9 | 7.4 | 1 | 0.0166 |
| Cholesterol (mg/dl) | 94.0 | 462.0 | 218.4 | 204.5 | 79.1 | 112.5 | 2 | 0.0100 |
| Calcium (mg/dl) | 7.3 | 13.1 | 10.4 | 10.2 | 1.1 | 1.4 | 1 | 0.1500 |
| AST (U/L) | 7.0 | 104.0 | 25.5 | 24.0 | 13.3 | 9.0 | 6 | 0.0100 |
| ALT (U/L) | 1.0 | 651.0 | 54.6 | 46.0 | 67.5 | 23.8 | 5 | 0.0100 |
| ALP (U/L) | 31.0 | 288.0 | 88.0 | 73.0 | 50.5 | 45.8 | 9 | 0.0100 |
| Total protein (g/dl) | 5.2 | 8.6 | 6.8 | 6.8 | 0.6 | 0.8 | 2 | 0.0409 |

[a]Interquartile range.

**Table 2.** Comparison of sample statistics (95% confidence interval, CI) and reference intervals recommended by manufacturer

| Analyte | Parametric | Nonparametric | Bootstrap (2.5 and 97.5 percentile) | | Manufacturer's CI |
|---|---|---|---|---|---|
| | | | Parametric | Nonparametric | |
| BUN (mg/dl) | 14.8 (4.8, 24.7) | 14.9 (7.0, 24.1) | 14.8 (5.1, 24.9) | 14.7 (7.0, 24.2) | (4.5, 30.5) |
| Cholesterol (mg/dl) | 228.7 (0, 521.9) | 204.0 (121.0, 427.0) | 228.7 (0, 518.8) | 227.3 (120.7, 480.8) | (126, 359) |
| Calcium (mg/dl) | 10.4 (8.1, 12.6) | 10.2 (8.2, 12.4) | 10.9 (8.2, 12.5) | 10.9 (8.1, 12.5) | (9.7, 12.2) |
| AST (U/L) | 25.5 (0, 52.3) | 24.0 (12.0, 68.0) | 25.5 (0, 51.6) | 25.4 (11.8, 66.6) | (1, 37) |
| ALT (U/L) | 54.6 (0, 190.2) | 46.0 (10.0, 167.0) | 25.5 (0, 66.6) | 25.5 (11.7, 68.9) | (3, 50) |
| ALP (U/L) | 88.0 (0, 189.5) | 72.0 (31.0, 239.0) | 88.0 (0, 240.8) | 87.7 (31.1, 240.8) | (20, 155) |
| Total protein (g/dl) | 6.8 (5.5, 8.1) | 6.8 (5.6, 8.2) | 6.8 (5.6, 8.1) | 6.8 (5.6, 8.2) | (4.8, 6.6) |

* All negative estimates were assigned to zero for meaningful interpretation of the results.

population, but the variations need to be considered to redefine the confidence intervals in the future setting.

One of the important factors to derive reliable reference intervals is the sample size requirement because intervals constructed with fewer individuals frequently result in weak intervals that are often questioned by clinical observations. In human medicine, a minimum size of 120 for parametric (28,32) or 200 individuals for nonparametric method (17,21) has been suggested previously. In veterinary medicine, since obtaining such subjects is often not accomplished, a more realistic number of 40 (1) or 60 animals (23) were reported to be sufficient if a Gaussian distribution is present.

The bootstrap simulation technique has increasingly being employed in the medical literatures such as disease surveillance and modeling (27), reproductive performance (10), vaccine potency testing (2), molecular biology (18) and others (3, 19). This method essentially consists of repeated random sampling of the original values with replacement to model the unknown variables, and to quantify the sampling error, consequently confidence interval of a mean or a median of interest (8,11). It has been known to be useful especially when the size of data set is small or non-Gaussian, and the confidence intervals calculated by the bootstrap are not subject to a restriction of symmetry, making them much more versatile (15,26,30). It generally accepted that 200 iterations are sufficient to estimate standard error, although much larger values are necessary depending on the data (4,11,12). In this study, 2,000 samples were drawn with replacement, yielding 2,000 reference intervals. From these values, the observed 2.5 and 97.5 percentiles were used as a 95% confidence interval. Linnet (20) showed that the bootstrap procedures had the lowest root mean squared error of the procedures examined for all sample sizes in a simulation study. In addition, the author recommended that a sample size of at least 100 should be considered for nonparametric reference interval estimation, as showing that the differences between the procedures were most pronounced at low sample sizes of less than 40.

It is clear that there are many clinical situations to derive a reference interval for a clinical analyte where laboratory instrument is superseded by new one, reagent sources are changed, the number of observed data values is small, characteristics of the population has changed, and outlier values are frequently identified in serum chemistry parameters as seen in this study. Reference intervals recommended by manufacturer can be used as a guideline derived from specific study populations where the instrumentations are tested so that laboratory users need to establish their own reference intervals to reveal the changing patterns in the population, for instance, ageing process. Furthermore, verification of a previously established interval may be necessary for communicating or transferring the values within and between laboratories (22). This will be the basic requirement for patient care.

## Acknowledgments

This study was supported by a grant (No. 0904001-1-1) from the High-Technology Development Project for Agriculture and Forestry, Ministry of Agriculture and Forestry, Republic of Korea.

## References

1. Archer J. Interpretation of laboratory data. In: Villiers E, Blackwood L (eds.). BSAVA manual of canine and feline clinical pathology. 2nd ed. British Small Animal Veterinary Association. UK. 2005: 11-22.
2. Barnett PV, Statham RJ, Vosloo W, Haydon FT. Foot-and-mouth disease vaccine potency testing: determination and statistical validation of a model using serological approach. Vaccine 2003; 21: 3240-3248.
3. Birsan M, Molnar P, Burlando P, Pfaundler M. Streamflow

trends in Switzerland. J Hydrol 2005; 314: 312-329.
4. Booth JG, Sarkar S. Monte Carlo approximation of bootstrap variances. Am Stat 1998; 52: 354-357.
5. Boyd JC, Lacher DA. A multi-stage Gaussian transformation algorithm for clinical laboratory data. Clin Chem 1982; 28: 1735-1741.
6. Boyd JC, Lacher DA. The multivariate reference range: an alternative interpretation of multi-test profiles. Clin Chem 1982; 28: 259-265.
7. Brunden MN, Clark JJ, Sutter ML. A general method of determining normal ranges applied to blood values for dogs. Am J Clin Path 1970; 53: 332-339.
8. Diaconis P, Efron B. Computer-intensive methods in statistics. Sci Am 1983; 248: 96-108.
9. Diem K, Seldrup J, Lentner C. Geigy scientific tables. Introduction to statistics, statistical tables and mathematical formulae-percentiles. 8th ed. Ciba-Geigy Basel. 1982: 197.
10. Dohoo IR, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island. Prev Vet Med 2001; 50: 127-144.
11. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986; 1: 54-77.
12. Efron B, Tibshirani R. Statistical data analysis in the computer age. Science 1991; 253: 390-395.
13. Elveback LR, Taylor WF. Statistical methods of estimating percentiles. Ann NY Acad Sci 1969; 161: 538-548.
14. Harris EK, DeMets DL. Estimation of normal ranges and cumulative proportions by transforming observed distributions to Gaussian form. Clin Chem 1972; 18: 605-612.
15. Henderson AR. The bootstrap: A technique for data-driven statistics using computer-intensive analyses to explore experimental data. Clinica Chimica Acta 2005; 359: 1-26.
16. Holmes EW, Kahn SE, Molnar PA, Bermes EW. Verification of reference ranges by using a Monte Carlo sampling technique. Clin Chem 1994; 40: 2216-2222.
17. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. Clin Chem 1998; 44: 622-631.
18. Law AE, Gale KR, Minchin CM, Shkap V, Waal D. Phylogenetic analysis of the erythrocytic *Anaplasma* species based on 16S rDNA and GroEL (HSP60) sequences of *A. marginale*, *A. centrale*, and *A. centrale* vaccine strain. Vet Microbiol 2003; 92: 145-160.
19. Lesnoff M, Lancelot R, Tillard E, Dohoo IR. A steady-state approach of benefit-cost analysis with a periodic Leslie-matrix model: presentation and application to the evaluation of a sheep-disease preventive scheme in Kolda, Senegal. Prev Vet Med 2000; 46: 113-128.
20. Linnet K. Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. Clin Chem 2000; 46: 867-869.
21. Lott JA, Mitchell LC, Moeschberger ML, Sutherland DE. Estimation of reference ranges: how many subjects are needed? Clin Chem 1992; 38: 648-650.
22. Lumsden JH. Normal or reference values: questions and comments. Vet Clin Pathol 1998; 27: 102-106.
23. Lumsden JH. Reference values. In: Feldman BF, Zinkl JG, Jain NC (eds.). Schalm's veterinary hematology. 5th ed. Philadelphia: Lippincott Williams & Wilkins. 2000: 12-15.
24. NCCLS (National Committee for Clinical Laboratory Standards). How to define and determine reference intervals in the clinical laboratory: approved guideline. NCCLS document C28-A2, 2nd ed. Villanova, PA, 2000.
25. Naus AJ, Borst A, Kuppens PS. Determination of n-dimensional reference ellipsoids using patient data. J Clin Chem Clin Biochem 1982; 20: 75-80.
26. Oviedo M, Muònoz P, Domßnguez A, Carmona G. Estimated Incidence of Hepatitis A Virus Infection in Catalonia. Ann Epidemiol 2006; 16: 812-819.
27. Pawitan Y, Griffin JM, Collins JD. Analysis and prediction of the BSE incidence in Ireland. Prev Vet Md 2004; 62: 267-283.
28. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. Clin Chem 1971; 17: 275-284.
29. Reed AH, Wu GT. Evaluation of a transformation method for estimation of normal range. Clin Chem 1974; 20: 576-581.
30. Rodgers JL. The bootstrap, the jacknife, and the randomization test: a sampling taxonomy. Mult Behavior Res 1999; 34: 441-456.
31. Shultz EK, Willard KE, Rich SS, Connelly DP, Critchfield GC. Improved reference-interval estimation. Clin Chem 1985; 31: 1974-1978.
32. Solberg HE. Statistical treatment of reference values. Bull Mol Biol Med 1983; 8: 13-19.
33. Solberg HE. Statistical treatment of reference values in laboratory medicine: testing the goodness-of-fit of an observed distribution to the Gaussian distribution. J Clin Lab Med 1986; 46 (supp 184): 125-132.
34. Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert panel 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.
35. Solberg HE, Gräsbeck R. Reference values. Adv Clin Chem 1989; 27: 1-79.

# Bootstrap 기법을 이용한 개의 혈청검사 일부 항목의 참고범위 평가

김으뜸 · 박선일[1]

강원대학교 수의학부대학 임상병리학교실

**요 약** : 혈청검사항목의 해석기준으로 사용하는 참고범위는 측정 장비와 병원마다 차이를 보이기 때문에 병원 간 정보를 교환하고 해석하는데 어려움이 많다. 또한 동일한 병원에서도 내원한 환자의 특성을 고려하여 참고범위를 재설정하는 것이 일반모집단의 특성을 제대로 반영한다. 본 연구에서는 강원대학교 수의학부대학 동물병원에서 설정한 혈청화학 검사 항목의 참고범위를 재평가하기 위하여 2005-2006년 동안 본원에 내원한 임상적으로 건강한 개 100두 (1-8세, 체중 2.2-5.8 kg)의 혈청검사 일부 항목을 모수 및 비모수적 bootstrap 모의시험으로 분석하였다. 평가항목은 BUN (mg/dl), cholesterol (mg/dl), calcium (mg/dl), aspartate aminotransferase (AST, U/L), alanine aminotransferase (ALT, U/L), alkaline phosphatase (ALP, U/L) 및 total protein (g/dl)으로 Ektachem DT 60 분석기 (Johnson & Johnson)로 측정하였다. 칼슘을 제외한 모든 항목이 왜곡이 매우 심한 분포를 보였으며 특히 혈청 효소항목의 outlier는 전체 자료의 5-9%, 기타 항목은 1-2%를 보였다. 각 항목의 분포에 상관없이 모수적 방법에 비하여 비모수적 방법으로 추정한 참고범위가 임상적으로 유용하였으며 추정된 참고범위는 BUN 14.7 (7.0-24.2), cholesterol 227.3 (120.7-480.8), calcium 10.9 (8.1-12.5), AST 25.4 (11.8-66.6), ALT 25.5 (11.7-68.9), ALP 87.7 (31.1-240.8), and total protein 6.8 (5.6-8.2)로 나타났다. 이러한 결과는 모집단의 특성을 고려하여 참고범위를 재설정하는데 비모수적 모의시험이 매우 유용하며 특히 측정항목의 분포에 무관하게 사용할 수 있는 장점이 있는 것으로 사료된다.

**주요어** : 참고범위, 혈청화학, 모의시험, bootstrap, 개