Evaluation of an in-house enzyme-linked immunosorbent assay for quantitative measurement of serum total thyroxine concentration in dogs and cats

Jill C. Lurye, DVM; Ellen N. Behrend, VMD, PhD, DACVIM; Robert J. Kemppainen, DVM, PhD

Objective—To compare serum total thyroxine (T₄) concentrations obtained with an in-house ELISA and a validated radioimmunoassay (RIA).

Design—Laboratory trial.

Sample Population—50 canine and 50 feline serum samples submitted for measurement of total T₄ concentration with the RIA; samples were selected to represent a wide range of concentrations (≤6 to 167 nmol/L).

Procedure—Results of the ELISA and RIA were compared by calculating correlation coefficients, examining linearity, determining bias and precision, and evaluating clinical interpretations.

Results—Correlation coefficients for results of the 2 methods were 0.84 for the canine samples and 0.59 for the feline samples. Examination of bias plots revealed large variations in ELISA results, compared with RIA results. For the feline samples, the ELISA consistently overestimated total T₄ concentration obtained with the RIA. When results of the 2 methods were categorized (low, borderline low, normal, borderline high, or high), results were discordant for 24 (48%) and 29 (58%) of the canine samples and for 18 (36%) and 28 (56%) of the feline samples (depending on whether borderline high ELISA results were considered normal or high). Reliance on results of the ELISA would have led to inappropriate clinical decisions for 31 (62%) canine samples and 25 (50%) feline samples. The ELISA coefficients of variation for the pooled canine and feline samples were 18 and 28%, respectively.

Conclusions and Clinical Relevance—Substantial discrepancies between ELISA and RIA results for T₄ concentrations were detected. Thus, we concluded that the in-house ELISA kit was not accurate for determining serum total T₄ concentrations in dogs and cats. (J Am Vet Med Assoc 2002;221:243–249)

Hypothyroidism in dogs and hyperthyroidism in cats are 2 of the most frequently encountered endocrinopathies in small animal medicine. In dogs, hypothyroidism most commonly results from a loss of functional thyroid tissue (primary hypothyroidism) secondary to lymphocytic thyroiditis, idiopathic atrophy, or, much less commonly, neoplastic infiltration. In cats, hyperthyroidism is a result of a functional thyroid adenoma or, more rarely, a functional adenocarcinoma.

Various methods of testing for thyroid disease have been devised, including measurement of serum total and free thyroxine (T₄) and triiodothyronine (T₃) concentrations and serum thyrotropin (TSH) concentration, as well as several stimulation and suppression tests. Nonthyroidal illness, however, may have a substantial impact on results of thyroid function tests, complicating the diagnosis of thyroid disorders. Thus, although a diagnosis of thyroid disease is often made on the basis of clinical signs and results of thyroid function tests, definitive diagnosis of thyroid disease in veterinary patients can be difficult.

Initial screening of dogs and cats for thyroid disease and monitoring of dogs and cats undergoing treatment for thyroid disease often involves measurement of serum total T₄ concentration. A radioimmunoassay (RIA) incorporating radiolabeled hormone is considered to be the gold standard for measurement of total T₄ concentration and is the generally accepted reference method. The RIA is a simple, rapid, and sensitive method of measuring T₄ concentration, but the use of radioactive isotopes requires appropriate licensing, storage facilities, safety devices, and laboratory training. Thus, most veterinary practices do not perform such testing in-house and send samples to a commercial laboratory for testing. Although this increases the time and, usually, the cost of testing, veterinarians can have a high level of confidence in the accuracy of results from these laboratories. An in-house method for measuring serum total T₄ concentration could be beneficial if results were immediately available. However, because many decisions regarding the management of veterinary patients are made on the basis of laboratory tests results, an in-house test must not only be economical, quick, and easy to perform but must also provide reliable results. The purpose of the study reported here was to compare serum total T₄ concentrations obtained with a commercially available ELISA kit’ with values obtained with a previously validated commercially available RIA.

Materials and Methods

Sample collection and analysis—Fifty canine and 50 feline serum samples submitted for measurement of total T₄ concentration to the Auburn University Endocrine Diagnostic Service were used in the study. For all samples, T₄
Concentration had been measured with a previously validated, commercially available RIA, and samples were chosen so that low, normal, high, and borderline concentrations were represented for both dogs and cats. As the manufacturer's instructions for the ELISA state that test results are not affected by hemolysis or lipemia, no specific selection process was used in regards to these sample characteristics.

Canine serum samples included 5 samples with total T₄ concentrations less than the laboratory's lower reference limit (< 11 nmol/L), 12 samples with borderline concentrations (12 to 19 nmol/L), 27 samples with concentrations within reference limits (20 to 55 nmol/L), and 6 samples with concentrations greater than the upper reference limit (> 55 nmol/L). Feline serum samples included 5 samples with total T₄ concentrations less than the laboratory's lower reference limit (< 10 nmol/L), 32 samples with concentrations within reference limits (10 to 50 nmol/L), including 18 samples with concentrations in the lower half of the reference range and 14 samples with concentrations in the upper half of the reference range, and 13 samples with concentrations greater than the upper reference limit (> 50 nmol/L). Samples were stored at −20°C until analyzed.

For the present study, serum total T₄ concentration of each sample was measured a second time with a commercially available in-house ELISA kit and analyzing equipment, following the manufacturer's directions. During analysis of samples with this ELISA, the analyzing equipment must be preset to indicate whether results are expected to be in a "low dynamic" (defined by the manufacturer as 6.4 to 45 nmol/L) or "high dynamic" (defined by the manufacturer as 23.7 to 90.0 nmol/L) range. The manufacturer recommends that all feline serum samples tested to screen cats for hyperthyroidism or to assess response to methimazole treatment be analyzed with the equipment preset to indicate that results are expected to be in the high dynamic range. To determine whether results of the ELISA could be improved when testing feline serum samples, an additional 20 feline samples were evaluated with the ELISA kit with the equipment preset to indicate that results were expected to be in the low dynamic range. Samples used included 2 with total T₄ concentrations, as determined with the RIA, less than the lower reference limit, 11 with concentrations within reference limits (6 with concentrations in the lower half of the reference range and 5 with concentration in the upper half of the reference range), and 7 with concentrations greater than the upper reference limit. All canine samples were tested as recommended by the manufacturer.

To determine test precision, 2 pooled serum samples were created, 1 by pooling 5 canine samples and 1 by pooling 5 feline samples. Thyroxine concentrations of the pooled canine and feline samples, as determined with the RIA, were 31 and 37 nmol/L, respectively. The pooled samples were tested 10 times each with the in-house ELISA kit.

Statistical analyses—Data were analyzed with standard statistical software. Concentrations reported by the ELISA kit to be < 90 nmol/L, the highest value reported, were arbitrarily assigned a concentration of 91 nmol/L for statistical purposes.

Concentrations reported as < 6.0 nmol/L by the RIA, the lowest value reported, were arbitrarily assigned a concentration of 5 nmol/L for statistical purposes. No sample concentrations measured with the RIA were greater than the upper limit of the standard curve.

Several feline samples analyzed with the testing equipment preset to the high dynamic range (as directed by the manufacturer) resulted in concentrations less than the lower limit of the testing range. These samples were reanalyzed with the equipment preset to the low dynamic range, and results were included in statistical analyses. No sample concentrations measured with the ELISA kit were less than the lower limit of the low dynamic range. Additionally, several of the 20 feline samples used in the second portion of the study resulted in concentrations higher than the upper limit of the testing range. These samples were reanalyzed with the equipment preset to the high dynamic range, and results were included in statistical analyses.

Results of the ELISA kit were compared with results of the RIA by means of least-squares linear regression. Additionally, linearity of the relationship between results of the 2 testing methods was evaluated with Passing-Bablok regression by use of theCUSUM test for linearity. Agreement between results of the 2 testing methods was also evaluated by constructing bias plots to create a visual assessment of the agreement of test results. For construction of these bias plots, the difference between concentration obtained with the RIA and concentration obtained with the ELISA kit was plotted against the concentration obtained with the RIA.

Clinical agreement between results of the 2 testing methods was assessed by categorizing results of the ELISA kit as low, borderline low, normal, borderline high, or high, according to ranges provided by the manufacturer, and results of the RIA as low, borderline low (canine samples only), normal, or high, according to established reference ranges of the testing laboratory. Percentages of discordant results were then calculated. Because results of the RIA were not categorized as borderline high, percentages of discordant results involving values obtained with the ELISA that were categorized as borderline high were calculated in 2 ways, first with borderline high results of the ELISA classified as normal and then with borderline high results of the ELISA classified as high. Similarly, results of the RIA for feline samples were not categorized as borderline low. Therefore, percentages of discordant results for feline samples involving values obtained with the ELISA that were categorized as borderline low were calculated first with borderline low results classified as normal and then with borderline low results classified as low.

Finally, numbers of ELISA results that would have led to incorrect clinical decisions (therapeutic or diagnostic) were also determined. For the canine samples, ELISA results were considered to have led to an incorrect clinical decision if hyperthyroidism would not have been identified in a dog suspected to have hyperthyroidism on the basis of results of the RIA, if results failed to suggest the need for follow-up or additional testing that would normally have been deemed appropriate on the basis of RIA results (eg, low or borderline T₄ concentration), if results suggested the need for additional diagnostic testing that would normally have not been deemed appropriate on the basis of RIA results, or if results indicated a need for an increase or decrease in thyroid supplementation that would normally be considered inappropriate on the basis of RIA results. For the feline samples, ELISA results were considered to have led to an incorrect clinical decision if results suggested a need for an increase or decrease in the dosage of methimazole that would normally be considered inappropriate on the basis of RIA results. For the canine samples, ELISA results were considered to have led to an incorrect clinical decision if results suggested a need for an increase or decrease in thyroid supplementation that would normally be considered inappropriate on the basis of RIA results. For evaluation of test precision,
standard statistical techniques were used to generate mean, SD, and coefficient of variation for results of the pooled samples.

**Results**

Correlation coefficients (R²) for results of the ELISA kit compared with results of the RIA were 0.84 for the canine samples and 0.59 for the feline samples. For both sets of samples, the fitted regression line appeared to differ significantly from the identity line (i.e., the line expected if results of the 2 test methods were exactly the same; Fig 1). Visual inspection of the scatterplots suggested that the relationship between results of the 2 test methods was nonlinear for both canine and feline samples, as large numbers of observations fell on both sides of the fitted regression line. Use of Passing-Bablok regression confirmed that the relationship was nonlinear.

Examination of the bias plots revealed large deviations in test results throughout the testing range, with several outliers for both canine and feline samples (Fig 2). On these plots, samples for which results of the ELISA kit were in agreement with results of the RIA would fall on the zero bias line. For the canine samples, values for individual samples fell above and below the zero bias line, indicating that the ELISA kit overestimated and underestimated the serum T4 concentration obtained with the RIA. By comparison, for most of the feline sam-

---

**Figure 1**—Scatterplots of total T4 concentrations obtained with a commercially available in-house ELISA and a previously validated radioimmunoassay (RIA). A—Scatterplot for 50 canine serum samples. B—Scatterplot for 50 feline serum samples. In each plot, the solid line represents the linear regression line (R² = 0.84 and 0.59 for canine and feline samples, respectively), and the dashed line represents the line of identity (i.e., the line obtained if results of the 2 methods had been equal).

**Figure 2**—Bias plots of total T4 concentrations obtained with a commercially available in-house ELISA and an RIA. A—Plot for 50 canine serum samples. B—Plot for 50 feline serum samples. Bias plots were created by plotting, for each sample, the difference between ELISA and RIA results versus the RIA result. In each plot, the zero bias line represents the line expected if results of the 2 methods had been equal, the bias line represents the direction and degree of any bias between results of the ELISA and results of the RIA, and the dotted-and-dashed lines represent the 95% agreement limits.
samples, values for individual samples fell above the zero bias line, indicating that the ELISA kit consistently overestimated the serum T4 concentration obtained with the RIA. In general, variations were much greater for the feline samples than for the canine samples.

For the canine samples, serum T4 concentrations measured with the ELISA kit were categorized as low, borderline low, normal, borderline high, or high. When results of the 2 methods were compared, results of the ELISA kit were found to be discordant with results of the RIA for 24 (48%) of the canine samples when borderline high ELISA results were categorized as normal and for 29 (58%) of the canine samples when borderline high ELISA results were categorized as high. Similarly, for the feline samples, serum T4 concentrations measured with the ELISA kit were categorized as low, normal, borderline high, or high (none of the samples yielded concentrations categorized as borderline low), and results of the RIA were categorized as low, normal, or high. When results of the 2 methods were compared, results of the ELISA kit were found to be discordant with results of the RIA for 18 (36%) of the feline samples when borderline high ELISA results were categorized as normal and for 28 (56%) of the feline samples when borderline high ELISA results were categorized as high.

Number of ELISA results that would have led to inappropriate clinical decisions was also determined. Results for 31 of 50 (62%) canine samples and 25 of 50 (50%) feline samples would have led to inappropriate clinical decisions.

Coefficients of variation for ELISA results obtained with pooled canine and feline samples were 18 and 28%, respectively.

For the 20 feline samples tested with the ELISA analyzing equipment preset to the low dynamic range, the correlation coefficient ($R^2$) for results of the ELISA kit, compared with results of the RIA, was 0.6. Visual examination of the scatterplot suggested that the relationship was nonlinear, which was confirmed with use of Passing-Bablok regression (Fig 3). Examination of a bias plot indicated that although overall bias was near 0, large deviations in test results throughout the testing range were evident; 1 outlier was observed. Samples with normal and low concentrations, as determined by use of the RIA, consistently fell above the bias line (ie, concentration was overestimated by the ELISA), whereas samples with high concentrations, as determined by use of the RIA, consistently fell below the bias line (ie, concentration was underestimated by the ELISA). When results were categorized, results of the ELISA kit were discordant with results of the RIA for 7 of 20 (35%) samples when borderline low ELISA results were categorized as normal and for 8 of 20 (40%) samples when borderline low ELISA results were categorized as low (none of the ELISA results were categorized as borderline high). Thirteen of 20 (65%) of the ELISA results would have led to an inappropriate clinical decision.

Discussion

Results of the present study revealed substantial discrepancies between serum total T4 concentrations obtained with the in-house ELISA and concentration obtained with the RIA. Thus, we conclude that the in-house ELISA kit is not an accurate method of determining serum total T4 concentrations in dogs and cats.

Although a moderate degree of correlation between results of the 2 testing methods was suggested by regression analysis, there are limitations to this statistical method of comparing test results. In particular, high correlation coefficients only indicate that results of the 2 methods are associated and do not indicate accuracy of results of the new method (ie, the ELISA kit) relative to results of the reference method (ie, the RIA).24 Furthermore, the correlation coefficient increases as the range of values measured increases,24 and T4 concentrations measured in this study covered a wide range (< 6 to 167 nmol/L). This limitation of the use of correlation coefficients for
comparing results of 2 testing methods was apparent in the present study, in that calculated correlation coefficients were much lower when a particular range of values was analyzed. For example, for canine samples with borderline concentrations (12 to 19 nmol/L), an important range when screening for hypothyroidism, the correlation coefficient was only 0.1, compared with a correlation coefficient of 0.84 when all canine samples were analyzed. Further evaluation of the regression data indicated that many observations deviated widely from the fitted regression line and that large numbers of observations appeared on both sides of the regression line. This suggests not only that there is a poor fit with the regression line but also that the relationship may not be linear, making the ELISA kit of limited or no practical value in predicting serum total T4 concentrations. The fact that the relationship was not linear was confirmed with Passing-Bablok regression by use of the CUSUM test for linearity.

Bias plots allow for a more straightforward interpretation of the difference between the results of 2 methods and may reveal differences that correlation coefficients do not. Bias plots graphically show the difference between values obtained with 2 test methods (a new method vs a reference method) as a function of the reference method. In this study, the RIA was the reference method, as it is considered to be the best estimate of the true value. With bias plots, results of the 2 methods can be assessed for agreement and bias. In addition, any relationship between imprecision and sample concentration can be identified. The zero bias line is a horizontal line that represents no difference between testing methods. Bias is present if results of the new test method are consistently higher or lower than results of the reference method, and the bias line on the plot demonstrates the degree and direction of this difference. A new test method that produces results that agree with results of the reference method without any bias would result in a plot on which the zero bias and bias lines would overlap and all data points would fall close to this horizontal line. In the present study, the distance of data points from the zero bias line indicated relatively large differences between serum T4 concentrations obtained with the ELISA kit and concentrations obtained with the RIA for a large portion of test samples. Substantial under- and overestimation of serum T4 concentrations was demonstrated.

The manufacturer of the ELISA equipment defines 2 dynamic test ranges, and the analyzer must be preset to 1 of these test ranges when a sample is assayed. Testing instructions recommend that the high dynamic range (25.7 to 90.0 nmol/L) be used when screening cats for hyperthyroidism, monitoring serum T4 concentrations in cats being treated with methimazole, and analyzing samples collected 4 to 6 hours after administration of L-thyroxine supplements to dogs. Instructions recommend that the low dynamic range (6.4 to 45 nmol/L) be used when screening dogs for hypothyroidism. Accordingly, the initial 50 feline samples tested in the present study were analyzed with the equipment preset to the high dynamic range. However, T4 concentration was overestimated for a large number of these samples, resulting in a substantial positive bias. Therefore, we attempted to determine whether results could be improved for feline samples by using the low dynamic range. However, even though the substantial positive bias was eliminated when the low dynamic range was used, an improvement in test accuracy was not observed, as data points were widely distributed on either side of the zero bias line. In general, when the low dynamic range was used, the ELISA kit overestimated T4 concentrations of samples with low or normal concentrations and underestimated T4 concentrations of samples with high concentrations. Given this and given the overlap between the 2 dynamic ranges and the fact that T4 concentration is not typically known prior to sample testing, there would appear to be no benefit in using the low dynamic range versus the high dynamic range when testing feline samples.

Evaluation of categorical data resulted in extremely discouraging results. The ELISA provided categorical results (ie, results categorized as low, borderline low, normal, borderline high, or high) that were different from those provided by the RIA for approximately half of the canine samples and for a third to a half of the feline samples, depending on whether borderline high ELISA results were considered normal or high and whether borderline low ELISA results were considered normal or low. Often, serum total T4 concentration is measured as a screening test for thyroid dysfunction. In addition, it is frequently used as a method for therapeutic monitoring. In these instances, laboratory results leading to inappropriate categorization of T4 concentration as low, normal, or high may result in an incorrect diagnosis or incorrect treatment. This was demonstrated by the large number of ELISA results that would have led to incorrect clinical decisions, compared with results of the RIA. Typically, measurement of serum total T4 concentration is recommended as a method for ruling out hypothyroidism in dogs and of monitoring the results of thyroid supplementation. Interpretation guidelines of the reference laboratory that performed the RIA in this study recommend that when screening dogs for hypothyroidism, serum total T4 concentrations < 19 nmol/L (ie, low or borderline-low values) be retested or tested with an alternative method. Dogs with serum total T4 concentrations between 20 and 55 nmol/L (ie, within the reference range) should be considered normal, and a diagnosis of hypothyroidism should be considered unlikely. Laboratory interpretive guidelines for samples from dogs receiving L-thyroxine supplementation suggest serum total T4 concentrations 4 to 12 hours after pill administration be between 30 and 60 nmol/L. On the basis of these guidelines, ELISA results would have led to incorrect clinical decisions for 31 of the 50 (62%) canine samples in the present study.

In cats, serum total T4 concentration is often measured to confirm a diagnosis of hyperthyroidism. When clinical and physical findings consistent with
hyperthyroidism are identified, treatment is often instituted on the basis of a high serum total T₄ concentration. However, a normal total T₄ concentration in a cat with appropriate clinical signs does not rule out the diagnosis, as cats with early or mild hyperthyroidism may have serum total T₄ concentrations in the upper half of the reference range. Current reference laboratory recommendations suggest that in cats, a serum total T₄ concentration, measured with the RIA, > 50 nmol/L is consistent with hyperthyroidism, and initiation or alteration of treatment is indicated. Retesting or additional testing is recommended when the concentration is in the upper half of the reference range (30 to 50 nmol/L) and clinical signs consistent with hyperthyroidism are present. On the basis of these recommendations, ELISA results would have led to incorrect clinical decisions for 25 of 50 (50%) feline samples analyzed with the equipment preset to the high dynamic range and for 13 of 20 (65%) feline samples analyzed with the equipment preset to the low dynamic range.

The degree of precision of the ELISA kit was also disappointing; coefficients of variation were 18 and 28% for canine and feline samples, respectively. Additional evidence of poor precision may be suggested by the number of outlier data points in the data sets. Although outlier data points cannot be definitively confirmed to be a result of poor test precision, other explanations are not readily identifiable. The higher precision during testing of canine, versus feline, samples may be related to species differences or other unknown variables. In general, intra- and interassay coefficients of variation < 6% have been recommended for measurement of serum total T₄ concentration in humans. Coefficients of variation reported for measurement of serum total T₄ concentration with various commercial RIA in veterinary medicine may be slightly higher, likely because of the wider variations in T₄ concentration regularly observed in veterinary patients, compared with human patients. Measurement of samples with lower T₄ concentrations is typically associated with higher coefficients of variation, compared with measurement of samples with higher T₄ concentrations. Reported intra- and interassay coefficients of variation for T₄ concentrations measured with a RIA reported in the veterinary literature range from 3.5 to 12.8%, with an average of approximately 5 to 8%. Intra- and interassay coefficients of variation for the RIA used in the present study have previously been reported to be approximately 5 and 10%, respectively.

References

25. Tummon I, Stemp J, Rose C, et al. Precision and method...


