The ability to form appropriate conclusions from experimental and observational studies depends on many factors including the suitability of the study design and appropriate analysis of data. Authors should ensure that the statistical analyses they perform are appropriate for the hypotheses being tested and the data being examined. A number of excellent textbooks are available as resources for study design, data presentation, statistical analysis and reporting of the results of these analyses in the biological sciences. Authors are strongly advised to consult with a biostatistician or epidemiologist when planning studies so that the study has optimal design and that the intended statistical analyses are appropriate for the hypotheses being tested.

Study design:
Designing the study is the most critical step in any investigation, and authors should note that a flawed study design is the most common reason for rejection of papers by journal editors. While true for all studies, clinical and epidemiological investigations (observational studies) can be even more challenging to appropriately design than classical experiments. Poor study design cannot be corrected through statistical analysis of the data and investigators should make a regular practice of consulting appropriate references on study design and statistical analysis when planning a new investigation. Consultation with individuals having a comprehensive understanding of the strengths and weaknesses of the various types of study designs, statistical analyses and the biological relationships under examination is essential during the design phase of the study. Epidemiologists with a strong interest in quantitative analysis and biostatisticians frequently fill this role.

A priori hypotheses (those formulated before the data are collected) should be stated, and clearly differentiated from hypotheses formulated after the data have been collected; just as primary study aims and outcomes should be clearly differentiated from secondary aims and outcomes. Bias in sample estimates will be reduced by formal random assignment of animals to groups and by masked (blind) evaluation of each outcome. Inability or failure to adhere to these principles must be noted in manuscripts along with a discussion of the potential impact on the results and validity of the study. Estimates of sample sizes required to achieve appropriate study power should be undertaken during the process of study design, not after the statistical analysis has been completed.

Reporting statistical methods and results:
Comprehensive recommendations for conducting and reporting results of statistical analyses can be found in numerous authoritative references (see below for a short list of suggested references). The following is a list of points for consideration that are specific to manuscripts submitted to the Journal of Veterinary Internal Medicine. These items address issues that are commonly identified during the review process.

1. Statistical software used for analysis should be cited as an endnote, including the version of the software used in the analysis.
2. Statistical methods for complex or unusual statistical methods beyond t-tests, correlation, chi-square, stratified analysis, analysis of variance, regression and survival analysis should be referenced.
3. Wherever possible, the data and main findings of the study should be summarized in the form of informative, carefully designed and labeled tables and graphs.
4. If the number of observations is small, all the data should be presented for example as a scatterplot in which values of each animal are depicted. Note that the Journal does not publish tables of values of variables for individual animals.
5. Appropriate descriptive statistics should be provided for data being analyzed.
   a. For a single continuous variable derived from a simple sample of individuals, these should include the following: the number of observations; a measure of central tendency (such as a mean or median); and a measure of variability (standard deviation [SD], range or interpercentile ranges [e.g. deciles, quartiles]), as appropriate for the data. The SD, rather than the standard error of the mean (SEM), should be used for describing variability among individuals or individual responses. The SEM or 95% confidence intervals are appropriate statistics for reporting the reliability of estimated parameters, including mean effects such as odds ratios.
   b. For categorical data, the numerator and denominator should be provided for each proportion /percentage in each group and category. Cut-points used to create categorical variables from continuous data need to be explained and justified. Calculation of proportions to summarize small samples (less than 20) is uninformative and should be avoided.
6. The number of significant figures in a value provides an indication of the reliability and precision of the measurement. Care should be taken when reporting values that an unreasonable precision is not implied by use of excessive significant figures. For instance, reporting serum creatinine concentration to 3 decimal places (1.175 mg/dl) implies
greater precision than is actually possible. Similarly, reporting white blood cell counts to 5 significant figures (11,575 cells/μl) suggests greater precision than is either achieved or relevant. Reporting of most values should be limited to 3 significant figures. A source for detailed instructions for reporting significant figures is provided in suggested reading material listed below.

7. All P values should be reported to two significant digits (e.g. P=0.032), whether or not the value is statistically significant. Values less than P=0.001 provide an exception to this rule, and instead should be reported as P<0.001. The practice of reporting P values as "less than" or "greater than" an arbitrary cut-point (e.g., 0.05) should be avoided, when possible.

8. When multiple hypotheses are evaluated, such as in the analysis of epidemiological studies, report the final and any competing models in tables that include: number of observations, measures of effect (e.g., odds ratios, relative risks, least square means), confidence intervals, and the P values.

9. Frequently clinical investigators test for differences between groups (e.g., treatment and control groups) with respect to a list of characteristics (e.g., a panel of hematological and biochemical tests) or risk factors. If this is done without having established an *a priori* hypothesis and using an approach such as multiple t-tests then the problem of "multiple testing" is encountered. A variety of strategies are available for dealing with this problem but it is preferable that the problem and the strategy are identified during the design phase of the study. Examples of appropriate strategies include, but are not limited to, controlling either for the family-wise type I error rate (such as with the method of Bonferroni) or for the false-discovery rate. Authors should consult with experienced data analysts or reference texts to ensure that their analytical methods have appropriately dealt with the issue of multiple comparisons.

10. If data transformation is performed to ensure that assumptions of statistical modeling are met, then an interpretation of the results of analysis in the context of the transformed data should be provided.

11. Analysis of repeated observations on an individual require the use of appropriate methods of statistical analysis that account for the correlation among observations.

12. Care must be taken in reporting or discussing results that do not achieve statistical significance (as defined by the *a priori* Type 1 error rate specified in the manuscript) to ensure that there is no implication or inference of a difference. The Journal does not permit use of terms such as "trend" to describe differences that approach but do not meet the type 1 error rate e.g. “there was a trend for fat cats being larger (p = 0.07)”, except when “trend” is used to describe a statistical difference over time (“there was a statistically significant (P = 0.011) trend in heart rate with duration of drug administration”). Similarly, terms such as “numerically but not statistically significantly higher” should not be used as they imply a difference when one does not exist. Rather, simply state the results and type 1 error rate (e.g. XX +/- YY and AA +/- BB, respectively, in groups 1 and 2, p = 0.055) and then discuss the reasons for the lack of a difference, if this is important to the study.

13. Randomized controlled trials (RCTs) in veterinary medicine often involve small sample sizes and may be underpowered, resulting in a type II (false negative) error. Manuscripts for RCTs should specify the primary outcome of interest, describe how the sample size was determined (including all elements used in calculations), and report confidence intervals (CIs) around observed treatment effects. Please consult “Type II error and statistical power in reports of small animal clinical trials” by Guiffrida et al (reference #7 in suggested reading) for additional information.

SUGGESTED READING
Both science and art are involved in the development of an optimal study design, data presentation and statistical analysis. There are many useful publications in these areas. The short list of references listed below should provide some guidance to novice and experienced investigators alike.


