Enhancing Clinical Decision Making

Systematic evaluation of scientific research for clinical relevance and control of bias to improve clinical decision making

Brad J. White, DVM, MS, and Robert L. Larson, DVM, PhD

Veterinary clinicians typically read the scientific literature to become better educated and thereby improve the quality of their clinical decisions. Whether this objective is met depends on the scientific validity of the studies evaluated and the similarity of the hypotheses tested to a specific clinical question. A focused approach to answering a specific clinical question or questions is important to avoid placing undue emphasis on preliminary data or extrapolating results beyond the populations to which they pertain. One approach to minimize potential misinterpretations and optimize time spent reading the literature is to use a strategic method of literature evaluation.

Veterinary clinicians have limited amounts of free time, and a streamlined method for selection and interpretation of the available literature on a given topic can increase the efficiency with which reported information is used to enhance clinical decision making. Clinicians learn strategies for performing thorough but efficient physical examinations by focusing on the problems most commonly encountered in practice, and a similar approach can be used to perform thorough but efficient literature evaluations. By focusing on the problems most commonly encountered in scientific research, clinicians can quickly identify research articles that can be immediately disregarded, thereby saving time and preventing the introduction of invalid information into the decision-making process. The purpose of the present article is to introduce the first 3 steps of a 5-step method for time-efficient literature evaluation designed to help clinicians obtain information to address a clinical question (**Figure 1**).

Determination of Clinical Relevance

The first step of a time-efficient approach to literature evaluation is to determine whether an identified research article is relevant to a particular clinical question and the current decision-making process. This process requires a clear understanding of the specific information needed to answer the clinical question as well as correct identification of the hypothesis tested in the reported study, given that hypotheses drive choice of study design, selection of outcome variables, and interpretation of the results.

One of the first things clinicians need to determine is whether the research described in an article is relevant to the current clinical scenario, which requires that clinicians develop a clear and concise clinical question.[1] Relevance can be evaluated by comparing the study animals and outcome of interest with those in the clinical scenario. Research articles should contain a clear, testable hypothesis that includes a description of the outcome variables that were measured in the study.[2] Research involving clinically meaningful outcome variables such as risk of disease onset, risk of death, or duration or quality of life can yield findings that can be directly applied to clinical scenarios, provided that the study in which they were evaluated was well designed and the study subjects were comparable to those in the clinical scenario (**Table 1**). Other research involving indirect measurements for outcome variables typically requires clinicians to make inferences about the implications of the findings for the animals they treat. For example, a study designed to evaluate the influence of a preventative measure on the likelihood of disease (outcome) might include indirect outcome measurements such as serum antibody titers or other physiologic or immunologic variables and not direct measurements such as presence or absence of clinical signs of disease. Although findings for such outcome variables might imply clinically important benefits (eg, high serum antibody titers suggest a high degree of immunity, which might infer a high degree of protection against disease), they should not be interpreted as proof of clinical benefits (eg, that disease is actually prevented). Consequently, clinicians should use caution when considering a treatment for which no data exist regarding directly measured clinical outcomes.[3]

Patient comparability is a term used to describe the relevance of the study population (ie, population from which the study animals were selected) to the patients of the clinical scenario (ie, clinical population). Biological, economic, and epidemiological differences between the study population and clinical population should be considered.[4] When the populations are similar, research results are directly applicable to the clinical scenario. However, when the study population differs from the clinical population and attempts are made to apply the results to

From the Department of Clinical Sciences, College of Veterinary Medicine, Kansas State University, Manhattan, KS 66506.
Address correspondence to Dr. Larson (rlarson@vt.k-state.edu).

This is the third installment of a series of articles intended to help practitioners enhance their ability to convert research information into clinical application. Please watch for additional articles in the series in future issues of the *JAVMA*.
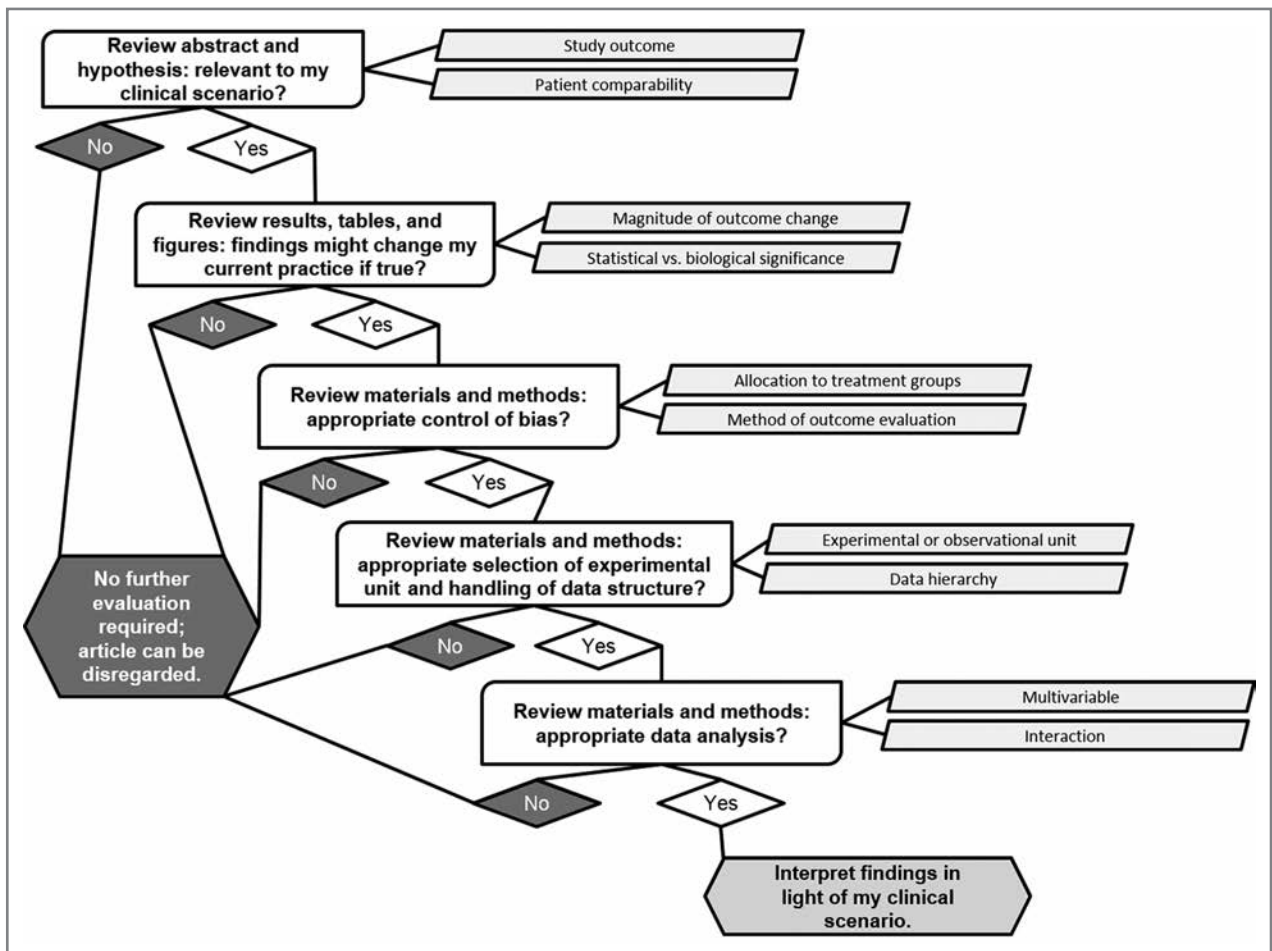
Figure 1—Diagram illustrating 5 steps of a systematic, time-efficient approach to evaluation of the scientific literature to improve clinical decision making.

the clinical scenario, potential errors in interpretation can be made. A clear example of low patient comparability is when results of research involving one animal species are extrapolated to another species; however, even results of research involving the same species may not be directly applicable when the study and clinical populations differ with respect to breed, age, production system, environmental conditions, risk of disease exposure, or other clinically important variables. For example, when the effectiveness of a vaccine against a naturally acquired infectious disease is evaluated in a group of animals, the study population from which those animals were selected might be comparable to the animals of the clinical population in many aspects. However, if the animals used in the study were at high risk for infection and the animals of the clinical population are at low risk, results may not be directly applicable to the clinical scenario. Indeed, for some clinical questions, little research has been performed and clinicians must use results with very low patient comparability (ie from other species or from in vitro studies) to estimate the potential effect of an intervention or factor on their own clinical population.

In many situations, clinical relevance of research articles can be determined by reviewing the title and abstract. The hypothesis and study population are commonly specified, and busy clinicians can use this information to quickly determine whether a given article might be clinically relevant and therefore worth reading. Although research articles with low direct clinical relevance (because of indirect outcome measurements or low patient comparability) may be interesting, the information they provide may lead clinicians to unfounded conclusions about treatment effects that might be expected for their own patients.

Examination of Study Results, Tables, and Figures

When the title and abstract of a research article suggest the article is likely to be clinically relevant, the next step is to consider how the magnitude of the reported effects of the investigated intervention or factor would influence future clinical decisions. Research articles that fail to provide evidence that might meaningfully influence the clinical decision-making process could be interesting but may not merit the time required for further evaluation. The most efficient way to determine whether an article will provide meaningful evidence is to review the tables and figures, and then to review the text of the results section as needed. This strategy does not involve the assumption that the results are true, but rather involves evaluation of the results to determine whether the findings, if true, would provide meaning-

Table 1—Selected components and methods of a systematic, time-efficient approach to evaluation of the scientific literature to improve clinical decision making.

Component	Evaluation method	Evidence to prompt further valuation of article	Evidence to suggest article can be disregarded
Clinical relevance			
Study outcome	List important outcomes for my clinical scenario and determine whether the outcomes evaluated in the study are similar.	Study outcome directly suggests that the intervention or factor was effective.	Study outcome lacks clinical relevance or would not influence clinical decisions.
Patient comparability	Examine the description of the population from which study animals were selected (study population) and compare with with animals in my clinical scenario (clinical population).	Study population is similar to the clinical population.	Study population differs sufficiently from the clinical population and raises questions as to whether similar results could be expected for my clinical scenario.
Control of bias			
Allocation of subjects to (experimental studies) or selection of subjects for (observational studies) study groups.	Examine the materials and methods and compare animal characteristics among groups at study entry to determine whether the groups are similar in all aspects (eg, through randomization), except for the intervention or factor of interest.	Data suggest that groups are comparable with respect to characteristics other than the intervention or factor of interest.	Random assignment or selection of subjects is not mentioned, or evidence of differences among groups at study initiation exists. If all animals in one group share a factor not present in the other groups, results may be confounded and inferences regarding effects of the intervention or factor cannot be made.
Method of outcome evaluation	Examine the materials and methods to determine the method of outcome evaluation (subjective, objective, or both).	Outcome evaluators were effectively blinded with regard to the group to which subjects were assigned or outcomes were measured with objective methods, both of which strategies decrease the likelihood that researchers' preexisting beliefs regarding the effect of the intervention or factor on outcome would inadvertently influence (bias) the results.	No evidence of blinding or objective outcome measurements, which increases the opportunity for researchers' preexisting beliefs to influence the results.

ful evidence to support clinical decision making. If the results would not influence the clinical decision, then there is no need to examine the article further and no further conclusions should be drawn from it.

Evaluation of the reported results should focus on the direction (increase or decrease), magnitude, and precision of effect estimate for each outcome variable. If an effect (ie, difference between treatment groups) is identified but the size of that effect is small enough to be considered clinically unimportant, no further evaluation of the article is required. In other words, detecting significant differences among treatment groups by statistical analysis does not necessarily mean that the difference is clinically important. For example, consider a large randomized controlled trial in which statistical analysis reveals a significant difference in treatment success rates between product A (92%) and product B (91%). Although the difference is significant from a statistical perspective, it may not be biologically or clinically meaningful, and this information would probably not persuade clinicians to change their current method for selection between the 2 products.

Careful examination of the legend of each table or figure is important for understanding the data and results displayed. When a legend is unclear, portions of the text in the results section may need to be reviewed to facilitate interpretation. Tables and figures that contain descriptive data can provide an overview of the animals used in the study or of the study population, which can be useful for determining clinical relevance. Tables and figures that contain statistical summaries of results can be useful for identifying differences between groups; however, because researchers define the cutoff for identifying significant differences (often through use of P values), care should be taken to identify the definition used. A value of $P \leq 0.05$ is commonly used to indicate significant differences; however, researchers may use larger values (eg, $P < 0.10$) to indicate significance, which increases the opportunity for type I error (ie, identifying a difference when one does not truly exist). Significant differences are commonly denoted in figures or tables by use of superscript letters, symbols, or bold text.

Examination of the title, abstract, tables, and figures in a research article should provide clinicians with a sufficient overview of the main study findings to determine whether the article merits further evaluation. However, final conclusions about the study findings should be reserved for later in the evaluation process, after the scientific validity of the research is determined.

Examination of Methods Used to Control Bias

Scientific research is optimally designed to generate an unbiased estimate of the effect of an observed or assigned intervention or factor on a specific outcome. Bias can be defined as an additional factor other than the observed or assigned intervention or factor that influences the study outcome. For a simple example, consider an article regarding a randomized controlled clinical trial conducted to evaluate the effectiveness of a new drug (Wonderdrug), compared with that of an

existing drug (Standarddrug), for treatment of animals with naturally acquired disease. The hypothesis is that Wonderdrug provides a better outcome as judged by fewer relapses of disease within a specific period after treatment than does Standarddrug. The study involves enrollment of 100 diseased animals for each treatment group, administration of the treatment, and determination of relapse risk after treatment. Statistical analysis reveals that treatment with Standarddrug is associated with a significantly and substantially lower relapse risk (21%) than is treatment with Wonderdrug (64%). Because this article appears to be clinically relevant and the results provide evidence of a clinically meaningful magnitude of effect, further evaluation is warranted to ensure that the results were not influenced by factors other than the treatment.

In the aforementioned example, there are several ways by which factors other than the treatment might influence or bias the findings. For example, if only severely diseased animals were given Wonderdrug and only less severely diseased animals were given Standarddrug, then animals given Wonderdrug would likely have a higher relapse risk regardless of treatment. In that situation, researchers may incorrectly conclude that use of Wonderdrug results in a higher relapse risk than does use of Standarddrug, when in fact the true reason for the difference in treatment effect was the bias during allocation of animals to treatment groups.

Although many opportunities exist for introduction of bias into study results, a few methods for minimizing bias can be easily identified when evaluating the scientific literature (Table 1). The process of assigning subjects to treatment groups (for experimental studies) or selecting subjects with particular risk factors or outcomes (for observational studies) is important because valid interpretation of the study results requires that a representative sample of the study population is assigned to each treatment or observation group. This can be achieved through random assignment (for experimental studies) or random selection (for observational studies) of subjects,[5] which promotes even dispersion among the groups of any and all other known or unknown factors that might influence study outcomes, except for the factor or factors of interest. Several appropriate methods for random allocation or selection of study subjects are described in detail elsewhere.[6] Clinicians can determine whether subjects have been appropriately allocated to groups by reviewing the materials and methods in an article. When randomization is used in a study, that detail is usually mentioned in the abstract. Further evaluation of the randomization process involves examination of descriptive data that characterize the study groups at the time of study initiation. This information may appear in the initial paragraphs of the results section or within the first few tables provided. When randomization has been properly implemented and the study sample size is adequate (ie, a sufficient number of animals are included), any differences identified between experimental or observational groups would be due to random chance rather than to systematic bias.

Another point to consider is that reported results are limited to data that were collected during the study; therefore, the possibility exists that factors other than those evaluated might have had an effect on the outcome and that those factors might not have been accounted for in the study. For example, when reviewing a research article to determine whether a given treatment improves weight gain, clinicians should confirm that no meaningful differences in body weight existed between study groups prior to initial treatment administration. If initial body weights were not recorded or reported for each treatment group, then no data exist by which to evaluate the effectiveness of randomization of treatment, and the study should be considered severely flawed and the article not evaluated further.

Confounding is a special type of bias by which a factor other than the intervention or factor of interest is associated with the study outcome, but this factor is not distributed evenly among study groups. Common examples of potential confounders include sex, breed, housing or environment, and age or body weight at the time of study initiation. When a sufficient number of animals are included in an experiment, randomization will eliminate the influence of unidentified confounding factors. The same is not necessarily true for observational studies because data may have been collected from only a specific population or in a manner that influences the results; therefore, when observational data are biased, adding more animals to the study does not eliminate the issue with potential confounding factors. However, sample size is often limited by budgetary constraints, and with small sample sizes, confounding could still exist even in experiments involving random allocation of subjects to treatment groups. When potential confounders are known by the investigators before the study begins, techniques such as blocking or matching by the confounding factors can be used as a part of the subject allocation or selection strategy. For example, when effects of a treatment or other factor of interest are known to be influenced by sex, researchers can randomly allocate subjects to blocks within treatment groups (experiments) on the basis of sex, match subjects on the basis of sex, or analytically control for the effect of sex to ensure that the impact of the external factor on the outcome is appropriately evaluated.

Although confounding can be controlled in many situations through study design or statistical methods, total confounding obfuscates study results. Total confounding occurs when a factor other the intervention or factor of interest differs completely between treatment or observation groups. Consider the example involving a comparison of disease relapse rates between Wonderdrug and Standarddrug. If animals given Wonderdrug were selected from farm A, whereas those given Standarddrug were selected from farm B, then the treatment effect would be totally confounded by the effect of farm on relapse rates. This is because animals of each farm could be expected to differ in important ways (eg, prior care, disease history or risk, immunologic status, or genetics), which might influence relapse risk. Consequently, the observed relapse risk might have been attributable to the drug administered, some unmeasured effect, or both. When study results are totally confounded, no interpretation regarding treatment or factor effects can be made. Opportunities for confounding can be identified by scrutinizing the methods of al-

location or selection for an imbalance in characteristics between study groups that might have influenced the outcome.

Outcome evaluation is another important component of study design that can be influenced by bias. Study outcomes can be subjectively or objectively measured. Subjective measurements, such as those based on human interpretation or opinion, can be influenced by preexisting beliefs regarding the effect of an intervention or factor on an outcome and the role that other factors might play in that outcome. For example, in the Wonderdrug versus Standarddrug study, relapse might have been defined as the presence of clinical signs (eg, signs of depression) of the disease for which the animals received treatment, which would involve subjective rather than objective measurements. If the outcome evaluators were aware of the treatment each animal received, their findings might be subconsciously influenced by preexisting beliefs regarding treatment effects, introducing bias. Even seemingly objective measurements such as mortality rates can have a subjective component when researchers control their definition (eg, choice of follow-up period for mortality rate calculation) or the study design allows researchers to make exceptions in certain conditions (eg, removal of animals from a study prior to death when the animal is identified as chronically ill).

The technique of blinding, by which methods are used to ensure outcome evaluators have no knowledge of the groups to which subjects have been assigned, greatly reduces opportunities for inadvertently biased outcome assessment. Indeed, when possible, all personnel involved in caring for study animals as well as those responsible for observing them should not be aware of treatment group (for experimental studies) or factor or outcome group (for observational studies) to which animals have been assigned. Blinding allows clinically relevant subjective measurements such as detection of clinical illness by only 1 evaluator to serve as valid study outcomes (provided that the outcome measured is suitable for the tested hypothesis) because the probability of mismeasurement is similar for all animals.

Despite the importance of randomization and blinding, many research articles do not include information regarding whether these bias-controlling techniques were used.[7,8] When randomization and blinding are not used in studies that should include them, the results should be considered unreliable and thereby disregarded when making clinical decisions. By nature, retrospective observational studies do not include randomized allocation and blinding; however, other less-effective methods can and should be used to reduce the bias introduced by confounding in these types of studies, such as matching, inclusion of only animals with a known confounding factor, or analytic control of confounding variables.

Clinical Summary

A systematic method for evaluation of the scientific literature includes assessment of the clinical relevance of the research described therein and an understanding of common methods used to control bias. Review of the abstract, results, tables, and figures will provide clinicians with enough information to determine whether the research is clinically relevant and important to the decision-making process. Clinicians should also consider the methods used to select or allocate subjects among groups and to evaluate outcomes. A systematic approach to literature review can prevent potential misinterpretation of study results while providing an efficient process for gaining clinical information from research articles. The method proposed in the present article was created on the basis of the authors' experience and contains steps that we have found beneficial, but each clinician should determine the method that best fits their own situation.

References

1. Larson RL, White BJ. First steps to efficient use of the scientific literature in veterinary practice. *J Am Vet Med Assoc* 2015;247:254–258.
2. Johnson PD, Besselsen DG. Practical aspects of experimental design in animal research. *ILAR J* 2002;43:202–206.
3. Bucher HC, Guyatt GH, Cook DJ, et al. Users' guides to the medical literature. XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA* 1999;282:771–778.
4. Dans AL, Dans LF, Guyatt GH, et al. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA* 1998;279:545–549.
5. Shott S. Designing studies that answer questions. *J Am Vet Med Assoc* 2011;238:55–58.
6. Boothe DM, Slater MR. Standards for veterinary clinical trials. *Adv Vet Sci Comp Med* 1995;39:191–252.
7. Kilkenny C, Parsons N, Kadyszewski E, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* 2009;4:e7824.
8. Elbers AR, Schukken YH. Critical features of veterinary field trials. *Vet Rec* 1995;136:187–192.