

Enhancing Clinical Decision Making

Systematic evaluation of scientific research for appropriateness of data analysis to improve clinical decision making

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

Many research articles published in well-respected journals describe studies with serious design flaws that limit their value for clinical decision making; therefore, it is important that clinicians critically evaluate the scientific literature they read.¹ Methods have been developed to guide critical evaluation of published articles concerning randomized clinical trials²⁻⁴ and observational studies.⁵ The REFLECT² (Reporting guidelines For randomized controlled trials for livestock and food safety) and STROBE⁶ (Strengthening the Reporting of Observational studies in Epidemiology) statements are guidelines intended to encourage standardized reporting of research and facilitate clinician evaluation of research integrity to gauge validity of the results. When clinicians are seeking to answer clinical questions, an effective and efficient systematic approach to evaluation of the scientific literature can reduce the amount of time spent reading the literature and minimize potential misinterpretation of research findings, thereby improving the value of the information gleaned.

In a previous article,⁷ the first 3 steps of a 5-step systematic method for time-efficient literature evaluation were described (Figure 1). The purpose of the present article is to describe the last 2 steps to further help veterinary clinicians obtain information to address their clinical questions.

Identification of Experimental Unit and Data Hierarchy

Proper interpretation of study results requires an accurate understanding of the primary unit of interest in experimental or observational studies (also referred to as experimental or observational units) and the inherent data hierarchy (Table 1). An experimental or observational unit can be an individual animal in circumstances in which an intervention or factor of interest is observed and evaluated at the level of the individual animal and the outcome for one animal is independent of the outcome for another animal. An experimental or observational unit can also be a group of animals housed together or even an entire farm or herd when interventions or factors of interest are observed and evaluated at that group level. For example, when calves on a farm are randomly assigned to receive one vaccine or another vaccine and the outcome is independent

among calves, then the experimental unit would be the individual calf. In that situation, whether a calf receives or responds to a particular vaccine would be independent of the vaccine that the other calves receive. On the other hand, when pens of calves on a farm are randomly assigned to receive one of the vaccines, with every calf in a pen receiving the same vaccine (without the possibility of receiving the other vaccine), then the pen and not the individual calf is the experimental unit of interest. In that situation, calves in the same pen are not independent of each other, but the outcome for one pen would be independent from the outcome for another pen. A study can also involve 2 types of experimental or observational units in which different outcomes are measured (eg, individual animal evaluated for weight gain and pen of animals evaluated for feed efficiency).

Sample size (ie, the number of experimental or observational units needed for each study group) and statistics should be calculated on the basis of the experimental or observational unit used and not necessarily on the basis of individual animals. To determine the sample size necessary to test whether differences exist in outcomes between treatment or observation groups, an estimate is needed for the magnitude of differences observed. An estimate is also needed for the variability in outcome measurements within each group (ie, within-group variability), given that no 2 units within a group will likely respond exactly alike, provided they are truly independent of each other. Estimates of within-group variability must be obtained by including > 1 unit/study group and performing (replicating) the intervention or observation on each unit; when only 1 unit is included per group, no variability in outcome measurements exists.

The concepts of experimental or observational units and replication are interrelated. A common error in scientific research is to incorrectly identify the experimental or observational unit and, in so doing, fail to appropriately replicate the interventions or observations. Improper identification of the experimental unit could lead to erroneous handling of multiple observations of the same experimental or observational unit as valid individual observations independent of each other. For example, consider a hypothesis that coat color contributes to the likelihood of illness. If 6 animals (3

From the Department of Clinical Sciences, College of Veterinary Medicine, Kansas State University, Manhattan, KS 66506. Address correspondence to Dr. White (bwhite@vet.ksu.edu).

This is the fourth installment of a series of articles intended to help practitioners enhance their ability to convert research information into clinical application.

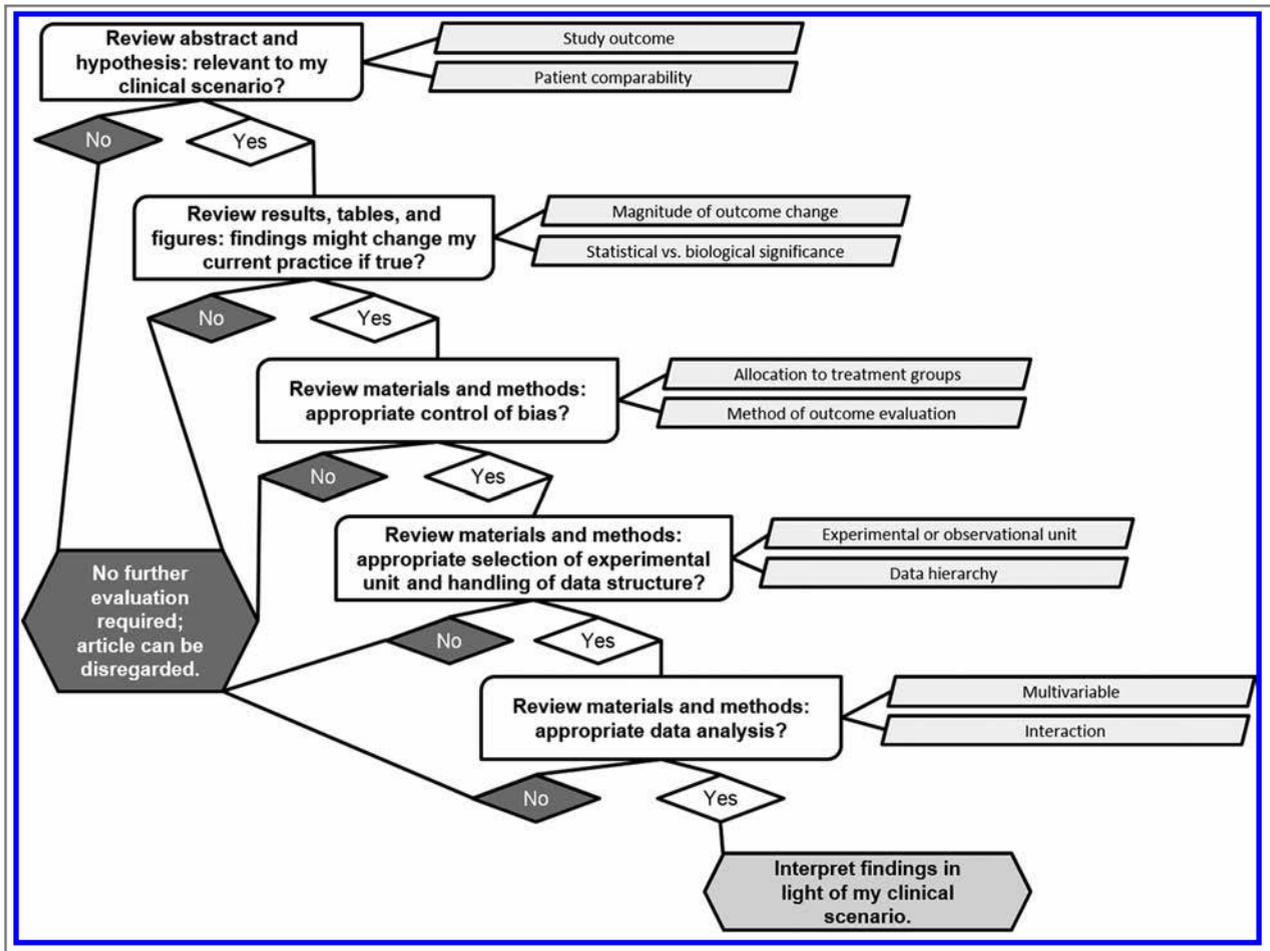


Figure 1—Diagram illustrating 5 steps of a systematic, time-efficient approach to evaluation of the scientific literature to improve clinical decision making.

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

Component	Evaluation method	Evidence to prompt further evaluation of article	Evidence to suggest article can be disregarded
Experimental or observational unit	Examine the materials and methods to identify the units to which treatments were allocated (experimental studies) or for which observations were made (observational studies) and outcomes were evaluated.	Treatments were applied and evaluated (experimental studies) or variable of interest was observed and evaluated (observational studies) at an appropriate level (eg, individual or group), which represents appropriate replication of the intervention or observation.	Treatments were applied or observations were made at one level (eg, pen or herd), but statistical evaluation was performed at a different level (eg, individual animal), which represents inappropriate replication of the intervention or observation.
Data hierarchy	Examine the description of the statistical analysis in the materials and methods.	Factors that threaten independence among observations (eg, repeated measurements on individuals or similar housing for groups of individuals) have been appropriately accounted for in the study design or statistical analysis.	Study design involves a data hierarchy that limits independence among observations, and this lack of independence is not accounted for in the statistical analysis.

sick and 3 healthy) were evaluated, and 2 of the 3 sick ones were identified as black whereas 1 of the 3 healthy ones was identified as red, then the conclusion might be that the difference between the groups is not significant (ie, $P = 0.22$), providing no evidence that coat color is associated with illness. However, as an example of improper identification of the experimental unit, consider that if 1,000 hairs were collected from each of those

animals, and each hair was evaluated as an independent observation, then 2,000 of 3,000 hairs from the sick animals would be black, compared with 1,000 of 3,000 hairs from the healthy animals. The difference would then become significant ($P < 0.01$). Although the proportions are similar between these examples, the considerably larger sample size in the second example (3,000 hairs/group vs 3 hairs/group) results in a lower

estimate of within-group variability (or SE). Indeed, the amount of within-group variability is substantially lower in the second example than in the first example because multiple specimens were obtained from the same animals.

A more realistic example might be a study designed to evaluate the efficacy of a feed additive for prevention of disease in groups of calves. If the calves were housed in pens or groups, the additive would by necessity be applied at the pen level, making pen the experimental unit. If this study involved 6 pens with 100 calves in each pen, with calves in 3 pens receiving the feed additive (treatment group) and calves in the other 3 pens receiving a placebo (control group), then an appropriate outcome measurement would be morbidity rate per pen (with 3 morbidity rates [1 per pen] obtained for each study group). However, if an individual calf were inappropriately identified as the experimental unit, comparison of the likelihood of a calf becoming ill without accounting for pen-level differences may lead to a different conclusion because of the high number of apparent replicates (3 pens \times 100 calves/pen = 300 units/study group). Such an error would again reduce the estimate of within-group variability and increase the likelihood of identifying a difference between the study groups when none exists (ie, type I error). To reiterate, these apparent replicates are not true replicates because all calves within the same pen received the same treatment (and could not have received the other treatment); therefore, observations made would not be independent of each other.

Evaluation of Interactions Between Study Factors

Scientific studies are not always straightforward such that only a single factor differs between treatment or observation groups. Findings from studies in which the effects of > 1 treatment or factor are compared between groups (ie, multivariable analysis) are often valuable, yet care should be taken when evaluating the results. For example, consider a study conducted to evaluate the effect of 2 drugs (A and B) for treatment of an infectious disease. The study could be designed to explore the potential impact of each drug separately, the impact of both drugs combined, and the impact of no treatment (negative control group), which results in 4 treatment groups. Animals enrolled in the study could be allocated to the 4 groups such that group 1 received drugs A and B, group 2 received drug A only, group 3 received drug B only, and group 4 received no drug. Several comparisons could be made, including whether the effect of drug A is modified by the presence or absence of drug B. The hypothesis could be stated as, "We believe the effect of drug A is modified in the presence or absence of drug B (ie, that an interaction exists between the 2 drugs)." The term interaction in this sense describes the situation that occurs when the effect of one treatment or factor on the outcome variable is modified by the effect of another treatment or factor.

Consider the results of 2 randomized controlled trials conducted to compare proportions of animals that recovered from an infectious disease within a specified

follow-up period (Figure 2). In the first trial, the effect of drug B alone can be evaluated by determining that 20% of the animals in the control group (no drug A or drug B) recovered, whereas when drug B but not drug A was given, that percentage increased to 30%, representing an absolute increase of 10% (ie, $30\% - 20\%$) or actual increase of 50% (ie, $[30\% - 20\%]/20\% \times 100\%$). The effect of drug A alone can be similarly evaluated, whereby 60% of animals receiving drug A recovered, which represented an absolute increase of 40% (ie, $60\% - 20\%$) or an actual increase of 200% (ie, $[60\% - 20\%]/20\% \times 100\%$) from the control value of 20%. Graphic depiction of these percentages reveals that responses did not appear to influence each other, given that the lines representing each drug are parallel, which suggests that the drug effects were independent of each other and without an interaction. In this circumstance, results of this first trial could be interpreted by comparing the individual effects of each treatment alone.

In the second trial, 30% of animals given drug B alone and 20% of animals given drug A alone recovered (Figure 2). However, this time 95% of animals given both drugs recovered, which is considerably greater

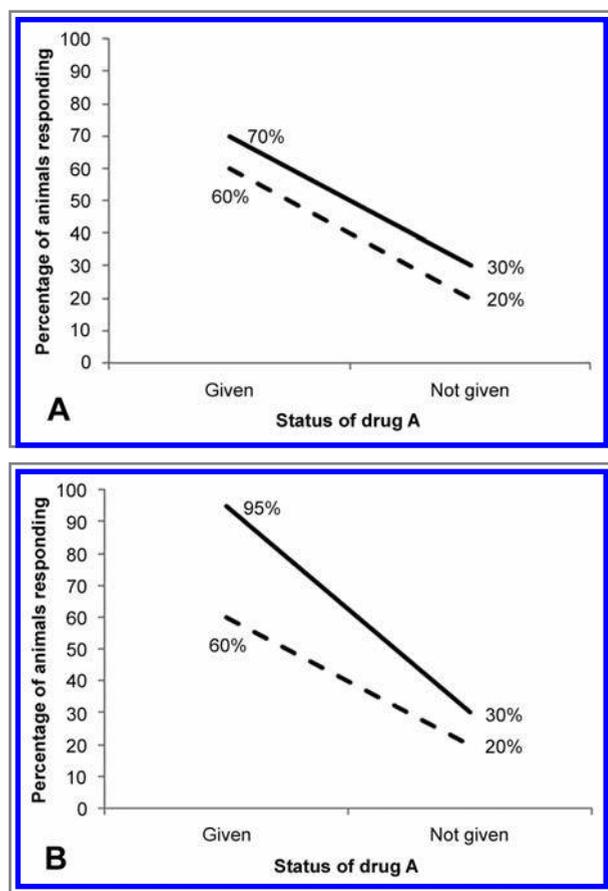


Figure 2—Percentage of animals that responded to treatment with or without drug A and with (solid line) and without (dashed line) drug B in a hypothetical randomized controlled trial. A—The lines for the treatment response are parallel, which suggests that no interaction exists between drugs A and B. B—The lines for the treatment response are not parallel, and the response to combined treatment with drugs A and B appears to be greater than when only 1 drug is administered, which suggests that an interaction exists between drugs A and B.

than the recovery percentage for animals given both drugs in the first trial (60%), when the effects of the drugs were independent of each other. The difference between trials in recovery percentages for animals given both drugs A and B suggests that the effect of drug A was modified (improved in this situation) by the presence of drug B and vice versa. For this second trial, graphic depiction of the results reveals that the lines representing the results for each drug are not parallel, confirming that an interaction exists. This example emphasizes the importance of considering the effects of each drug alone as well as their interaction when reporting study results or making clinical decisions.

Consideration of Bias in Clinical Interpretations

Results of studies that meet the relevant criteria for valid, unbiased research should be incorporated into the clinical decision-making process. However, clinicians have preexisting beliefs when reviewing the scientific literature, which may influence confidence in a given study's results, regardless of how well the study was designed or conducted. This influence is acceptable provided that the preexisting beliefs were based on valid research and open to modification as new information is encountered. Formation of scientific opinions is an iterative process, and knowledge builds over time.⁸ Confidence in those opinions increases as the number of studies that support them also increases.

Unfortunately, it is not always possible to identify multiple well-designed studies that address a particular clinical question, and findings from a single study often are the only information available to support a clinical decision. When research results are applied to a clinical scenario, clinicians should consider the study outcome measured, patient comparability, and number of appropriate studies available to assist the decision-making process. Clinicians should also interpret results within the limits of the study, being careful to avoid extrapolation of those results beyond the study population. This is challenging and is one of the reasons that keeping oneself apprised of scientific findings to aid decision making is a fairly slow process.

Evaluation of the scientific literature in a systematic manner facilitates identification of potential problem areas in study design or reporting. If severe flaws such as those described in previous articles^{7,8} are identified when reading a research article, then clinicians can disregard the article at that point, without reading further. Although the decision to ignore published literature that is severely flawed may be unfamiliar or uncomfortable initially, clinicians are encouraged to remember that severely flawed research is unable to yield valid conclusions. Clinicians may underestimate the importance of potential errors in studies, particularly when the findings support their preexisting beliefs, and information obtained from such studies may further support a clinical impression that is not necessarily true.

Consider a clinician who reviews a research article in which diagnostic test A was determined to be

more accurate than diagnostic test B, which supports the clinician's prior belief that test A is superior. However, after reading further, the clinician learns that the outcome evaluators were not blinded to animal disease status when the diagnostic tests were used. The clinician might realize that the lack of blinding represents a severe flaw in the study but might still trust the findings because they support his or her preexisting beliefs. That clinician might therefore continue to use test A with more confidence than before reading the article. On the other hand, if a clinician begins reviewing the same article with the belief that diagnostic test B is more accurate than diagnostic test A, he or she might choose to ignore the study results because no attempt was made to blind outcome evaluators. In that situation, the clinician might therefore continue to use test B. In these examples, different preexisting beliefs led to different interpretations for the same study results. This emphasizes the point that regardless of preexisting beliefs, results of any study with severe flaws that bias those results are therefore unreliable and should be ignored.

Clinical Summary

When clinicians evaluate the scientific literature to answer a clinical question, they should consider the clinical relevance, clinical importance, and validity of the research. Common errors that may invalidate research findings can be identified by becoming familiar with the typical methods used to control for bias, ensure appropriate allocation and handling of experimental units, and deal with the structure of research data or populations. Research articles can be reviewed in a time-efficient and meaningful manner when a systematic approach is used.

References

1. Kastelic JP. Critical evaluation of scientific articles and other sources of information: an introduction to evidence-based veterinary medicine. *Theriogenology* 2006;66:534–542.
2. O'Connor AM, Sargeant JM, Gardner IA, et al. The REFLECT statement: methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety by modifying the CONSORT statement. *Zoonoses Public Health* 2010;57:95–104.
3. Sargeant JM, Elgie R, Valcour J, et al. Methodological quality and completeness of reporting in clinical trials conducted in livestock species. *Prev Vet Med* 2009;91:107–115.
4. Kilkenny C, Browne W, Cuthill IC, et al. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br J Pharmacol* 2010;160:1577–1579.
5. Sargeant JM, O'Connor AM. Issues of reporting in observational studies in veterinary medicine. *Prev Vet Med* 2014;113:323–330.
6. Universität Bern. STROBE statement: strengthening the reporting of observational studies in epidemiology. Available at: www.strobe-statement.org. Accessed May 15, 2015.
7. White BJ, Larson RL. Systematic evaluation of scientific research for clinical relevance and control of bias to improve clinical decision making. *J Am Vet Med Assoc* 2015;247:496–500.
8. Larson RL, White BJ. Importance of the role of the scientific literature in clinical decision making. *J Am Vet Med Assoc* 2015; 247:247:58–64.