

# How to Read a Scientific Paper

David W. Ramey, DVM

Biomedical literature is expanding at a phenomenal pace. At the same time, the time that's available to read that literature is becoming increasingly hard to find. Thus, if you want to try to stay current with medical developments and you don't want to get overwhelmed, it is important to develop a system for reading and evaluating papers of interest. This paper will look at *why* you might read a scientific paper, suggest ways to decide *whether* to read a particular paper, and *how* to interpret the evidence presented. Author's address: PO Box 5231, Glendale, CA 91221. © 1999 AAEP.

## 1. Why Read a Scientific Paper?

If you want to continue to try to do more good than harm for the horses in your care, you need to recognize the need to change and/or improve your diagnostic and therapeutic interventions so that they remain consistent with valid new knowledge. If you don't stay current, you run the risk of falling short in your clinical practice. Clinical journals are generally the most accessible means of obtaining the information that you need.

There are any number of reasons why you might read a clinical journal; 10 of them are listed in Table 1. Of particular interest to practitioners are items 5–7; these will be discussed in more detail later in this article.

## 2. Whether to Read a Particular Scientific Paper

It's not possible for an individual veterinary practitioner to know everything about veterinary medicine or even *equine* medicine and surgery. Given the demands of clinical practice and the desire to maintain some sort of a nonpractice life, it seems reasonable to assume that you're already behind in your reading and that you will *never* have more time to read than you do right now. Thus, to make the most

out of your reading time, you should consider focusing on the *few* articles that are both valid and applicable to your area of interest and rejecting *most* articles almost immediately. The following guidelines should help you do that (Figure 1).

## 3. Is the Study Relevant?

If the study doesn't apply to you or your practice, it may not be worth reading at all. Here are a few suggestions on how you might be able to tell if an article is worth a more thorough evaluation.

1. *Look at the title.* Is the article one that is potentially useful in your practice? For example, do you have a reproductive practice and is the article about a new orthopedic technique? If so, you may consider rejecting the article out of hand and go on to the next article.

2. *Read the summary.* Here your objective is simply to decide if the conclusion, if valid, would be important to you as a clinician. The issue here is not whether the results are true; rather, it is whether the results, if true, would be useful. Most summaries can be found in the abstracts that precede full-text articles.

---

## NOTES

**Table 1. Ten Reasons to Read Clinical Journals**

1. To impress others
2. To keep abreast of news in the profession
3. To understand pathobiology
4. To find out how a seasoned clinician handles a particular problem
5. To find out whether to use a new or existing diagnostic tests on your patients
6. To learn the clinical features and course of a disorder
7. To distinguish useful from useless or even harmful therapies
8. To determine etiology or causation
9. To sort out claims concerning new therapies
10. To read the letters to the editor

3. *Consider the site.* Here, the question is if the site of the study is such that it might apply to your practice. For example, if a new technique for laparoscopic renal biopsy is proposed, would you have the expertise, or access to the required facilities and equipment to perform the technique? Similarly, it's useful to look at the horses that are included in the study; a technique that is applied to racehorses may

not be useful for foals. Putting it another way, are the patients that are included in the study similar to those in your own practice and could you apply the results even if you wanted to?

**4. Is the Study Valid?**

Unfortunately, you can't assume that a study is worth a look just because it appears in a reputable journal. The review and editorial policies of even the best journals do not protect the reader from errors. In fact, it's reasonable to be skeptical of the conclusions of almost *any* article from the start. That's because the mere fact that the article got printed in the first place makes it subject to potential biases. A list of these can be found in Table 2.

You cannot accept an article solely on the basis of its conclusions. If you do so, you run the risk of accepting false information. A summary can sometimes tell you that an article is *invalid* but it can almost never tell you if it is valid. Thus, if you've made the decision to read an article, there is no alternative to investing your first efforts in reviewing the Materials and Methods sections of an article.

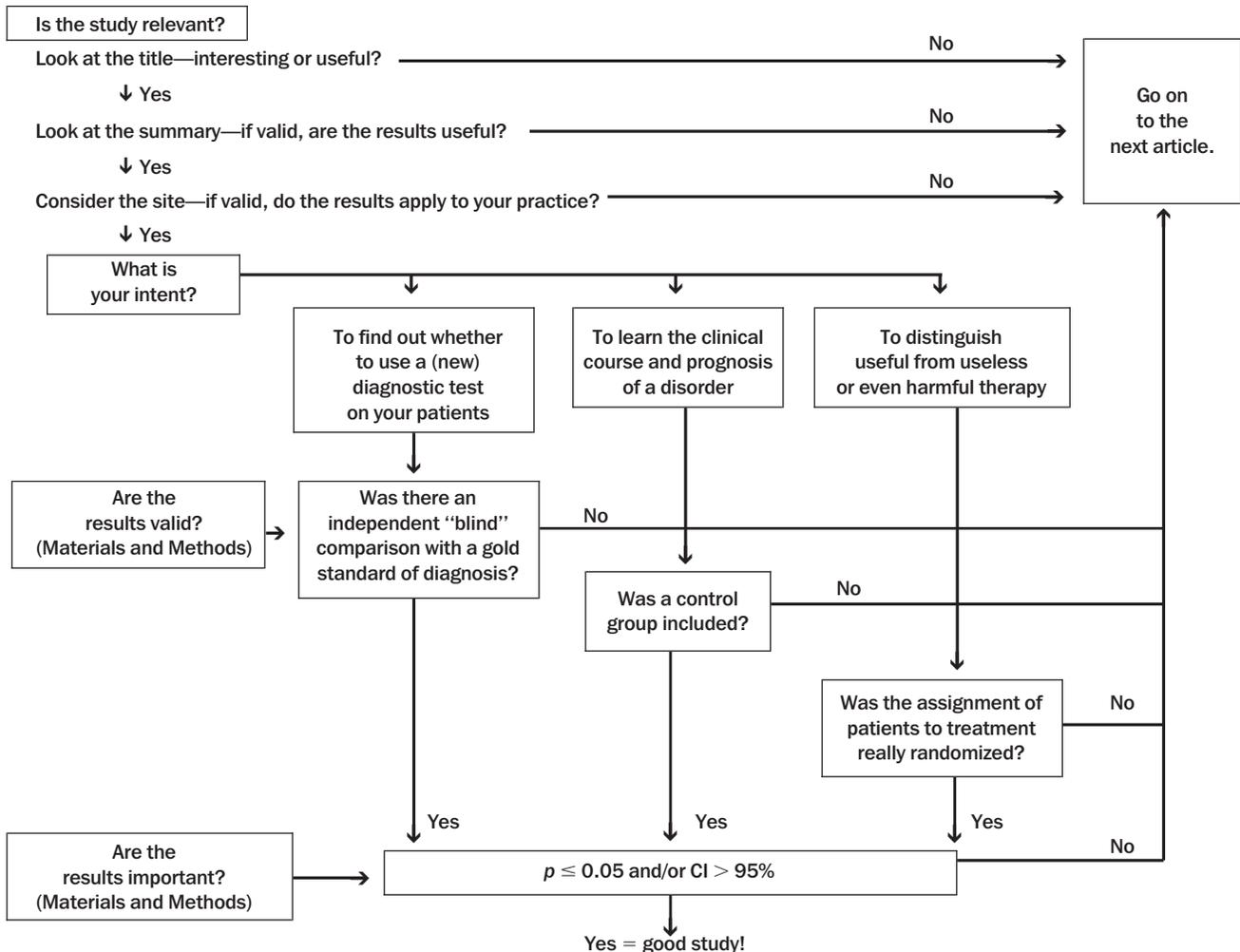


Fig. 1. Flow chart for selecting scientific reading material.

## BACK TO BASICS

Table 2. Sources of Positive Bias in the Reporting of Controlled Trials<sup>a</sup>

Source of Bias	Cause
Submission bias	Research workers are more strongly motivated to complete, and submit for publication, positive results
Publication bias	Editors are more likely to publish positive studies
Methodological bias	Methodological errors such as flawed randomization produce positive biases

<sup>a</sup>Data from Gray.<sup>2</sup>

It has been noted that the “Conclusion giveth, but the Materials and Methods taketh away.” That’s where the meat of any article can be found.

The reasons for reading a paper of most general interest to practicing clinicians are likely to be numbers 5–7 in Table 1. Of most particular interest is item 7, “To distinguish useful from useless or even harmful therapy.” This paper will focus only briefly on items 5 and 6, with more detail provided on item 7. More thorough discussions can be found about how to critically read papers on any of these, as well as several other, subjects.<sup>1</sup>

1. *To find out whether or not to use a (new) diagnostic test on your patients.* When encountering an article about diagnostic tests, the first thing to look for is if there was an independent “blind” comparison with a “gold standard” of diagnosis. That is, patients must be shown, by the application of some *objective* method of determination, such as a biopsy, surgery, postmortem examination or long-term follow-up, to have had the disease in question. There must be a second group of patients shown by the same standard *not* to have had the disease. Then, the test should have been interpreted by clinicians who didn’t know whether or not a given patient really had the disease (that is, they were “blind”). Afterward, these test results should be compared to the gold standard. If this litmus test of validity isn’t followed, you may consider rejecting the paper (and the test) out of hand.

2. *To learn the clinical course and prognosis of a disorder.* When you’re reading to find out about the clinical course or prognosis of a particular treatable condition, the first thing to look for is if there was a control group assembled. Control groups are used for comparison with the group being studied. Ideally, the control group is identical to the study group, except that it does not possess the characteristic or has not been exposed to the treatment under study. Controls should ideally include no treatment groups, as well as groups that receive placebo treatments. Controlled studies may be divided into *case-control* studies, in which the cases are horses that have already developed the disease and the controls are those that have not, or *cohort* studies, in which the horses are separated into study and control groups before the investigator is aware of whether the horses have or will develop the condition being studied.

Controls are especially important in studying subjective analyses (did the horse get “better?”) where there might be a significant observer bias (that is, where the observer might be inclined to believe that there was improvement merely because the patient was treated). If the study fails to include controls, the results of the study will be unpredictable. Accordingly, you may think about passing by that particular paper.

3. *To distinguish useful from useless, or even harmful, therapy.* When trying to determine if a new therapy is worth trying in your practice, the first thing to look for is if the study was randomized. Randomization means that some method, such as flipping a coin, was used to assign patients to the treatment groups. The reason that randomization is so important is that it is the best way to group patients at the start of a trial who are identical in their risk of events that are going to be studied. It does this in two ways.

First, randomization tends to balance the groups for prognostic factors, such as the severity of the disease in question. If prognostic factors are unevenly distributed, they can exaggerate, cancel, or even counteract the effects of therapy. This can lead to false-positive or false-negative results.

Second, if the studying clinicians are unaware of the randomization, that is, if they are “blinded” to it, they won’t know which treatment the next patient will receive. Thus, they won’t be able to distort, consciously or unconsciously, the balance between the two groups being compared. If randomization isn’t concealed, it tends to lead to a situation where patients with a more favorable prognosis receive the experimental therapy. This can exaggerate the benefits of therapy and perhaps even lead to a false-positive conclusion.

Ideally, the study will be “double-blinded.” That is, neither the patient nor the clinician will know who is receiving treatment until after the study is over. Unfortunately, blinding isn’t possible in all studies (for example, surgical ones); however, it’s still a goal that should be pursued wherever possible so as to try to eliminate subtle biases that may influence study results.

How much do things like randomization, blinding, and controls really matter? Quite a bit, actually. Table 3 estimates the effect that ignoring these critical factors may have on study results.

Table 3. Effects of Poor Design of Controlled Trials on Estimates of Treatment Effects (Trials with Poor Evidence of Randomization vs. Trials with Adequate Randomization.<sup>a</sup>

Design Fault	Exaggeration of Odds Ratio
Inadequate method of treatment allocation	Larger by 41%
Unclear method of treatment allocation	Larger by 30%
Trials not double blind	Larger by 17%

<sup>a</sup>Data from Schultz et al.<sup>3</sup>

It's usually easy to tell if a study is randomized, controlled, and blinded. It's something to be proud of and it's usually right there in the title and/or abstract of the paper. If a paper doesn't meet these criteria, you may consider rejecting it out of hand. However, if you're still curious about the validity of the paper, then you'll want to look critically at the evidence provided in support of the conclusions and at how the conclusions were drawn.

**5. What About the Evidence?**

Not all, perhaps not most, research papers in equine medicine are randomized, blinded, and controlled. If that's the case, you're in a bit of a bind. For example, you may end up trying to determine the usefulness of a potential therapy based on evidence that you know may not be all that good. To help sort things out, it may be useful to think of the *quality* of the evidence with which you are being presented.

All evidence is *not* created equal. A hierarchy of evidence quality has been developed by the Centre for Evidence-Based Medicine in Oxford, England. Using it may help you decide if the results presented in the study are likely to be true. In order of importance, the levels of evidence are indicated in Table 4.

Even if the quality of evidence is poor, there may even be a couple of things that you can glean from poor studies.

**Table 4. Quality of Scientific Studies**

Grade of Recommendation	Level of Evidence
A	1a Systematic review (with homogeneity) of randomized, controlled trials
	1b Individual randomized, controlled trial (with narrow confidence interval)
	1c All or none studies (studies in which the effects are so great that they could not have been attributed to chance; treatments for disorders in which all died and now some live, or some went on to a bad outcome and now none do so)
B	2a Systematic review (with homogeneity) of cohort studies
	2b Individual cohort study (including low quality randomized, controlled trial; e.g., <80% follow-up)
	2c "Outcomes" research (nonexperiments linking outcomes, e.g., death, to treatments, e.g., surgery)
	3a Systematic review (with homogeneity) of case-control studies
	3b Individual case-control study
C	4 Case-series (and poor quality cohort and case-control studies)
D	5 Expert opinion without explicit critical appraisal, or conclusions based on physiology, bench research (in vitro or lab animal studies), or basic biologic principles

1. *See if the treatment effect is so huge that you can't imagine that it could be a false-positive study.* This usually only happens when the prognosis is uniformly terrible; it's a very rare situation.

2. *If the nonrandomized study concluded that a therapy was useless or harmful, it's usually safe to accept that conclusion.* False-negative conclusions from studies are less likely than false-positive ones.

**6. Are the Results Important?**

A. Statistical Evaluation

Finally, it's worth your time to take a look at the statistical conclusions reached in the paper. If you've decided that a paper is worth reading and you understand the evidence that's provided in it, you should take a look to see the likelihood that the results have some significance. In that regard, there are two particularly useful indicators.

1. *p value.* Perhaps the most commonly used indicator of significance is the *p* value. Basically, the *p* value is a measurement of the likelihood that the data obtained merely arose by chance or natural variation. A *p* value of <0.05 has been arbitrarily chosen as the standard of significance, but it is not a magic number. You can define to whatever level of probability you want to choose. Still, a paper with a *p* of >0.05 may not be worth paying much attention to; on the other hand, you'd like to see a *p* of as far below 0.05 as possible.

However, the *p* value does have some limitations. On the good side, a *p* of <0.05 means that there is less than a 5% probability that the evidence gathered in the study occurred by chance. (It should be noted that a *p* of <0.05 does *not* mean that there is a 95% chance that the results of the study are valid.) However, looking at it the other way, that means that there is still a 5% probability (1 in 20) that the results *did* occur by chance. Statistical significance tests can measure probabilities, but no matter what level of significance is chosen, there is *always* some probability of seeing a difference between studied groups when none really exists.

The *p* value has a couple of other limitations as well. By itself, the *p* value tells nothing about the size of the difference in the study groups. A statistically significant difference may have slight clinical relevance. For example, in one study of approximately 40,000 people comparing clot dissolving agents there was a reduced risk of 0.5 deaths per hundred from myocardial infarction when the more expensive agent was chosen. This difference was highly significant statistically, but it translated to an actual difference of 1 life per 200 people treated. (Whether the benefit is worth the cost is not a question that can be answered statistically.) Nor does the *p* value tell you the direction of the statisti-

## BACK TO BASICS

cal difference; it doesn't tell you if treatment "a" is better than treatment "b," for example. Other study factors must be interpreted for that assessment.

2. *Confidence intervals.* Perhaps a more useful, although less commonly reported, indicator of statistical significance is the confidence interval (CI). The CI gives a measure of the precision (or uncertainty) of study results for making inferences about the population of similar individuals. Confidence intervals combine information about the strength of an association with information about the effects of chance on the likelihood of obtaining the results. It is possible to calculate the confidence interval for any percentage confidence from 0 to 100; however, most studies choose a CI of 95 per cent. A 95% CI is an interval of numerical values within which you can be 95% confident that the mean of the population will be included. However, you *cannot* say that the studied parameter is in the 95% CI with a 95% probability.

Confidence intervals place a clear emphasis on quantification of the effect, in direct contrast to the *p* value approach (which arises from significance testing). CIs indicate the strength of the evidence about quantities that are directly relevant, such as

treatment benefits. Thus, they are particularly important to practitioners.

### 7. Conclusion

By using the aforementioned guidelines, you will almost certainly be able to dramatically reduce your reading time. Unfortunately, if you apply them strictly, you may find that you have virtually nothing to read. Nevertheless, using a systematic approach to reading scientific papers will help you sort out the wheat from the chaff and help make sure that the precious time that you devote to reading is well spent.

The author would like to thank Drs. David Sackett and Martha Lee for their assistance in the preparation of this paper.

### References

1. Sackett DL, et al. *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill-Livingstone, 1998.
2. Gray JAM. *Evidence-based healthcare*. London: Churchill Livingstone, 1997.
3. Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-412.