

EBVM: a quick guide to evaluating veterinary evidence

by Cartlidge, H.

Copyright, publisher and additional information: .This is the authors' accepted manuscript. The published version is available via MAG Online.

Please refer to any applicable terms of use of the publisher

[DOI link to the version of record on the publisher's site](#)



**Harper Adams
University**

EBVM: A QUICK GUIDE TO EVALUATING VETERINARY EVIDENCE

Abstract:

The ability to evaluate evidence is a key skill for veterinary professionals pursuing an evidence-based approach to patient care. The evidence available on a particular topic in the veterinary field may be of variable quality though and the strengths and weaknesses of the type of evidence should be considered. The way a research study is conducted can also impact on the validity and reliability of the results presented and aspects of study design such as control groups, representative samples, sample size, elimination of bias and outcome measures should be evaluated. This article aims to give further insight into the evaluation of research studies including examples to aid understanding.

Key Words: Bias, Evidence, Confounding, Control group, Reliability, Representative sample, Validity

Key Points:

- Evidence Based Veterinary Medicine is the use of current best evidence in making clinical decisions
- The ability to evaluate the strengths and weaknesses of research evidence is a key skill for veterinary professionals.
- A control group allows the relative merits of a treatment to be appreciated.
- Selection bias or allocation bias in a study can impact on the validity of the results.
- Outcome measures should be valid and reliable.

INTRODUCTION

In the past the evaluation of published research may have been perceived as something that was reserved for those undertaking academic studies. However, in recent times there has been recognition of the need to justify clinical decisions and to ensure the best treatment decisions are being made for our patients. This is better known as Evidence Based Veterinary Medicine (EBVM); the use of current best evidence in making clinical decisions (Cockcroft and Holmes 2003).

Unfortunately, the pursuit of EBVM is not just as simple as finding relevant research for a case, noting the results and implementing the care suggested. The quality of research available is variable so the ability to evaluate the strengths and weaknesses of sources is key to the practice of EBVM and is therefore a key skill that all veterinary professionals should possess.

This article will consider the different sources of evidence you may come across and the relevant aspects of study design that should be evaluated to assess the strength of the research presented.

TYPES OF EVIDENCE –

Evidence can come in many forms (see Box 1). A randomised controlled trial (RCTs) is often considered the 'gold standard' of experimental research however in the authors experience the availability of RCTs for some topics in the veterinary field can be limited. This may be in part due to ethical restrictions, lack of funding or lack of availability of cases. When collecting evidence on a topic you may come across forms of evidence such as case studies and expert opinion. As EBVM utilises the 'best available evidence', where evidence is limited such forms of evidence may be useful in shaping our clinical decisions however, we must appreciate the potential weaknesses in the forms of evidence we encounter. For example:

- Expert opinion is as it states, 'an opinion' and has therefore been influenced by that person's positive and negative experiences so will be biased.
- A case report is based on the treatments and outcome for just one individual; It is not possible to conclude from one case if the treatment approach used would be the most beneficial to use in other cases.

The evidence pyramid (Figure 1) is a good tool to help us understand the relative strength of different forms of evidence that may be available on a topic.

BOX 1: FORMS OF EVIDENCE

Systematic Review – a systematic search is used to find all research studies conducted on a particular topic. The quality of the studies are reviewed, and an overall conclusion made.

Randomised Controlled Trial (RCT) – a study where the subjects are randomly selected to be involved and randomly allocated to treatment groups. There is also a control group for the treatment to be compared to.

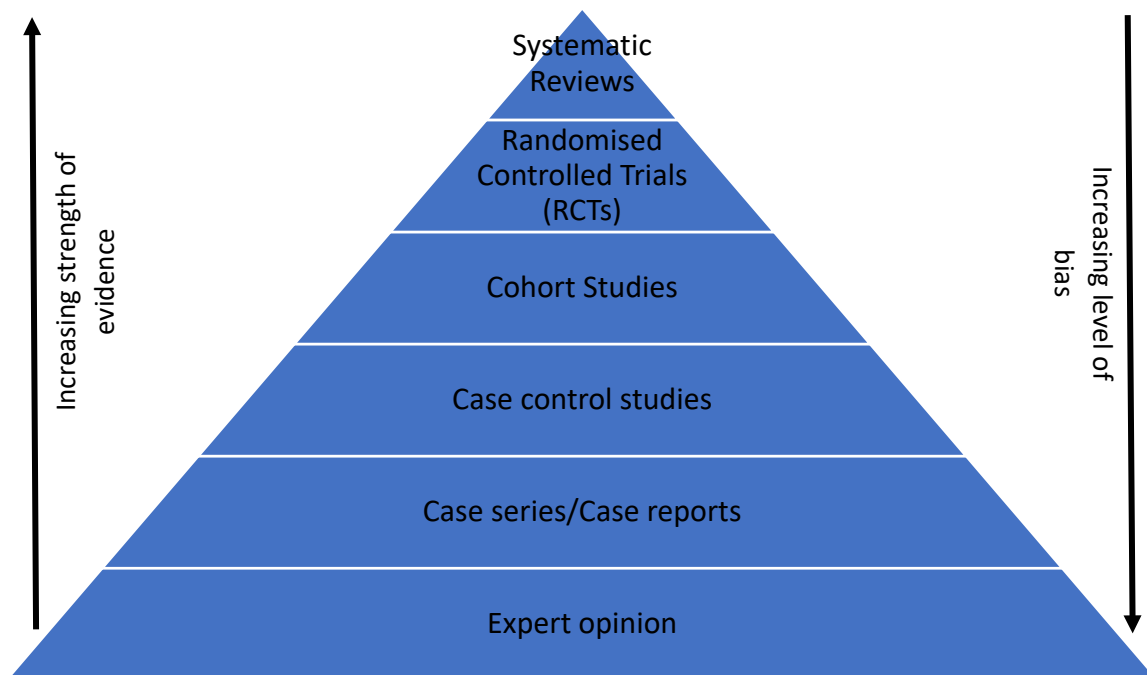
Cohort Studies – A longitudinal study which follows a group of subjects all with a similar characteristic i.e. particular disease diagnosis / particular breed characteristic. Outcomes are measured periodically over time.

Case Control Studies – Subjects are selected based on whether they have a specific characteristic or not such a as disease diagnosis. Retrospective information is collated to look for risk factors for that characteristic.

Case Report/Case Series – The details of the diagnosis, treatment and outcome of an individual case or a selection of similar cases is reported.

Expert Opinion – the opinion of someone with specialist knowledge in a particular field and who may possess further qualifications in that field.

Figure 1: Evidence Pyramid



(Cockcroft and Holmes, 2003)

STUDY DESIGN –

A research study can be broken down into many smaller elements that can impact on the validity and reliability of the outcomes reported. We can consider these elements to help us to evaluate the quality of the research presented and how applicable it is to the clinical situation we are considering.

Box 2 summarises some useful questions to consider when critiquing study design.

BOX 2: SUGGESTED CRITIQUE QUESTIONS

- What was the population? and was the sample representative of that population?
- How were the subjects selected? Was it random?
- How were the subjects allocated to groups? Was it random?
- Was there a control group?
- Were groups treated identically other than the intervention of interest?
- Was any potential for bias limited?
- Was the outcome measure subjective? Was blinding used?
- Were outcome measures reliable and valid?

What is the population of interest?, who are the results of the study applicable to?

A research study will observe a sample of a population with a view to inferring the results from that sample onto the population of interest. For the results of the study to be truly applicable to the population the sample must be representative (reflect the characteristics) of the population. In many cases you will find that researchers have set specific inclusion and exclusion criteria. These are specific criteria that the subject being studied must meet to be included, or if they meet a specific criterion they are excluded from involvement. This is often done to help reduce variation within a study or due to ethical considerations; a particular subgroup of the population may have the potential to suffer pain, distress or lasting harm if included in the study or by reducing the variation the sample size can be reduced meaning less subjects are required. However, it may impact on how applicable results from this study are to other cases. For example, Fritsch and Jewell (2015) conducted a study to look at the acceptance of therapeutic renal food in cats with chronic kidney disease (CKD) finding positive impacts of the diet and ease of transition onto the diet. Their study population included cats that had clinical signs of CKD however they needed to be in good health otherwise and not emaciated. Such exclusion criteria are understandable but, the results of this study would therefore not be truly applicable to cases where the cat's health was further deteriorated. Not to say that the diet would not be beneficial to the individual in this case, but you may not see the outcomes that the research has presented. This should be appreciated and communicated to owners as appropriate.

Sample Size

Sample size is important because if the sample size is too small then there may be a risk of not recognising a treatment effect where there in fact is one (Type II error) (Petrie and Watson 2013). This would have a considerable impact on the validity of a study's results. A good study will use statistical calculations to estimate appropriate minimum sample size that will reduce the chances of a Type II error. Further details on sample size calculations can be found in Mesquita et al. (2012).

Is there a control group?, is the only difference between groups the variable under investigation?, is there any confounding present?

A control group is a key element of good study design. It allows the intervention under investigation to be compared to the scenario where either no intervention is made (negative control group) or where an alternative intervention is made (positive control group). Changes will naturally occur over time irrespective of any intervention so without this comparable group it would be difficult to know if the new intervention was more, less or just as effective as other interventions. For example, Maruhashi et al. (2016) present a pilot study looking into the use of honey in the treatment of otitis externa. A positive response was found however, there was no control group to compare the effect of the honey against therefore it is not possible to know if it is better or as good as other treatments or whether the subjects would have improved over the same time without treatment anyway.

It is also important that the control group does not differ significantly from the intervention group(s) other than the aspect of care under investigation. If there were significant differences in the groups such as difference in the average age of the patients, severity of condition, breed representation or other additional treatments etc. then these differences could be the reason for any deviation in the outcomes shown and not as a result of the care being scrutinised. This is known as confounding; where the effect of one factor (intervention) cannot be separated from the effect of another factor (the difference in the groups) (Cockcroft and Holmes 2003).

Has any potential for bias been limited?

Bias can lead to distortion of a study's results (Petrie and Watson 2013) so it is crucial to look for any opportunity for bias (conscious or unconscious) within a study's design. It may not be feasible to eliminate bias entirely, but the researcher should have made steps to reduce bias where possible.

There is the potential for bias to be introduced to a study when the subjects are selected and when they are allocated to treatment groups. An example of selection bias is seen in a study conducted by

Ortiz et al. (2018) which considered the effect of including metronidazole alongside amoxicillin-clavulanate in the treatment of haemorrhagic diarrhoea in canines. After the study was completed the researchers noted that the overall severity of the cases (measured in days to discharge from the veterinary clinic) included in the study was less than had been measured during their piloting. It was recognised that selection of patients put forward to be involved in this trial was being made by the consulting veterinarians and had resulted in bias towards less severe cases. Consequently, this would impact on the application of these results to more severe cases.

Random selection of subjects and random allocation of subjects to the different treatment groups should eliminate chance of bias but it is also useful to check baseline information for subjects (if presented) to check for consistency between groups.

Another point of a study where bias can be introduced is the measurements taken if they are subjective (such as pain scores, stress scores etc.). Such measures have the potential to be influenced by the person taking the measurement and this is more likely where the person taking the measurement is aware of the intervention that has been made (or not made). The potential for this can be eliminated by 'blinding' the person to the intervention(s). An example of this can be seen in a study by Hamilton et al. (2014). The study considered the impact of surgical approach to orchidectomy on development of post-operative complications. Wound observations were made by veterinary nurses who were unaware of the surgical approach that had been used; they were 'blinded' to the treatment.

It may also be considered appropriate to blind owners to the intervention(s) applied if they may for example adapt the way in which they interact with their pet or if they are involved in outcome measurements such as client questionnaires. Due to the nature of some interventions blinding may not always be possible but should be implemented if feasible. The study by Fritsch and Jewell (2015) discussed earlier that investigated the impact of therapeutic diet on cats with CKD used blinding of owners; they were not aware of the nature of the diet that their pet was being transitioned onto. In

this study blinding of owners was considered important as they were asked to assess their pet's quality of life. The study was however lacking a control group, given an appropriate placebo diet, that would have allowed greater appreciation of the relative impact of the therapeutic diet. This is because any change in diet may prompt owners to perceive change in their pet even if they are unaware of the intended effect of the diet.

Where measurements are objective (such as weights, heart rates, blood analysis etc.) these cannot be influenced by the observer. Therefore, measurement bias is not a concern and 'blinding' is not necessary.

Are outcome measures valid and reliable?

Reliability is concerning how repeatable a measure is; if the measurement were to be taken on multiple occasions would you get the same every time. When assessing study design, you should consider whether the equipment (if any) used to measure the outcome is reliable or if there is the potential for technical errors? or human errors? The researcher may have addressed some of these points in the study design through calibrating equipment, ensuring consistent use of the same piece of equipment, consistent measuring techniques and using the same individual to take measurements.

Validity concerns whether a measure is a true reflection of the outcome it intends to measure. For example are lameness scores a true reflection of the mobility of the animal or is a pain score a true reflection of the level of pain an animal is experiencing. For subjective measures this can be difficult to assess however there are several research studies that have considered the validity (and reliability) of some commonly used subjective measures such as Hudson et al. (2004), Walton et al. (2013) and Shipley et al. (2019). When evaluating studies that have used subjective measures you should consider if those measures have been validated.

Where a measure is objective you can still consider if it is a valid measure or not. Such as the study presented by Ortiz et al. (2018) looking at the impact of metronidazole on cases of haemorrhage diarrhoea in canines. The outcome measure used in this case was number of days to discharge from the veterinary clinic. When evaluating the study you should consider is this a suitable measure to reflect the impact of metronidazole on these patients?

UNDERSTANDING THE RESULTS

It is outside of the remit of this article to discuss the wide variety of statistical appraisals you may see presented by researchers. However, a common feature of statistical analysis is the use of the P value to convey the significance of any results found so an understanding of the interpretation of this value is important. The P value is a probability value and conveys the probability of getting the results found by chance alone. Therefore, the smaller the P value the smaller the likelihood that the results could be by chance and not because of the intervention applied. It is standard practice to consider a P value of less than 0.05 (5%) as significant; meaning there is less than 5% probability that results are by chance. It is essential that results found are interpreted alongside evaluation of the study design though because, as this article has discussed, this can impact on the reliability and validity of the results.

Conclusion:

EBVM is the use of current best evidence in making clinical decisions. Although limitations may be recognised in a research study or a form of evidence, it may still provide you with the 'current best evidence' on that topic. As a professional you should use your skills in research evaluation to assess the strengths and weaknesses of that evidence and help you utilise it in an appropriate manner.

References

Cockcroft PD, Holmes MA. 2003. Handbook of evidence based veterinary medicine. Oxford:Blackwell

Frisch DA, Jewell DE. 2015. Acceptance and effects of a therapeutic renal food in pet cats with chronic kidney disease. *Vet Rec Open*. 2(2)

Hamilton KH, Henderson ER, Toscano M, Chanoit GP. 2014. Comparison of postoperative complications in healthy dogs undergoing open and closed orchidectomy. *J Small Anim Pract*. 55(10):521-526

Hudson JT, Slater MR, Taylor L, Scott M, Kerwin C. 2004. Assessing repeatability and validity of a visual analogue scale questionnaire for use in assessing pain and lameness in dogs. *American Journal of Veterinary Research*. 65(12):1634-1643.

Maruhashi E, Braz BS, Nunes T, Pomba C, Belas A, Henrique J, Correlá D, Lourenço AM. 2016. Efficacy of medical grade honey in the management of canine otitis externa-a pilot study. *Vet Dermatol*. 27(2):93-98

Mesquita JR, Nobrega C, Vala H, Sousa SIV. 2012. Statistics in veterinary nursing research: what to know before starting the study. *The Veterinary Nurse*. 3(10):594-598.

Ortiz V, Klein L, Channell S, Simpson B, Wright B, Edwards C, Gilbert R, Day R, Caddy SL. 2018. Evaluating the effect of metronidazole plus amoxicillin-clavulanate versus amoxicillin-clavulanate alone in canine haemorrhagic diarrhoea: a randomised controlled trial in primary care practice. *J Small Anim Pract*. 59(7):398-403

Petrie A, Watson PF. 2013. *Statistics for veterinary and animal science*. 3rd edition. Oxford:Wiley-Blackwell

Shipley H, Guedes A, Graham L. 2019. Preliminary appraisal of the reliability and validity of the collarado state university feline acute pain scale. *J Feline Med Surg*. 21(4):335-339.

Walton MB, Cowderoy E, Lascelles D, Innes JF. 2013. Evaluation of construct and criterion validity for the 'Liverpool Osteoarthritis In Dogs' (LOAD) clinical metrology instrument and comparison to two other instruments. Plos One. 8(3)