Equine Veterinary Education

EQUINE VETERINARY EDUCATION
*Equine vet. Educ.* (2018) •• (••) ••-••
doi: 10.1111/eve.12896



*Review Article*

# How to critically appraise a paper

***G. L. Pinchbeck* and D. C. Archer***

*Department of Veterinary Epidemiology and Population Health, Institute of Infection and Global Health and Institute of Veterinary Science, Faculty of Health and Life Sciences, University of Liverpool, Leahurst Campus, Neston CH64 7TE, UK*
**Corresponding author email: Ginap@liv.ac.uk*

**Keywords:** horse; critical appraisal; study design; bias; confounding; validity; evidence-based veterinary medicine

## Summary

**Critical appraisal skills enabling assessment of the validity and importance of evidence are essential for clinicians to make informed decisions regarding what new information should be incorporated into their clinical practice. This review highlights key points to consider in a critical review concentrating on common study designs used in the equine literature.**

## Introduction

A vast number of veterinary papers are published every year in a variety of journals. Although many of these are in reputable journals, and have been through a process of peer-review by experts, the quality may still be variable (Pocock *et al.* 2004). In addition, there are many sources of nonpeer-reviewed literature, including textbooks, reports and proceedings and the internet. Formal systematic reviews are beginning to appear in the equine veterinary literature (Allen *et al.* 2012; Sullivan *et al.* 2015; Dominguez *et al.* 2016), and RCVS Knowledge and the Equine Veterinary Education are beginning to publish knowledge summaries (or critically appraised topics) (RCVS Knowledge 2017). However, it is usually up to the individual reader to assess the scientific validity, strength of evidence and practical relevance of results presented in a paper and the extent to which they can be applied to the particular question they are interested in. To assist the process of critical appraisal a number of veterinary and non-veterinary organisations have online resources and checklists that can be used as aids (CASP 2017; CEBM 2017; CEVM 2017; EBVMA 2017). More detailed checklists exist for examination of different specific study types such as randomised controlled trials (RCT), validating diagnostic or screening tests, qualitative studies and systematic review or meta-analyses (Greenhalgh 2001c; CASP checklist 2017; CEBM 2017). In addition, a number of reporting guidelines have also been developed for various study designs (More 2010), many of which can be found on the EQUATOR network website (www.equator-network.org). Specific guidelines, with checklists, widely in use include those for RCTs (CONSORT) and for strengthening the reporting of observational studies (STROBE) (von Elm *et al.* 2008) available at https://www.strobe-statement.org/index.php?id=available-checklists.

Evidence-based medicine, defined by Sackett *et al.* (1996) as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients", is much more than just critically assessing papers. However, critical appraisal is a key skill that must be mastered to practice evidence-based medicine. The Royal College of Veterinary Surgeons recognises "*How to evaluate evidence*" as an essential day one competence required of all veterinary surgeons (RCVS) and "*critical analysis of new information and research findings relevant to veterinary medicine*" is a core competency for accreditation by the American Veterinary Medical Association (Anon AVMA). These requirements recognise the essential role of critical appraisal for clinicians' decision-making regarding what new information, among the vast array available, should be incorporated into their clinical practice.

## Critical appraisal

Critical appraisal is a formal, unbiased, systematic approach to assessing the quality and relevance of evidence presented in a paper and its applicability to decision making for our patients. It includes evaluation of the appropriateness of the study design for the research question, and a careful assessment of whether the study conforms to specific criteria, related to the study design. It should be a balanced assessment of benefits and strengths of research against its flaws and weaknesses. The assessment of methodological quality should be done without consideration (or even knowledge) of the results, so as to avoid interpretation bias (Kaptchuk 2003), which arises because interpretation of findings is rarely completely independent of our previous beliefs or preconceptions. For example, higher standards of evidence may be required in the case of study findings that contradict an individual's initial expectations, compared with a study that agrees with these expectations (confirmation bias).

For the equine clinician with limited epidemiological or statistical knowledge, it can be difficult to critically appraise the study design, statistical analyses used and whether the conclusions drawn can be justified based on the material presented by the authors. The following sections highlight key points that should be considered in a critical review (**Table 1**), concentrating on randomised controlled trials and observational studies, which are the most common study designs used in the equine literature.

### *What is the main purpose of the study?*

This includes the clinical question the study seeks to address and what hypothesis is being tested. Not all research studies aim to test a single definitive hypothesis and qualitative research studies investigate particular issues in a broad, open-ended way (Christley and Perkins 2010). The relevance of the study question and whether it measures an outcome that is relevant to your clinical practice, e.g. return to athletic

**TABLE 1: Checklist to aid critical appraisal of a paper**

| | | |
|---|---|---|
| 1. What is the main purpose of the study | | What was the main aim/hypothesis of the study?<br>What was the exposure or intervention?<br>What was the outcome and how was it measured?<br>What were the main results?<br>What was the study population? What, where, when, inclusion/exclusion criteria? |
| 2. What type of study design was used | | What was the study design and was this appropriate? |
| 3. Is the study internally valid? | Bias and confounding | Are the results likely to be affected by selection/sampling bias?<br>Are the results likely to be affected by observation/measurement bias?<br>If an intervention trial, were the assessments of outcomes blinded?<br>Are the results likely to be affected by confounding?<br>If a clinical trial, was the allocation adequately randomised? |
| | Statistical methods | Is there sufficient statistical power?<br>Are appropriate statistical methods used and results interpreted appropriately? |
| | Causality | Can a temporal relationship be ascertained?<br>Is the relationship important/strong?<br>Is there a dose–response relationship?<br>Can the results be explained by noncausal explanations? |
| 4. External validity | | Can the study results be applied more widely to other populations and to the population under your care?<br>Were all clinically important outcomes considered?<br>Are the outcomes assessed of relevance to your patients? |
| 5. Other factors | | Who are the authors and is there any potential for conflict of interest?<br>Are the results consistent with other evidence?<br>Did the study have appropriate ethical approval? |

function or reduced mortality, and whether it adds anything new to the literature is also important to consider.

Most research studies will evaluate one of the following (Greenhalgh 2001a): Therapy (efficacy of a drug treatment, surgical procedure or other intervention); Causation (if a suspected risk factor is related to development of a particular disease); Prognosis (outcome of a disease following treatment/diagnosis); Diagnosis (the validity and reliability of a new diagnostic test and superiority to any existing tests) or Screening (tests applied to a population to detect disease).

### *What type of study design was used (and is this the most appropriate for the question addressed)?*

Understanding the type of study that has been performed is a prerequisite to evaluation of the strength of evidence provided by the study. In addition, knowledge of the design will enable the reader to determine if the study in question has been appropriately designed and conducted (assessing the internal validity), and, if not, whether this should decrease the strength of belief in the results. Sometimes, authors will state that they have performed a particular study design but careful reading of the methods may contradict this.

The hierarchy of evidence (**Fig 1**), which ranks the relative strength of evidence carried by the different types of study when making decisions about clinical interventions, is well recognised. The pinnacle of the hierarchy is reserved for papers in which all the primary studies on a clinical question or subject are critically appraised according to rigorous criteria (the systematic review) and meta-analysis, which integrates the numerical data from more than one study. The latter studies are relatively infrequent in equine veterinary research (Calzetta *et al.* 2017), but are likely to become more common in the future as a greater number of high quality primary clinical research studies (particularly randomised controlled trials) are reported. Primary clinical research studies can be experimental or observational. In experimental studies, such as the RCT, the investigator controls the allocation of the intervention (e.g. a new treatment vs. no treatment or an existing treatment) to a randomly selected subset of the study subjects and then compares between groups of study animals to make inferences about the effect of the intervention on the outcome of interest. The RCT can provide strong evidence and is often regarded as the highest standard of evidence to guide clinical decision-making due to key aspects of design that aim to avoid bias. Although still infrequent in equine clinical research, RCTs are becoming more common (Sabate *et al.* 2009; Talbot *et al.* 2013; Higler *et al.* 2014). The observational studies (cohort, case control and cross-sectional) are next in the hierarchy, and these are relatively frequently used in the equine research literature. Descriptive studies such as case reports and case series are very common in the equine research literature. However, these are not designed to test an association between a therapy and a treatment (Grimes and Schulz 2002). If an author draws conclusions about the merits of a particular therapy from a case report or series, this should be regarded as the author's opinion/conjecture only, as these provide limited evidence to support this.

### *Is the study internally valid?*

Internal validity is concerned with the quality of the study as it applies to the population that is being studied, and its assessment involves asking the question: did the researchers

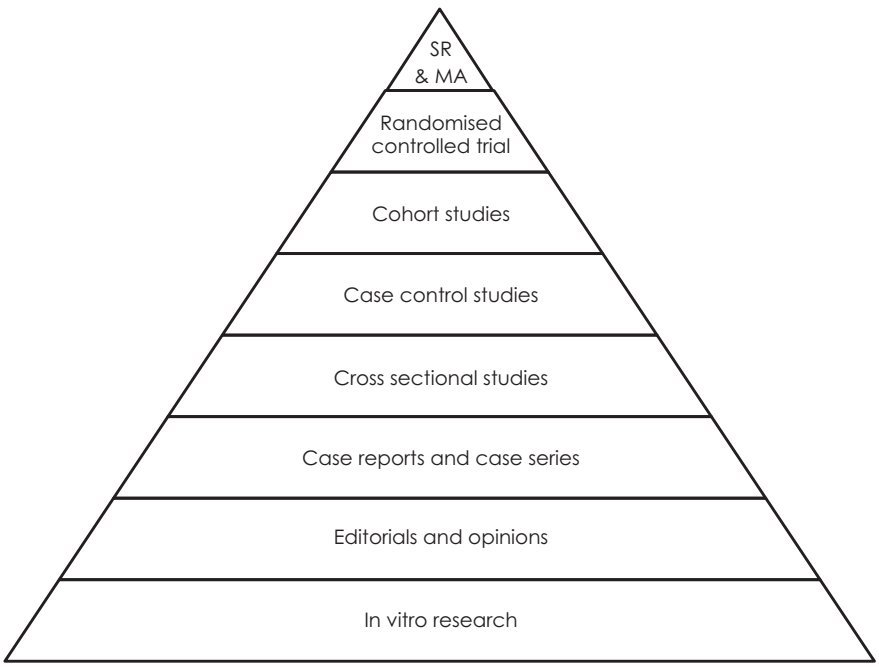


***Fig 1: Hierarchy of evidence. SR & MA = Systematic reviews and meta-analyses. Adapted from: SUNY Downstate Medical Center. Medical Research Library of Brooklyn. Evidence Based Medicine Course. A Guide to Research Methods: The Evidence Pyramid: http://library.downstate.edu/EBM2/2100.htm].***

do things properly? Once you have established that the paper addresses a relevant clinical question, and that an appropriate study design has been used, the methodology should be critically appraised for quality and the strength of evidence of the work presented. This should consider whether potential sources of bias or confounding were addressed, whether the study was performed according to the original protocol and if appropriate statistical methods were performed correctly. In some instances, the reader may be presented with insufficient information to appraise the reliability and such papers should be interpreted with caution.

*Bias and confounding*

Many of the aspects of design which need to be considered to determine internal validity of a study are concerned with assessing whether systematic bias has been avoided or minimised throughout the study. Bias can be defined as the *systematic (nonrandom) error in design, conduct or analysis of a study resulting in mistaken estimates*, and different study designs require different steps to reduce bias (see next section). Bias can occur due to the way populations are sampled, data are collected, or are analysed. Unlike random error, increasing the sample size will not decrease systematic bias. There are numerous types of bias (Sackett 1979; Dohoo et al. 2009) but these can be considered under three headings:

- Selection bias: This bias occurs when the composition of study subjects or participants in a research project systematically differs from the source population. There are many different sub-types of this type of bias including choice of comparison groups, nonresponse bias (e.g. respondents differ compared with those who do not respond in a questionnaire study), follow-up bias (loss to follow-up is different between groups being compared), selective entry bias (e.g. use of horses that are currently racing is biased towards a healthy horse population) and detection bias (controls wrongly classified when they have the disease of interest because, for example, they did not receive the same examination protocol as cases).
- Information bias: Occurs when the outcomes, exposures of interest (factors measured) or other data are incorrectly classified or measured. This might be, for example, due to use of poor diagnostic criteria or tests, or differing application of data collection techniques or tests depending on the outcome or exposure status of the subject (for example, administering a questionnaire face-to-face for cases but by post for controls)
- Confounding bias: This is the mixing of the effects of two or more factors. We might think that we are measuring the association between an exposure factor and an outcome, but the association we observe actually includes the effect of one or more other variables. Hence, our assessment of the association between the exposure factor and the outcome is biased (or confounded). For example, if we wished to assess whether transport, as an exposure factor, was associated with the risk of colic, we would need to take into account the effects of one or more potential confounding factors (e.g. change of forage or change in time at pasture), that are hypothesised risk factors for colic but are also likely to be associated with transport. To avoid bias and get an accurate estimate of the effect of transport on colic, these confounding factors must be taken into account. Once you have identified the study type and considered the exposures and outcome factors, you can make a list of possible confounders and then examine the paper to see how the authors have dealt with these (if at all).

Particular types of bias may be more likely with specific study designs, as outlined in the next section. Specific points

that should be considered in the appraisal of these studies are also summarised (adapted from Young and Solomon 2009):

*Randomised controlled trials (RCTs).* The RCT is a prospective study designed to assess the effect of one (or more) treatments or interventions compared with a control group that may consist of no treatment, a placebo or a comparator treatment (for example, a current standard therapy). Key design aspects that provide the best means of avoiding bias include the process of random allocation to treatment groups, which aims to ensure that treatment groups are equivalent in terms of both known and unknown confounding factors and hence any differences in outcomes can therefore be ascribed to the effect of treatment. The process of blinding in RCTs, whereby participants and those who are assessing the outcomes (and in some cases also those analysing the data) are unaware of intervention assignment, reduces information or measurement bias. The CONSORT guidelines (Anon 2017) provide a comprehensive checklist specifically for RCTs but specific points for consideration should include:

- Was the process of treatment allocation properly described and truly random?
- Were the groups comparable in all important aspects except for the variable being studied?
- Were participants and researchers 'blinded' to participants' treatment group?
- Were primary and secondary outcome measures, properly defined and objectively assessed?
- Were all participants who were randomly allocated a treatment accounted for in the final analysis?

Papers describing comparative studies in which subjects are allocated to intervention or control groups in a nonrandom manner are not randomised trials, and can be termed "*other controlled clinical trials*" (Greenhalgh 2001a). There is a high risk of bias and confounding in these types of studies. For example, if a horse is not randomised to a particular therapy or surgical procedure, a clinician may decide upon the therapeutic/surgical treatment of that patient based on particular characteristics or specific features of the disease, such as severity (leading to selection bias); hence, any difference in outcome between the groups may relate to biased selection (grouping) of the patients themselves rather than the therapy/surgical intervention performed. If the paper you are looking at is a nonrandomised controlled clinical trial, you must use your judgement to decide if the baseline differences between the groups are likely to have been so great as to undermine claims regarding any differences (or lack of differences) ascribed to the intervention.

*Cohort studies.* Cohort studies involve the follow-up of study participants with varying exposures forward, to observe which animals develop the outcome(s) of interest (e.g. clinical disease or mortality), and to determine the effect of exposures on the outcome. These studies are commonly prospective, where data are specifically collected for the purposes of the study, but may also be retrospective in nature, using data that have been routinely collected for another purpose e.g. clinical or race records. There is a potential for confounding in these types of studies and for bias due to loss to follow-up. Where potential confounding factors are identified prior to prospective cohort studies, these can be measured and taken into account in the analysis, but in retrospective studies this may be difficult or impossible if these factors were not measured or were measured poorly or inconsistently. Your checklist for these studies should consider:

- Is the study prospective or retrospective?
- Is the cohort representative of a defined group/ population?
- Were all important confounding factors identified, adequately measured and adjusted for in the design or analysis?
- Were all important exposures and outcomes measured accurately, objectively and equivalently in all the cohort subjects (including irrespective of exposure status)?
- Were there any losses to follow-up and could this bias the findings?

*Case–control studies.* Case–control studies are common in the equine literature and are ideal for investigation of risk factors where the outcome of interest (e.g. a specific disease) is rare and it would be impractical to perform a prospective cohort study (e.g. too many horses needed/would take too long to perform). In case–control studies, cases are only recruited once they have developed the outcome of interest and exposure data are then evaluated to determine whether exposures of interest differ between the cases and controls. Appropriate control selection is often the most difficult aspect of a case–control study as controls must come from the same population of cases to avoid bias. Controls should be subjects who would have entered the study and become cases if the outcome had occurred. Exposure data are often collected by questionnaire or from existing records. Recall bias is another potential issue in these types of studies where owners/ carers cannot remember exact details about historical factors, and is a particular concern where this may differ between groups (e.g. owners of horses who developed the disease of interest may recall specific factors more accurately than owners of control horses). In addition, recording of subjective rather than objective data may result in different responses from case or control owners leading to bias. Specific points to consider in these studies are therefore:

- Were cases clearly defined and eligibility and ascertainment described?
- Were controls appropriately selected and drawn from the same population as the cases, using the same eligibility criteria?
- Were exposures measured equivalently for cases and controls?
- Is recall bias likely?
- Were all important confounding factors identified and adjusted for in design (for example by matching) or analysis?

*Cross-sectional studies.* These studies provide a 'snapshot' in time to determine information (for example, disease prevalence) about a population of interest. Concurrent measurement of exposures of interest may also be used to

elucidate factors associated with the outcome. Specific aspects of study design that should be considered include:

- Were the eligibility criteria and methods of selection of the study sample clearly defined?
- Was a representative sample obtained (e.g. sufficiently high response rate)
- Were all relevant exposures, potential confounders and outcomes measured accurately?
- Is it certain that the exposure of interest occurred prior to development of the outcome?

*Statistical methodology*

Although it is unlikely that most readers will be able to critically evaluate every type of statistical analysis that may be used in clinical studies, it is still important to be able to critique key issues. If the statistical tests in the paper are uncommon, the authors should provide justification why they have used them and describe them in detail or include a relevant reference.

For many studies, and in particular RCTs, a sample size calculation is a crucial prerequisite. The sample size should be clearly defined and justified using appropriate calculations and should be big enough to have a high chance of detecting, as statistically significant, a worthwhile effect if it exists. If ultimately the study did not then reach the described sample size, then it may be underpowered and lead to erroneous conclusions.

The statistical methods used for any comparisons of the data should be clearly stated and any key assumptions, and evidence that these have been met, should be described (e.g. assumptions of normal distributions, multiple testing). Statistical tests are either parametric (i.e. they assume that the data were sampled from a normal distribution) or nonparametric (i.e. they do not assume that the data were sampled from any particular distribution). Inappropriate statistical tests commonly used include use of parametric tests when data are not normally distributed, conducting multiple testing and ignoring clustered data. Greenhalgh (2001b) provides further details, including some frequently used statistical 'tricks' you should be aware of.

The results section should include sufficient description of the data to enable readers to understand how the authors arrived at their conclusions. This usually entails provision of point estimates of effect (or difference) and measures of variation (usually the 95% confidence interval). The Equine Veterinary Journal provides useful guidelines and a statistical checklist (Christley 2015) but specific points to consider include:

- Are groups comparable and, if necessary, adjusted for baseline differences?
- Have the data been analysed according to the original protocol?
- Are statistical tests appropriate for the types of data described?
- If the statistical tests in the paper are uncommon, why have the authors chosen to use them?
- Have confidence intervals been calculated and do the authors' conclusions reflect these?
- Have assumptions been made about the nature and direction of causality? Remember statistical association does not provide direct evidence of causation.

*Strength of evidence and causality*

Other factors that should be considered include an interpretation of whether the results infer causality (Hill 1965):

- *Temporal relationship* – does the study demonstrate that the exposure/intervention preceded the disease/outcome of interest? It is often not possible to demonstrate this in cross-sectional or case–control studies.
- *Strength of association* – Is the relationship sufficiently 'strong' to be clinically/biologically important? This can usually be evaluated by looking at the size of ratio measures (e.g. odds/risk ratio or differences in risk/rate of disease) and the precision of the estimate of effect. Whether the effect would still be sufficiently strong at either end of the 95% CI should be considered.
- *Dose response* – does the risk of disease or outcome of interest increase with increasing levels of the exposure? The most commonly cited example of this is in human medicine and the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily. Demonstrating a strong dose–response relationship adds a great deal to the simpler evidence that an exposure increases the risk of disease.
- *Noncausal explanations* – does the association between an exposure and outcome of interest make biologically and epidemiological sense (and is it consistent with other studies or what is known about disease mechanisms) or could this relationship be due to something else?

### *How externally valid is the study and is it applicable to patients under your care?*

Once you have determined a study is of good quality (i.e. internally valid), the next step is to evaluate if the study has external validity. External validity relates to how well the results can be generalised to other populations and in particular your target population, i.e. patients under your care. The horses studied may differ from those under your care if, for example, they had more or less co-morbidities; they were of different breed or under very different management routines. In addition, you should consider if the outcomes assessed are of relevance to your patients, or if clinically relevant outcomes have been ignored. Remember, however, a study can only have external validity if it is also internally valid.

### *Other factors to consider*

The findings of the study should then be considered together with other evidence to determine whether they are consistent with other studies or whether the findings contradict previous studies. It is also important to consider whether the authors have any potential conflicts of interest and, if so, whether these have been recognised. For example, commercial funding of a study might lead to a conflict of interest (e.g. if there is potential for increased sales of a commercial product based on study findings, any negative findings may not be presented). If potential conflicts of interest exist this should be addressed e.g. for commercially funded research there should be a statement about whether the study design, analysis and interpretation of results was performed independently of the funding body. In addition, the reader should consider whether the study has been performed with appropriate ethical approval (Bertone 2013), including informed owner consent, although ethical

guidelines and standards (including for reporting) do vary between different countries and journals, and in some settings local institutional ethics committees may not exist.

## Conclusion

Ultimately, it is often up to the reader to determine the extent to which s/he feels the results presented in a paper are reliable and if, and how the findings may potentially benefit patients under their care. There is no single tool that can be used to perform critical appraisal of all types of study, however there are numerous online resources and checklists that can be used as an aid. The critical appraisal checklist (**Table 1**) is not exhaustive but can be used as a guide to identify whether the research documented in a paper conforms to certain criteria that are important in epidemiological studies.

## Authors' declaration of interests

No conflicts of interest have been declared.

## Ethical animal research

Not applicable to this review article.

## Source of funding

None.

## Authorship

Both authors contributed to the writing and approved the final manuscript.

## References

Allen, K.J., Christley, R.M., Birchall, M.A. and Franklin, S.H. (2012) A systemic review of the efficacy of interventions for dynamic intermittent dorsal displacement of the soft palate. *Equine Vet. J.* **44**, 259-266.

American Veterinary Medical Association. https://www.avma.org/ProfessionalDevelopment/Education/Accreditation/Colleges/Pages/coe-pp-requirements-of-accredited-college.aspx. Accessed 09/08/2017

Bertone, J. (2013) Ethical, scientific and practical standards in clinical practice research. *Equine Vet. J.* **45**, 391-393.

Calzetta, L., Roncada, P., di Cave, D., Bonizzi, L., Urbani, A., Pistocchini, E., Rogliani, P. and Matera, M.G. (2017) Pharmacological treatments in asthma-affected horses: a pair-wise and network meta-analysis. *Equine Vet. J.* **49**, 710-717.

Centre for Evidence-Based Medicine. (2017) http://www.cebm.net/critical-appraisal/ Accessed: 14/7/2017

Centre for Evidence-based Veterinary Medicine, University of Nottingham. (2017) http://www.nottingham.ac.uk/cevm/evidence-synthesis/resources.aspx Accessed 10/8/2017

Christley, R. (2015) Statistical guidelines for Equine Veterinary Journal. *Equine Vet. J.* **47**, 131-132.

Christley, R.M. and Perkins, E. (2010) Researching hard to reach areas of knowledge: qualitative research in veterinary science. *Equine Vet. J.* **42**(**4**), 285-286.

CONSORT – Consolidated Standards Of Reporting Trials. (2017) http://www.consort-statement.org/ Accessed 20/6/2017

Critical Appraisal Skills Programme. (2017) CASP Checklists. Available at http://www.casp-uk.net/casp-tools-checklists/ Accessed: 14/7/2017.

Dohoo, I., Martin, W. and Stryhn, H. (2009) Validity in observational studies. *Veterinary Epidemiological Research*, VER Inc., Charlottetown, Prince Edward Island, Canada. pp 243-279.

Dominguez, M., Munstermann, S., De Guindos, I. and Timoney, P. (2016) Equine disease events resulting from international horse movements: systematic review and lessons learned. *Equine Vet. J.* **48**, 641-653.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandenbroucke, J.P. (2008) STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. "http://download.journals.elsevierhealth.com/pdfs/journals/0895-4356/PIIS0895435607004362.pdf" \t "_blank" \o "Opens external link in new window" *J. Clin. Epidemiol.* **61**, 344-349.

Equine Vet. J. Clinical Evidence in Equine Practice online collection. Edited By: C. M. Marr. http://onlinelibrary.wiley.com/journal/10.1001/(ISSN)2042-3306/homepage/clinical_evidence_in_equine_practice_online_collection.htm Accessed: 17/8/2017

Evidence Based Veterinary Medicine Association (2017) https://ebvma.org/continuing-education/ Accessed: 10/8/2017.

Greenhalgh, T. (2001a) Chapter 3: getting your bearings (what is this paper about?). *How to read a paper: the basics of evidence based medicine*, 2nd edn., B.M.J. Books, London. pp 39-58.

Greenhalgh, T. (2001b) Chapter 5: statistics for the non-statistician. *How to read a paper: the basics of evidence based medicine*, 2nd edn., B.M.J. Books, London. pp 76-93.

Greenhalgh, T. (2001c) Appendix 1: checklists for finding, appraising, and implementing evidence. *How to read a paper: the basics of evidence based medicine*, 2nd edn., B.M.J. Books, London. pp 200-209.

Grimes, D.A. and Schulz, K.F. (2002) Descriptive studies: what they can and cannot do. Epidemiology Series. *Lancet* **359**, 145-148.

Higler, M.H., Brommer, H., L'Ami, J.J., de Grauw, J.C., Nielen, M., van Weeren, P.R., Laverty, S., Barneveld, A. and Back, W. (2014) The effects of three-month oral supplementation with a nutraceutical and exercise on the locomotor pattern of aged horses. *Equine Vet. J.* **46**, 611-617.

Hill, A.B. (1965) The environment and disease: association or causation. *Proc. R. Soc. Med.* **58**, 295-300.

Kaptchuk, T.J. (2003) Effect of interpretive bias on research evidence. *BMJ* **326**, 1453-1455.

More, S. (2010) Improving the quality of reporting in veterinary journals: how far do we need to go with reporting guidelines? *Vet. J.* **184**, 249-250.

Pocock, S.J., Collier, T.J., Dandreo, K.J., de Stavola, B.L., Goldman, M.B., Kalish, L.A., Kasten, L.E. and McCormack, V.A. (2004) Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* **329**, 883-887.

RCVS Knowledge (2017) https://veterinaryevidence.org/index.php/ve/announcement/view/4 Accessed 14/7/2017.

Royal College of Veterinary Surgeons, Day One Competences statement. Essential competences required of the veterinary surgeon pp 3. http://www.rcvs.org.uk/document-library/day-one-competences-statement: Accessed: 10/8/2017

Sabate, D., Homedes, J., Salichs, M., Sust, M. and Monreal, L. (2009) Multicentre, controlled, randomised and blinded field study comparing efficacy of suxibuzone and phenylbutazone in lame horses. *Equine Vet. J.* **41**, 700-705.

Sackett, D.L. (1979) Bias in analytic research. *J. Chronic Dis.* **32**, 51-63.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B. and Richardson, W.S. (1996) Evidence based medicine: what it is and what it isn't. *BMJ* **312**, 71-72.

Sullivan, S.L., Whittem, T., Morley, P.S. and Hinchcliff, K.W. (2015) A systematic review and meta-analysis of the efficacy of furosemide for exercise-induced pulmonary haemorrhage in Thoroughbred and Standardbred racehorses. *Equine Vet. J.* **47**, 341-349.

Talbot, W.A., Pinchbeck, G.L., Knottnebelt, D.C., Graham, H., and Mckane, S.A. (2013) A randomised, blinded, crossover study to assess the efficacy of a feed supplement in alleviating the clinical signs of headshaking in 32 horses. *Equine Vet. J.* **45**, 293-297.

Young, J.M. and Solomon, M.J. (2009) How to critically appraise an article. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **6**, 82-91.