BMJ Open

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011458
Article Type:	Research
Date Submitted by the Author:	15-Feb-2016
Complete List of Authors:	Downes, Martin; Griffith University School of Medicine, Centre for Applied Health Economics Brennan, Marnie; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science Williams, HC; University of Nottingham, Centre of Evidence Based Dermatology Dean, Rachel; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Medical education and training, Patient-centred medicine, Research methods
Keywords:	Cross sectional studies, Critical appraisal, Evidence-based Healthcare, Delphi

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Martin J. Downes[1¶] Marnie L Brennan[2¶], Hywel C Williams[3], Rachel S Dean[2*¶]

[1]Centre for Applied Health Economics, School of Medicine, Griffith University, Meadowbrook, Australia

[2]Centre for Evidence-based Veterinary Medicine, The University of Nottingham, Sutton Bonington Campus, United Kingdom.

[3]Centre of Evidence Based Dermatology, The University of Nottingham, Kings Meadow Campus, Lenton Lane, Nottingham, United Kingdom.

* Email: Rachel.Dean@nottingham.ac.uk; (RSD)

[¶]These authors contributed equally to this work

ABSTRACT

Objectives: The aim of this study was to develop a critical appraisal tool that addressed study design and reporting quality as well as the risk of bias in cross sectional studies. In addition the aim was to produce a help document to guide the non-expert user through the tool.

Design: An initial scoping review of the published literature and key epidemiological texts was undertaken prior to the formation of a Delphi panel to establish key components for a critical appraisal tool for cross sectional studies. A consensus of 80% was required from the Delphi panel for any component to be included in the final tool.

Results: The Appraisal tool for Cross-Sectional Studies (AXIS tool) was developed consisting of 20 components agreed upon by consensus of the Delphi panel. A detailed explanatory document was also developed with the tool, giving expanded explanation of each question and providing simple interpretations and examples of the epidemiological concepts being examined in each question to aid non-expert users.

Conclusion: Critical appraisal of the literature is a vital step in evidence synthesis and therefore evidence-based decision making in a number of different disciplines. Until now there has been no critical appraisal tool available to aid the reader in the assessment of study design and reporting quality as well as the risk of bias in this type of study. The AXIS tool is therefore unique and was developed in a way that it can be used across disciplines to aid the inclusion of cross sectional studies in systematic reviews, guidelines and clinical decision making.

Strengths and Limitations of this study

Strengths of this study include:

- The development of a novel critical appraisal tool that can be used across discipline
- A multimodal evidence-based approach was used to develop the tool
- Expertise was harnessed from a number of different disciplines

Limitations of this study:

- The Delphi panel was based on convenience and may not encompass all eventual users of the tool.
- A numerical scale to reflect quality was not included in the final tool, which may be perceived as a limitation.

INTRODUCTION

Critical appraisal (CA) is a skill central to undertaking evidence based practice which is concerned with integrating the best external evidence with clinical care. When reading any type of evidence, being critical of all aspects of the study design, execution and reporting is vital for assessing its quality before being applied to practice (Cockcroft PD, 2003`, McColl et al., 1998, Rosenberg and Donald, 1995). Systematic reviews have been used to develop guidelines and to answer important questions for evidence based practice (Rosenberg and Donald, 1995, Silagy et al., 2001) and critical appraisal to assess the quality of studies included is a crucial part of this process (Higgins J, 2008). Teaching critical appraisal has become an important part of the curriculum in medical schools and plays a central role in the interpretation and dissemination of research for evidence based practice (Bennett et al., 1987, Mogford et al., 2011, Parkes et al., 2001, Sackett, 1990).

Traditionally evidence based practice has been about utilising systematic reviews of randomised control trials (RCTs) to inform the use of interventions (Ho et al., 2008). However other types/designs of research studies are becoming increasingly important in evidence based practice, such as diagnostic testing, risk factors for disease and prevalence studies (Ho et al., 2008), hence systematic reviews in this area have become necessary. Cross sectional studies are typically used for estimating the prevalence and severity of a particular disease within a population, as well as for exploratory analytical studies of possible risk factors for mainly chronic diseases.

Various reporting guidelines are available for the creation of scientific manuscripts involving observational studies which provide guidance for authors reporting their findings. In addition, well developed appraisal tools have been developed for readers assessing the quality of cohort and case control studies, however there is currently a lack of an appraisal tool specifically aimed at cross sectional studies (CSSs). Two systematic reviews failed to identify a standalone appraisal tool specifically aimed at cross sectional studies (Katrak et al., 2004, Sanderson et al., 2007). Katrak et al. (2004) identified critical appraisal tools had been formulated specifically for individual research questions but were not transferable to other cross sectional studies. We have identified an appraisal tool, developed in Spanish, which specifically examines cross sectional studies (JME-R, 2008). However this tool only appraises the quality of reporting of cross sectional studies and does not address risk of bias

or other aspects of study quality (Chalmers et al., 1981, JME-R, 2008). As the need for the inclusion of cross sectional studies in evidence synthesis grows, the importance of understanding the quality of reporting and assessment of bias of cross sectional studies becomes increasingly important. Therefore a robust critical appraisal tool to address the quality of study design and reporting to enable the risk of bias to be identified is needed. Delphi methods and use of expert groups are increasingly being implemented to develop tools for reporting guidelines and appraisal tools (Moher et al., 1999, Schulz et al., 2010).

The aim of this study was to develop a critical appraisal tool that was simple to use, that addressed study design quality (design and reporting) and risk of bias in cross sectional studies. A secondary aim was to produce a document to aid the use of the critical appraisal tool where appropriate.

METHODS

Development of the initial draft CA tool:

The authors completed a systematic search of the literature for critical appraisal tools of cross sectional studies (See supplementary Material). Areas that needed to be included in the critical appraisal tool were identified using the results from this review and key epidemiological texts. A first draft of the tool (Supplementary material Table 1) and accompanying help text was created using previously published critical appraisal tools for observational and other types of study designs and other reference documents (Cockcroft PD, 2003`, Katrak et al., 2004, JME-R, 2008, Shea et al., 2007, von Elm et al., 2007, BestBETs, 2009, Crombie, 1996, Dohoo IR, 2009, Fowkes and Fulton, 1991, Genaidy et al., 2007, Hanke et al., 2007, Health Evidence Bulletin, 2004, Rothman, 2002, The Joanna Briggs Institute, 2008, Wong et al., 2008). STROBE (reporting standards for observational studies) was also used to aid in the development of questions for the critical appraisal tool (von Elm et al., 2007). The help text was directed at general users and was developed in order to make the tool easy to use and understandable.

The first draft of the CA tool was piloted with colleagues within the Centre for Evidence-based Veterinary Medicine (CEVM), School of Veterinary Medicine and Science (SVMS), The University of Nottingham, and the Centre for Veterinary Epidemiology and Risk Analyses in

University College Dublin (UCD). The tool was piloted using different research papers of varying quality that used cross sectional study design methodology during journal clubs and research meetings and was used in the analysis of a systematic review of cross sectional studies (Downes et al., 2013). The CA tool was also sent via email to nine individuals experienced with systematic reviews in veterinary medicine and/or study design for informal feedback. Feedback from the different groups was assessed and any changes to the CA tool were made accordingly. The analysis identified components that were to be included in a second draft of the CA tool of cross sectional studies (Supplementary material Table 2) which was used in the first round of the Delphi process.

The Delphi panel:

The purpose of the Delphi panel was to reach consensus on what components should be present in the critical appraisal tool and aid the development of the help text. Participants for the Delphi panel were sought from the fields of evidence based medicine (EBM), evidence-based veterinary medicine (EVM), epidemiology, nursing and public health and were required to be involved in university education in order to qualify for selection. Personal contacts of the authors and well known academics in the EBM/EVM fields were used as the initial contacts and potential members of the panel. Email was used to contact potential participants for enrolment in the Delphi study. These potential participants were also asked to provide additional recommendations for other potential participants. All potential participants were contacted a second time if no response was received from the first email; if no response was received after the second email the potential participant was not included any further in the study.

The Delphi process:

Prior to conducting the Delphi process it was agreed that consensus for inclusion of each component in the tool would be set at 80%. This meant that the Delphi process would continue until at least 80% of the panel agreed a component should be included in the final tool. In each round if a component had 80% consensus it stayed in the tool. If consensus was lower than 80% but greater than 50%, the component was considered for modification or

was integrated into other components that were deemed to require reassessment for the next round of the Delphi. If consensus was ≤ 50%, components were removed from the tool.

The second draft (developed in phase one described above) of the critical appraisal tool (Supplementary material Table 2) was circulated in the first round of the Delphi process to the panel with an online questionnaire (SurveyGizmo©) asking participants: if each component of the tool should be included or not; if any component required alteration or clarification; or if a further component should be added. Participants were asked to add any additional comments they had regarding each component. A hyperlink to the online questionnaire with the tool was distributed to the panel using email. Participants were given four weeks to complete their assessment of the tool using the questionnaire. Participants were reminded about the work required one week, and again three days before the Delphi round was due to close. If participants failed to respond to a specific round they were still included in the following rounds of the Delphi process. The process was repeated, with a new draft of the critical appraisal tool circulated each time based on the findings and consensus of the previous round, until 80% consensus on all components of the tool was achieved.

On the third round of the Delphi process, a draft of the help text for the tool was also included in the questionnaire and consensus was sought as to whether the tool was suitable for the non-expert user, and participants were asked to comment on the text. The responses were compiled and analysed at the end of round 3. Consensus was sought for the suitability of the help text for the non-expert user and set at 80%. However if consensus was lower than 80% but greater than 50% the help text was considered for modification. If comments were given on the help text, these comments were integrated into the help text of the tool.

Ethical approval

The ethics Committee at the School of Veterinary Medicine and Science, The University of Nottingham reviewed and approved this study. Written consent (via email) from the panellists was received by the authors following invitation to be included in the study.

RESULTS

The initial review of existing tools and texts identified 39 components that were deemed relevant for critical appraisal of cross sectional studies and were included in the 1st draft of the tool (Supplementary material Table 1). Post-feedback modification after the pilot study identified 37 components to be included in 2nd draft of the CA tool (Supplementary material Table 2).

Twenty seven potential participants were contacted for the Delphi study. Eighteen experts (67%) agreed to participate in the Delphi panel. The most common reasons for not partaking was not enough time (n=5); of these, four were lecturers with research and clinical duties and one was a lecturer with research duties. Two contacts felt they were not suitably qualified for the Delphi panel (n=2); one was retired and the other was a lecturer with research and clinical duties. Two contacts did not respond to the emails; these were both lecturers with research duties. Of those that took part eight were involved in clinical, teaching and research duties and ten were involved in research and teaching, five of the participants were veterinary surgeons and six were medical doctors.

During round one of the Delphi process, 20 components reached consensus, 13 components were assessed to require modification and it was deemed appropriate to remove four components from the tool. General comments mostly related to the tool having too many components.

> *" the tool needs to be succinct and easy and quick to use if possible - too many questions could have an impact."*
>
> *"List is too long at present and contains too many things that are general to all scientific studies"*
>
> Comments voiced included the discussion as part of the critical appraisal process being unnecessary and potentially misleading:
>
> *"the interpretation should, in my opinion, come from the methods and the results and not from what the author thinks it means"*
>
> *"I don't believe a Discussion section should be part of a critical appraisal."*

Therefore round one the tool was modified in an attempt to reduce its size and to encompass all comments. For round two 11 components remained the same and did not require testing for consensus as this was established in round one; nine components that had previously reached consensus were incorporated with the 13 components that required modification to create 10 new components (Supplementary material Table 3).

In round two consensus was reached on a further two components, six components were assessed to require modification and it was deemed appropriate to remove two components from the tool. Comments from the panel regarding the components of the tool that related to the discussion, suggested further reduction in these components due to their limited use as part of the critical appraisal process.

> *"the discussion could legitimately be highly speculative and not justified by the results provided that the authors don't present this as conclusions."*

With the reduction in the number of questions and modification of the wording, comments in round 2 reflected positive nature to the usability of the tool.

> *"I like the fact that it is quite simple - not too overloaded with methodological questions."*

After round 2, the tool was further reduced in size and modified to create a 4th draft of the tool with 20 components incorporating 13 components with full consensus and seven modified components for circulation in round 3 of the Delphi process.

Following round 3 of the Delphi process there was consensus (81%) that all components of the tool were appropriate for use by non-expert users so no further rounds were necessary. The final critical appraisal tool for cross sectional studies (AXIS tool) consisting of 20 components is shown in Table 1. The comments from the panel regarding the help text were addressed and minor modifications to the text were made (supplementary material 4).

Table 1. The final Critical Appraisal tool for Cross Sectional studies following consensus on all components by the Delphi panel.

		Yes	No	Don't know/ Comment
Introduction				

1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			
13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results for the analyses described in the methods, presented?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

DISCUSSION

The AXIS Tool

A critical appraisal tool to assess the quality and risk of bias in cross-sectional studies (AXIS), along with supporting help text, was successfully developed by an expert panel using Delphi methodology. This is the first critical appraisal tool made available for assessing this type of evidence that can be incorporated in systematic reviews, guidelines and clinical decision making.

Key Delphi Findings

One of the key items raised in comments from the experts was assessing quality of design versus quality of reporting. It is important to note that a well reported study may be of poor quality and conversely a poorly reported study could be a well conducted study(Glasziou et al., 2004, Juni et al., 2001). It is also apparent that if a study is poorly reported it can be difficult to assess the quality of the study. Some information may be lacking due to poor reporting in studies making it difficult to assess the risk of biases and the quality of the study design. High quality and complete reporting, of studies is a pre-requisite for judging quality (Moher et al., 1999, von Elm et al., 2007, Stroup et al., 2000). For this reason the AXIS tool incorporates some quality of reporting as well as quality of design and risk of biases to overcome these problems.

Using the tool

The tool was also reduced in size on each round of the Delphi process as commentators raised concerns around developing a tool with too many questions. The comments suggested that a long questionnaire would lead to the tool being cumbersome and difficult to use, and for this reason efforts were made to develop a much more concise tool.

The AXIS tool focuses mainly on the presented methods and results. It was the view of the Delphi group that the assessment as to whether the published findings of a study are credible and reliable should relate to the aims, methods and analysis of what is reported and not on the interpretation (e.g. discussion and conclusion) of the study. This view is also seen in other appraisal tools, is shared by other researchers and can be seen by the absence of questions relating to the discussion sections in critical appraisal tools for other types of studies (Katrak et al., 2004, Shea et al., 2007, Wong et al., 2008, Chalmers et al., 1981, Wells, 2000).

A comprehensive explanatory text is often used in appraisal tools for different types of study designs as it aids the reviewer when interpreting and analysing the outputs from the appraisal tool (Katrak et al., 2004, Moher et al., 1999, Schulz et al., 2010, Shea et al., 2007, von Elm et al., 2007). This approach was also used in the development of AXIS where a reviewer can link to explanatory text for each question in the tool to aid in answering and interpreting the answers to the tool questions.

Study strengths and limitations

The tool was developed through a rigorous process incorporating evidence based methods, comprehensive review, testing and consultation. Using a similar process to other appraisal tools (Crowe and Sheppard, 2011), we reviewed the relevant literature to develop a concise background on critical appraisal of cross sectional studies and to ensure no other relevant tools existed. While numerous tools exist for critical appraisal, we found a lack of tools for general use in cross sectional studies and this was consistent with what others have found previously (Katrak et al., 2004, Sanderson et al., 2007). In order to ensure quality and completeness of the tool we utilized recognised reporting guidelines, other appraisal tools and epidemiology design text in the development of the initial tool which is similar process for appraisal tools of other types of studies (Katrak et al., 2004).

The use of a multidisciplinary panel with experience in epidemiology and evidence based medicine limits the effect of using a non-representative sample, and the use of the Delphi tool is well recognised for developing consensus in health care science (Fink et al., 1984). The selection of the Delphi group is very important as it effects the results of the process (BAS, 2007). As cross-sectional studies are used extensively in both human and veterinary research it was appropriate to use expertise from both of these fields. To ensure that the tool was developed to a high standard, a high level for consensus was required in order for the questions to be retained (BAS, 2007, Greenland, 2001). There was a high level of consensus between both veterinary and medical groups in this study which adds to the rigour of the tool but also demonstrates how both healthcare areas can co-operate effectively to produce excellent outcomes.

The Delphi study was conducted using a convenience sample of experts and as such is not representative of all possible users of the tool. However the purpose of a Delphi study is to purposely hand pick participants that have prior expertise in the area of interest (Mckenna,

1994). The Delphi members came from a multidisciplinary network of professionals from medicine, nursing and veterinary medicine with experience in epidemiology and EVM and exposure to teaching and areas of EBM that were not just focused on systematic reviews of randomised control trials. The panel was restricted to those that were literate in the English language and may therefore not be representative of all nationalities.

As the tool does not provide a scale for assessing the quality of the study, a degree of subjective assessment is required. This has implications for interpretation after using the tool as there will be differences in individuals' judgements. However it has been debated that quality scales can be problematic as the outputs from assessment checklists are not linear and as such are difficult to sum up or weight making them unpredictable at assessing study quality (Greenland, 2001, Higgins et al., 2011, Juni et al., 1999). The AXIS tool has the benefit of providing the user the opportunity to assess each individual aspect of study design to give an overall assessment of the quality of the study. Further studies would be needed to assess how practical this tool is when used by clinicians and if the critical appraisal of studies using AXIS is repeatable.

CONCLUSION

In conclusion a unique tool (AXIS) for the critical appraisal cross sectional studies was developed that can be used across disciplines. The components of the AXIS tool are based on a combination of evidence, epidemiological processes, experience of the researchers and Delphi participants.

As with other evidence based initiatives the AXIS tool is intended to be an organic item that can change and improve where required, as such the validity of the tool will be measured and continuously assessed. We would also invite any future user of the tool to provide feedback so that the tool can be developed if needed and can incorporate user experience to provide better usability.

ACKNOWLEDGMENTS:

The authors thank the following individuals who participated in the Delphi process: Peter Tugwell, Thomas McGinn, Kim Thomas, Mark Petticrew, Fiona Bath-Huxtell, Amanda Burls,

Sharon Mickan, Kevin Mackway Jones, Aiden Foster, Ian Lean, Simon More, Annette O'Connor, Jan Sargeant, Hannah Jones, Ahmed Elkhadem, Julian Higgins and Sinead Langan.

The authors would like thank those who piloted the tool in the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. The authors would also like thank Michelle Downes for designing the population diagram.

The Centre for Evidence-based Veterinary Medicine is supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Contributorship statement

MJD, MLB and RSD made substantial contribution to the conception and design of the work, as well as the acquisition, analysis, or interpretation of data. He was directly involved in the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

HCW made substantial contribution to the conception and design of the work, interpretation of data and the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

Competing interests

The authors have no competing interests.

Funding

The Centre for Evidence-based Veterinary Medicine and this project was supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Data sharing statement

The authors are willing to share any of the data collected during the Delphi process that is not in this published manuscript or the supplementary material.

REFERENCES

BAS, C.-C. H. 2007. The Delphi Technique: Making Sense Of Consensus. *Practical Assessment, Research & Evaluation,* 12**,** 1-8.

BENNETT, K. J., SACKETT, D. L., HAYNES, R. B., NEUFELD, V. R., TUGWELL, P. & ROBERTS, R. 1987. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA,* 257**,** 2451-4.

BESTBETS. 2009. *BETs CA Worksheets* [Online]. Available: http://bestbets.org/links/BET-CA-worksheets.php. [Accessed 01 April 2015.

CHALMERS, T. C., SMITH, H., JR., BLACKBURN, B., SILVERMAN, B., SCHROEDER, B., REITMAN, D. & AMBROZ, A. 1981. A method for assessing the quality of a randomized control trial. *Control Clin Trials,* 2**,** 31-49.

COCKCROFT PD, H. M. 2003`. *Handbook of Evidence-based Veterinary Medicine. ,* Oxford, UK, Blackwell Publishing Ltd.

CROMBIE, I. 1996. *The Pocket Guide to Critical Appraisal. annotated edition. Wiley-Blackwell.,* London, UK, BMJ Publishing group.

CROWE, M. & SHEPPARD, L. 2011. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *Journal of Clinical Epidemiology,* 64**,** 79-89.

DOHOO IR, M. S., STRYHN H 2009. *Veterinary Epidemiologic Research.,* Charlottetown, Canada, Ver Inc.

DOWNES, M. J., DEAN, R. S., STAVISKY, J. H., ADAMS, V. J., GRINDLAY, D. J. & BRENNAN, M. L. 2013. Methods used to estimate the size of the owned cat and dog population: a systematic review. *BMC Vet Res,* 9**,** 121.

FINK, A., KOSECOFF, J., CHASSIN, M. & BROOK, R. H. 1984. Consensus Methods - Characteristics and Guidelines for Use. *American Journal of Public Health,* 74**,** 979-983.

FOWKES, F. G. R. & FULTON, P. M. 1991. Critical-Appraisal of Published Research - Introductory Guidelines. *British Medical Journal,* 302**,** 1136-1140.

GENAIDY, A. M., LEMASTERS, G. K., LOCKEY, J., SUCCOP, P., DEDDENS, J., SOBEIH, T. & DUNNING, K. 2007. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. *Ergonomics,* 50**,** 920-960.

GLASZIOU, P., VANDENBROUCKE, J. P. & CHALMERS, I. 2004. Assessing the quality of research. *BMJ,* 328**,** 39-41.

GREENLAND, S., O'ROURKE K 2001. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostat* 2**,** 463-471.

HANKE, B. A., MOTSCHALL, E. & TURP, J. C. 2007. Association between orthopedic and dental findings: What level of evidence is available? *Journal of Orofacial Orthopedics-Fortschritte Der Kieferorthopadie,* 68**,** 91-107.

HEALTH EVIDENCE BULLETIN 2004. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional.

HIGGINS J, G. S. 2008. *Cochrane Handbook for Systematic Reviews of Interventions* The Cochrane Collaboration adn John Wiley and Sons Ltd.

HIGGINS, J. P. T., ALTMAN, D. G., GOTZSCHE, P. C., JUNI, P., MOHER, D., OXMAN, A. D., SAVOVIC, J., SCHULZ, K. F., WEEKS, L., STERNE, J. A. C., GRP, C. B. M. & GRP, C. S. M. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal,* 343.

HO, P. M., PETERSON, P. N. & MASOUDI, F. A. 2008. Evaluating the evidence: is there a rigid hierarchy? *Circulation,* 118**,** 1675-84.

JME-R, S. B. 2008. [A tool for the critical appraisal of epidemiological cross-sectional studies.]. *Gaceta sanitaria / SESPAS,* 22**,** 492-497.

JUNI, P., ALTMAN, D. G. & EGGER, M. 2001. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ,* 323**,** 42-6.

JUNI, P., WITSCHI, A., BLOCH, R. & EGGER, M. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. *Jama-Journal of the American Medical Association,* 282**,** 1054-1060.

KATRAK, P., BIALOCERKOWSKI, A. E., MASSY-WESTROPP, N., KUMAR, S. & GRIMMER, K. A. 2004. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol,* 4**,** 22.

MCCOLL, A., SMITH, H., WHITE, P. & FIELD, J. 1998. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ,* 316**,** 361-5.

MCKENNA, H. P. 1994. The Delphi Technique - a Worthwhile Research Approach for Nursing. *Journal of Advanced Nursing,* 19**,** 1221-1225.

MOGFORD, E., GOULD, L. & DEVOGHT, A. 2011. Teaching critical health literacy in the US as a means to action on the social determinants of health. *Health Promotion International,* 26**,** 4-13.

MOHER, D., COOK, D. J., EASTWOOD, S., OLKIN, I., RENNIE, D., STROUP, D. F. & GRP, Q. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet,* 354**,** 1896-1900.

PARKES, J., HYDE, C., DEEKS, J. & MILNE, R. 2001. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev***,** CD001270.

ROSENBERG, W. & DONALD, A. 1995. Evidence based medicine: an approach to clinical problem-solving. *BMJ,* 310**,** 1122-6.

ROTHMAN, K. 2002. *Chapter 5: Bias in Study Design. Epidemiology: An Introduction,* New York, Oxford Univeristy Press.

SACKETT, D. L. 1990. Teaching critical appraisal. *J Gen Intern Med,* 5**,** 272.

SANDERSON, S., TATT, I. D. & HIGGINS, J. P. 2007. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol,* 36**,** 666-76.

SCHULZ, K. F., ALTMAN, D. G., MOHER, D. & GRP, C. 2010. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology,* 63**,** 834-840.

SHEA, B. J., GRIMSHAW, J. M., WELLS, G. A., BOERS, M., ANDERSSON, N., HAMEL, C., PORTER, A. C., TUGWELL, P., MOHER, D. & BOUTER, L. M. 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *Bmc Medical Research Methodology,* 7.

SILAGY, C. A., STEAD, L. F. & LANCASTER, T. 2001. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. *BMJ,* 323**,** 833-6.

STROUP, D. F., BERLIN, J. A., MORTON, S. C., OLKIN, I., WILLIAMSON, G. D., RENNIE, D., MOHER, D., BECKER, B. J., SIPE, T. A. & THACKER, S. B. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA,* 283**,** 2008-12.

THE JOANNA BRIGGS INSTITUTE. 2008. *oanna Briggs Institute Reviewers' Manual 2008 edition* [Online]. Available: http://www.joannabriggs.edu.au/documents/jbireviewmanual_cip11449.pdf.

VON ELM, E., ALTMAN, D. G., EGGER, M., POCOCK, S. J., GOTZSCHE, P. C. & VANDENBROUCKE, J. P. 2007. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bulletin of the World Health Organization,* 85**,** 867-872.

WELLS, G., SHEA B, O'CONNELL D, PETERSON J, WELCH V, ET AL. 2000. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* [Online]. Available: http://www.medicine.mcgill.ca/rtamblyn/Readings/The%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf [Accessed 31 October 2013.

WONG, W. C., CHEUNG, C. S. & HART, G. J. 2008. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. *Emerg Themes Epidemiol,* 5**,** 23.

Supplementary material

Table 1. The 1st draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post review of the literature.

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Are the aims of the study clearly stated?			
Methods				
2	Is the type of study design appropriate for the stated aim?			
3	Is the sample size justified (based on pre-study considerations of statistical power)?			
4	Is the target or reference population clearly defined? (is it clear who the research was about)			
5	Is the sample frame taken from an appropriate population base so that it closely represents the overall population under investigation?			
6	In the selection process:			
	a. Were any inclusion/exclusion criteria used?			
	b. Was random selection used to obtain participants?			
7	Is the selection process likely to select subjects that were representative of the study population of interest?			
8	If appropriate, were measures undertaken to address and categorise non-responders?			
9	Do the variables measured, in the study, produce data that reflect the aims of the study? (Validity)			
	a. Are the outcomes of interest clearly measured?			
	b. Are the risk factors appropriately measured to be compared to the outcomes of interest?			
10	If appropriate, have the measurement instruments been trialled, piloted or published previously? (Reliability and reproducibility)			
11	Are the statistical methods clearly stated?			
12	If appropriate, is the means by which statistical significance is inferred stated?			
13	Are the methods sufficiently described to enable them to be repeated?			
Results				
14	Are the basic data adequately described?			
15	Is the response rate given, if appropriate?			
16	Is information about non-responders described, if appropriate?			
17	Are the results internally consistent?			
	a. Do the numbers add up?			
	b. Are any missing data acknowledged, or described?			
18	Are the results described objectively without author opinion?			
19	Are results pertaining to the study aim reported?			
20	If appropriate is the statistical significance level declared in the methods			

	adhered to?			
21	Are the results of all tests described in the methods presented?			
Discussion				
22	Are all results pertaining to the study aim discussed?			
23	Are the limitations of the study discussed?			
24	Is selection bias addressed?			
25	Is non-response addressed?			
26	Do the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?			
27	If appropriate are non-significant results discussed?			
	a. Do the authors consider issues around study design when interpreting non-significant results?			
	b. Do the authors consider issues around sample size when interpreting non-significant results?			
Conclusions				
28	Are the authors' conclusions justified by the results?			
Other				
29	Are any conflicts of interest/funding declared in the text?			
30	Was ethical aspect approval or consent of participants attained?			

Table 2. The 2nd draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post piloting with the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. This draft was used in the first round of the Delphi panel and the results of the consensus from the panel or each component are presented.

	Consensus
Introduction	
1. Is it clear what the aims of the study were?	94.12
Methods	
2. Was the type of study design appropriate for the stated aim?	94.12
3. Was the sample size justified (based on pre-study considerations of statistical power)?	76.47
4. Was the target or reference population clearly defined? (is it clear who the research was about?)	100.00
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	88.24
6.a. In the selection process: Were any inclusion/exclusion criteria used?	82.35
6.b. In the selection process: Was random selection used to obtain participants?	70.59
7. Was the selection process likely to select subjects that were representative of the study population of interest?	93.33
8. If appropriate, were measures undertaken to address and categorise non-responders?	87.50
9. Did the variables measured in the study, produce data that reflected the aims of the study? (Validity)	68.75
9.a. Were the outcomes of interest clearly measured?	86.67
9.b. Were the risk factors measured appropriate to the outcomes of interest?	66.67
10. If appropriate, had the measurement instruments been trialled, piloted or published previously?	93.33
11. Is it clear what statistical methods were used?	86.67
12. If appropriate is it possible to determine the means by which the statistical significance was inferred? (p-values, confidence intervals)	64.29
13. Were the methods sufficiently described to enable them to be repeated?	87.50
Results	
14. Were the basic data adequately described?	75.00
15. If appropriate, was the response rate sufficient to base conclusions on?	93.75
16. If appropriate, was information about non-responders described?	93.75
17. Were the results internally consistent?	57.14
17.a. Did the numbers add up?	56.25
17.b. Were any missing data acknowledged, or described, if appropriate?	93.75
18. Were the results described objectively without author opinion?	43.75
19. Were the results pertaining to the study aim reported?	81.25

20. If appropriate, was the statistical significance level declared in the methods adhered to?	68.75
21. Were the results of all tests described in the methods presented?	93.75
22. Were all results pertaining to the study aim discussed?	56.25
Discussion	
23. Were the limitations of the study discussed?	81.25
24. Was selection bias discussed appropriately?	62.50
25. Was non-response discussed appropriately?	68.75
26. Did the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?	53.33
27. If appropriate, were non-significant results discussed?	50.00
27.a. Did the authors consider issues around study design when interpreting non-significant results?	50.00
27.b. Did the authors consider issues around sample size when interpreting non-significant results?	50.00
28. Were the authors' conclusions justified by the results?	93.75
Other	
29. Were any conflicts of interest/funding declared in the text?	93.75
30. Was ethical approval or consent of participants attained?	81.25

Table 3. The 3rd draft of a CA tool created following round 1 of the Delphi study after comments and consensus was taken into account. Results on consensus for each question from the round 2 of a Delphi panel are presented.

	Consensus*
Introduction	
1. Is it clear what the aims of the study were?	
Methods	
2. Was the type of study design appropriate for the stated aim?	
3. If appropriate, was the sample size justified?	68.75
4. Was the target or reference population clearly defined? (Is it clear who the research was about?)	
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	
6. Was the selection process likely to select subjects that were representative of the study population of interest?	
7. Were measures undertaken to address and categorise non-responders?	81.25
8. Would the variables measured in the study produce data that reflected the aims of the study? (Validity)	62.5
9. Is it clear what statistical methods were used?	50
10. Is it clear how statistical significance was determined? (eg: p-values, confidence intervals)	62.5
11. Were the methods sufficiently described to enable them to be repeated?	
Results	
12. Were the basic data adequately described?	
13. If appropriate, was the response rate sufficient to enable sound conclusions to be drawn?	56.25
14. If appropriate, was information about non-responders described?	
15. Were the results internally consistent?	
16. Were all the results of the analyses described in the methods presented?	60
17. If a statistical significance level was declared in the methods, was it adhered to in the results?	31.25
Discussion	
18. Were the authors' discussions and conclusions justified by the results?	87.5
19. Were the limitations of the study discussed?	
Other	
20. Were there any funding sources or conflicts of interest that were likely to affect the authors' interpretation of the results?	75
21. Was ethical approval or consent of participants attained?	

**Where no consensus figure is given, consensus was reached on this question in the previous round.*

BMJ Open

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011458.R1
Article Type:	Research
Date Submitted by the Author:	12-May-2016
Complete List of Authors:	Downes, Martin; Griffith University School of Medicine, Centre for Applied Health Economics Brennan, Marnie; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science Williams, HC; University of Nottingham, Centre of Evidence Based Dermatology Dean, Rachel; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Medical education and training, Patient-centred medicine, Research methods
Keywords:	Cross sectional studies, Critical appraisal, Evidence-based Healthcare, Delphi

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Martin J. Downes[1¶] Marnie L Brennan[2¶], Hywel C Williams[3], Rachel S Dean[2*¶]

[1]Centre for Applied Health Economics, School of Medicine, Griffith University, Meadowbrook, Australia

[2]Centre for Evidence-based Veterinary Medicine, The University of Nottingham, Sutton Bonington Campus, United Kingdom.

[3]Centre of Evidence Based Dermatology, The University of Nottingham, Kings Meadow Campus, Lenton Lane, Nottingham, United Kingdom.

* Email: Rachel.Dean@nottingham.ac.uk; (RSD)

[¶]These authors contributed equally to this work

ABSTRACT

Objectives: The aim of this study was to develop a critical appraisal tool that addressed study design and reporting quality as well as the risk of bias in cross sectional studies. In addition the aim was to produce a help document to guide the non-expert user through the tool.

Design: An initial scoping review of the published literature and key epidemiological texts was undertaken prior to the formation of a Delphi panel to establish key components for a critical appraisal tool for cross sectional studies. A consensus of 80% was required from the Delphi panel for any component to be included in the final tool.

Results: An initial list of 39 components was identified through examination of existing resources. An international Delphi panel of 18 medical and veterinary experts was established. After 3 rounds of the Delphi process the Appraisal tool for Cross-Sectional Studies (AXIS tool) was developed by consensus and consisted of 20 components. A detailed explanatory document was also developed with the tool, giving expanded explanation of each question and providing simple interpretations and examples of the epidemiological concepts being examined in each question to aid non-expert users.

Conclusion: Critical appraisal of the literature is a vital step in evidence synthesis and therefore evidence-based decision making in a number of different disciplines. The AXIS tool is therefore unique and was developed in a way that it can be used across disciplines to aid the inclusion of cross sectional studies in systematic reviews, guidelines and clinical decision making.

Strengths and Limitations of this study

Strengths of this study include:

- The development of a novel critical appraisal tool that can be used across discipline
- A multimodal evidence-based approach was used to develop the tool
- Expertise was harnessed from a number of different disciplines

Limitations of this study:

- The Delphi panel was based on convenience and may not encompass all eventual users of the tool.
- A numerical scale to reflect quality was not included in the final tool, which may be perceived as a limitation.

INTRODUCTION

Critical appraisal (CA) is a skill central to undertaking evidence based practice which is concerned with integrating the best external evidence with clinical care. This is because when reading any type of evidence, being critical of all aspects of the study design, execution and reporting is vital for assessing its quality before being applied to practice [1-3]. Systematic reviews have been used to develop guidelines and to answer important questions for evidence based practice [3 4] and critical appraisal to assess the quality of studies included is a crucial part of this process [5]. Teaching critical appraisal has become an important part of the curriculum in medical schools and plays a central role in the interpretation and dissemination of research for evidence based practice [6-9].

Traditionally evidence based practice has been about utilising systematic reviews of randomised control trials (RCTs) to inform the use of interventions [10]. However other types/designs of research studies are becoming increasingly important in evidence based practice, such as diagnostic testing, risk factors for disease and prevalence studies [10], hence systematic reviews in this area have become necessary. Cross sectional studies are one of those study designs that are of increasing importance in evidence based medicine. A cross sectional study has been defined as: 'An observational study whose outcome frequency measure is prevalence. The basis of a cross sectional study design is that a sample, or census, of subjects if obtained from the target population and the presence or the absence of the outcome is ascertained at a certain point' [11] Various reporting guidelines are available for the creation of scientific manuscripts involving observational studies which provide guidance for authors reporting their findings. In addition, well developed appraisal tools have been developed for readers assessing the quality of cohort and case control studies [12 13]; however, there is currently a lack of an appraisal tool specifically aimed at cross sectional studies (CSSs). Two systematic reviews failed to identify a standalone appraisal tool specifically aimed at cross sectional studies [12 13]. Katrak et al. (2004) identified critical appraisal tools had been formulated specifically for individual research questions but were not transferable to other cross sectional studies. We have identified an appraisal tool, developed in Spanish, which specifically examines cross sectional studies [14]. Berra et al (2008) essentially converted each reporting item identified in the STROBE reporting guidelines and turned them into questions for their appraisal tool asking if this was reported

in the study. Therefore, this tool only appraises the quality of reporting of cross sectional studies and does not address risk of bias or other aspects of study quality [14 15]. As the need for the inclusion of cross sectional studies in evidence synthesis grows, the importance of understanding the quality of reporting and assessment of bias of cross sectional studies becomes increasingly important. Therefore a robust critical appraisal tool to address the quality of study design and reporting to enable the risk of bias to be identified is needed. Delphi methods and use of expert groups are increasingly being implemented to develop tools for reporting guidelines and appraisal tools [16 17].

The aim of this study was to develop a critical appraisal tool that was simple to use, that addressed study design quality (design and reporting) and risk of bias in cross sectional studies. A secondary aim was to produce a document to aid the use of the critical appraisal tool where appropriate.

METHODS

Development of the initial draft CA tool:

The authors completed a systematic search of the literature for critical appraisal tools of cross sectional studies (See supplementary Material). Areas that needed to be included in the critical appraisal tool were identified using the results from this review and key epidemiological texts. A first draft of the tool (Supplementary material Table 1) and accompanying help text was created using previously published critical appraisal tools for observational and other types of study designs and other reference documents [1 11 12 14 18-28]. STROBE (reporting standards for observational studies) was also used to aid in the development of questions for the critical appraisal tool [19]. The help text was directed at general users and was developed in order to make the tool easy to use and understandable.

The first draft of the CA tool was piloted with colleagues within the Centre for Evidence-based Veterinary Medicine (CEVM), School of Veterinary Medicine and Science (SVMS), The University of Nottingham, and the Centre for Veterinary Epidemiology and Risk Analyses in University College Dublin (UCD). The tool was piloted using different research papers of varying quality that used cross sectional study design methodology during journal clubs and research meetings and was used in the analysis of a systematic review of cross sectional

studies [29]. The CA tool was also sent via email to nine individuals experienced with systematic reviews in veterinary medicine and/or study design for informal feedback. Feedback from the different groups was assessed and any changes to the CA tool were made accordingly. The analysis identified components that were to be included in a second draft of the CA tool of cross sectional studies (Supplementary material Table 2) which was used in the first round of the Delphi process.

The Delphi panel:

The purpose of the Delphi panel was to reach consensus on what components should be present in the critical appraisal tool and aid the development of the help text. Participants for the Delphi panel were sought from the fields of evidence based medicine (EBM), evidence-based veterinary medicine (EVM), epidemiology, nursing and public health and were required to be involved in university education in order to qualify for selection. Personal contacts of the authors and well known academics in the EBM/EVM fields were used as the initial contacts and potential members of the panel. Email was used to contact potential participants for enrolment in the Delphi study. These potential participants were also asked to provide additional recommendations for other potential participants. All potential participants were contacted a second time if no response was received from the first email; if no response was received after the second email the potential participant was not included any further in the study.

The Delphi process:

Prior to conducting the Delphi process it was agreed that consensus for inclusion of each component in the tool would be set at 80%[30 31]. This meant that the Delphi process would continue until at least 80% of the panel agreed a component should be included in the final tool. In each round if a component had 80% consensus it stayed in the tool. If consensus was lower than 80% but greater than 50%, the component was considered for modification or was integrated into other components that were deemed to require reassessment for the next round of the Delphi. If consensus was ≤ 50%, components were removed from the tool.

The second draft (developed in phase one described above) of the critical appraisal tool (Supplementary material Table 2) was circulated in the first round of the Delphi process to the panel with an online questionnaire (SurveyGizmo©) asking participants: if each component of the tool should be included or not; if any component required alteration or clarification; or if a further component should be added. Participants were asked to add any additional comments they had regarding each component. A hyperlink to the online questionnaire with the tool was distributed to the panel using email. Participants were given four weeks to complete their assessment of the tool using the questionnaire. Participants were reminded about the work required one week, and again three days before the Delphi round was due to close. If participants failed to respond to a specific round they were still included in the following rounds of the Delphi process. The process was repeated, with a new draft of the critical appraisal tool circulated each time based on the findings and consensus of the previous round, until 80% consensus on all components of the tool was achieved.

On the third round of the Delphi process, a draft of the help text for the tool was also included in the questionnaire and consensus was sought as to whether the tool was suitable for the non-expert user, and participants were asked to comment on the text. The responses were compiled and analysed at the end of round 3. Consensus was sought for the suitability of the help text for the non-expert user and set at 80%. However if consensus was lower than 80% but greater than 50% the help text was considered for modification. If comments were given on the help text, these comments were integrated into the help text of the tool.

Ethical approval

The ethics Committee at the School of Veterinary Medicine and Science, The University of Nottingham reviewed and approved this study. Written consent (via email) from the panellists was received by the authors following invitation to be included in the study.

RESULTS

The initial review of existing tools and texts identified 39 components that were deemed relevant for critical appraisal of cross sectional studies and were included in the 1st draft of the tool (Supplementary material Table 1). Post-feedback modification after the pilot study identified 37 components to be included in 2nd draft of the CA tool (Supplementary material Table 2).

Twenty seven potential participants were contacted for the Delphi study. Eighteen experts (67%) agreed to participate in the Delphi panel. The most common reasons for not partaking was not enough time (n=5); of these, four were lecturers with research and clinical duties and one was a lecturer with research duties. Two contacts felt they were not suitably qualified for the Delphi panel (n=2); one was retired and the other was a lecturer with research and clinical duties. Two contacts did not respond to the emails; these were both lecturers with research duties. Of those that took part eight were involved in clinical, teaching and research duties and ten were involved in research and teaching, five of the participants were veterinary surgeons and six were medical clinicians. It was an international panel including ten participants from the United Kingdom, three from Australia, two from the United states of America, two from Canada, and one from Egypt.

During round one (undertaken in February 2013) of the Delphi process, 20 components reached consensus, 13 components were assessed to require modification and it was deemed appropriate to remove four components from the tool. General comments mostly related to the tool having too many components.

> *" the tool needs to be succinct and easy and quick to use if possible - too many questions could have an impact."*
>
> *"List is too long at present and contains too many things that are general to all scientific studies"*
>
> Comments voiced included the discussion as part of the critical appraisal process being unnecessary and potentially misleading:
>
> *"the interpretation should, in my opinion, come from the methods and the results and not from what the author thinks it means"*
>
> *"I don't believe a Discussion section should be part of a critical appraisal."*

Therefore, in round one the tool was modified in an attempt to reduce its size and to encompass all comments. For round two (undertaken in May 2013) 11 components remained the same and did not require testing for consensus as this was established in round one; nine components that had previously reached consensus were incorporated with the 13 components that required modification to create 10 new components (Supplementary material Table 3).

In round two consensus was reached on a further two components, six components were assessed to require modification and it was deemed appropriate to remove two components from the tool. Comments from the panel regarding the components of the tool that related to the discussion, suggested further reduction in these components due to their limited use as part of the critical appraisal process.

> *"the discussion could legitimately be highly speculative and not justified by the results provided that the authors don't present this as conclusions."*

With the reduction in the number of questions and modification of the wording, comments in round 2 reflected positive nature to the usability of the tool.

> *"I like the fact that it is quite simple - not too overloaded with methodological questions."*

After round 2, the tool was further reduced in size and modified to create a 4th draft of the tool with 20 components incorporating 13 components with full consensus and seven modified components for circulation in round 3 of the Delphi process.

Following round 3 (undertaken in July 2013) of the Delphi process there was consensus (81%) that all components of the tool were appropriate for use by non-expert users so no further rounds were necessary. The final critical appraisal tool for cross sectional studies (AXIS tool) consisting of 20 components is shown in Table 1. The comments from the panel regarding the help text were addressed and minor modifications to the text were made (supplementary material 4).

Table 1. The final Critical Appraisal tool for Cross Sectional studies following consensus on all components by the Delphi panel.

		Yes	No	Don't know/ Comment

Introduction				
1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			
13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results for the analyses described in the methods, presented?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

DISCUSSION

The AXIS Tool

A critical appraisal tool to assess the quality and risk of bias in cross-sectional studies (AXIS), along with supporting help text, was successfully developed by an expert panel using Delphi methodology. This is the first critical appraisal tool made available for assessing this type of evidence that can be incorporated in systematic reviews, guidelines and clinical decision making.

Key Delphi Findings

One of the key items raised in comments from the experts was assessing quality of design versus quality of reporting. It is important to note that a well reported study may be of poor quality and conversely a poorly reported study could be a well conducted study[32 33]. It is also apparent that if a study is poorly reported it can be difficult to assess the quality of the study. Some information may be lacking due to poor reporting in studies making it difficult to assess the risk of biases and the quality of the study design. High quality and complete reporting, of studies is a pre-requisite for judging quality [16 19 34]. For this reason the AXIS tool incorporates some quality of reporting as well as quality of design and risk of biases to overcome these problems.

Using the tool

The tool was also reduced in size on each round of the Delphi process as commentators raised concerns around developing a tool with too many questions. The comments suggested that a long questionnaire would lead to the tool being cumbersome and difficult to use, and for this reason efforts were made to develop a much more concise tool.

The AXIS tool focuses mainly on the presented methods and results. It was the view of the Delphi group that the assessment as to whether the published findings of a study are credible and reliable should relate to the aims, methods and analysis of what is reported and not on the interpretation (e.g. discussion and conclusion) of the study. This view is also seen in other appraisal tools, is shared by other researchers and can be seen by the absence of questions relating to the discussion sections in critical appraisal tools for other types of studies [12 15 18 27 35].

As with all critical appraisal tools it is only possible for the reader to be able to critique what is reported. If an important aspect of a study is not in the manuscript, it is unclear to the

reader whether it was done, and not reported, or not done at all. It is therefore the responsibility of the appraiser of the study to recognise omissions in reporting and identify omissions and consider how this affects the reliability of the results. (do I need to reference this?)

A comprehensive explanatory text is often used in appraisal tools for different types of study designs as it aids the reviewer when interpreting and analysing the outputs from the appraisal tool [12 16-19]. This approach was also used in the development of AXIS where a reviewer can link to explanatory text for each question in the tool to aid in answering and interpreting the answers to the tool questions.

Study strengths and limitations

The tool was developed through a rigorous process incorporating rigorous methods, including comprehensive review, testing and consultation through a Delphi panel. Using a similar process to other appraisal tools [36], we reviewed the relevant literature to develop a concise background on critical appraisal of cross sectional studies and to ensure no other relevant tools existed. While numerous tools exist for critical appraisal, we found a lack of tools for general use in cross sectional studies and this was consistent with what others have found previously [12 13]. In order to ensure quality and completeness of the tool we utilized recognised reporting guidelines, other appraisal tools and epidemiology design text in the development of the initial tool which is similar process for appraisal tools of other types of studies [12].

The use of a multidisciplinary panel with experience in epidemiology and evidence based medicine limits the effect of using a non-representative sample, and the use of the Delphi tool is well recognised for developing consensus in health care science [37]. The selection of the Delphi group is very important as it effects the results of the process [30]. As cross-sectional studies are used extensively in both human and veterinary research it was appropriate to use expertise from both of these fields. To ensure that the tool was developed to a high standard, a high level for consensus was required in order for the questions to be retained [30 31 38]. There was a high level of consensus between both veterinary and medical groups in this study which adds to the rigour of the tool but also demonstrates how both healthcare areas can co-operate effectively to produce excellent outcomes.

The Delphi study was conducted using a convenience sample of experts and as such is not representative of all possible users of the tool. However the purpose of a Delphi study is to purposely hand pick participants that have prior expertise in the area of interest [39]. The Delphi members came from a multidisciplinary network of professionals from medicine, nursing and veterinary medicine with experience in epidemiology and EVM and exposure to teaching and areas of EBM that were not just focused on systematic reviews of randomised control trials. The panel was restricted to those that were literate in the English language and may therefore not be representative of all nationalities. The interests and experiences of the panel will clearly have an effect on the results of this study as this is common to all Delphi studies [30 40]. The majority of Delphi studies are conducted using between 15 and 20 participants [30], so a panel of 18 is consistent with other published Delphi panels. We aimed to recruit a minimum of 15 participants and as it was anticipated that not all participants contacted would be able to take part, 27 potential participants were contacted.

As the tool does not provide a scale for assessing the quality of the study, a degree of subjective assessment is required. This has implications for interpretation after using the tool as there will be differences in individuals' judgements. However it has been debated that quality scales can be problematic as the outputs from assessment checklists are not linear and as such are difficult to sum up or weight making them unpredictable at assessing study quality [38 41 42]. The AXIS tool has the benefit of providing the user the opportunity to assess each individual aspect of study design to give an overall assessment of the quality of the study. By providing this subjectivity AXIS gives the user more flexibility in incorporating both quality of reporting and risk of bias when making judgment on the quality of a paper, whereas Berra et al (2008) [14]only allows the user to assess quality of reporting and tools such as the Cochrane risk of bias tool [5] do not address poor reporting. Further studies would be needed to assess how practical this tool is when used by clinicians and if the critical appraisal of studies using AXIS is repeatable.

CONCLUSION

In conclusion a unique tool (AXIS) for the critical appraisal cross sectional studies was developed that can be used across disciplines. The components of the AXIS tool are based on a combination of evidence, epidemiological processes, experience of the researchers and Delphi participants.

As with other evidence based initiatives the AXIS tool is intended to be an organic item that can change and improve where required, as such the validity of the tool will be measured and continuously assessed. We would also invite any future user of the tool to provide feedback so that the tool can be developed if needed and can incorporate user experience to provide better usability.

ACKNOWLEDGMENTS:

The authors thank the following individuals who participated in the Delphi process: Peter Tugwell, Thomas McGinn, Kim Thomas, Mark Petticrew, Fiona Bath-Huxtell, Amanda Burls, Sharon Mickan, Kevin Mackway Jones, Aiden Foster, Ian Lean, Simon More, Annette O'Connor, Jan Sargeant, Hannah Jones, Ahmed Elkhadem, Julian Higgins and Sinead Langan.

The authors would like thank those who piloted the tool in the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. The authors would also like thank Michelle Downes for designing the population diagram.

The Centre for Evidence-based Veterinary Medicine is supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Contributorship statement

MJD, MLB and RSD made substantial contribution to the conception and design of the work, as well as the acquisition, analysis, or interpretation of data. He was directly involved in the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

HCW made substantial contribution to the conception and design of the work, interpretation of data and the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

Competing interests

The authors have no competing interests.

Funding

The Centre for Evidence-based Veterinary Medicine and this project was supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Data sharing statement

The authors are willing to share any of the data collected during the Delphi process that is not in this published manuscript or the supplementary material.

REFERENCES

1. Cockcroft PD HM. *Handbook of Evidence-based Veterinary Medicine.* . 1st ed. Oxford, UK: Blackwell Publishing Ltd, 2003`.
2. McColl A, Smith H, White P, et al. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. Bmj 1998;**316**(7128):361-5.
3. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. Bmj 1995;**310**(6987):1122-6.
4. Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. Bmj 2001;**323**(7317):833-6.
5. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*: Wiley Online Library, 2008.
6. Bennett KJ, Sackett DL, Haynes RB, et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. Jama 1987;**257**(18):2451-4.
7. Mogford E, Gould L, Devoght A. Teaching critical health literacy in the US as a means to action on the social determinants of health. Health Promot Int 2011;**26**(1):4-13.
8. Parkes J, Hyde C, Deeks J, et al. Teaching critical appraisal skills in health care settings. The Cochrane database of systematic reviews 2001(3):CD001270.
9. Sackett DL. Teaching critical appraisal. Journal of general internal medicine 1990;**5**(3):272.
10. Ho PM, Peterson PN, Masoudi FA. Evaluating the evidence: is there a rigid hierarchy? Circulation 2008;**118**(16):1675-84.
11. Dohoo IR, Martin SW, Stryhn H. *Veterinary Epidemiologic Research*. Charlottetown, Canada: Ver Books, 2009.
12. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC medical research methodology 2004;**4**:22.
13. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International journal of epidemiology 2007;**36**(3):666-76.
14. Berra S, Elorza-Ricart JM, Estrada M-D, et al. [A tool (corrected) for the critical appraisal of epidemiological cross-sectional studies]. Gac Sanit 2008;**22**(5):492-7.
15. Chalmers TC, Smith H, Jr., Blackburn B, et al. A method for assessing the quality of a randomized control trial. Controlled clinical trials 1981;**2**(1):31-49.
16. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999;**354**(9193):1896-900.

17. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;**63**(8):834-40.
18. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology 2007;**7**.
19. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. B World Health Organ 2007;**85**(11):867-72.
20. BestBETs. BETs CA Worksheets. Secondary BETs CA Worksheets 2009. http://bestbets.org/links/BET-CA-worksheets.php.
21. Crombie I. *The Pocket Guide to Critical Appraisal. annotated edition. Wiley-Blackwell.* 1 ed. London, UK: BMJ Publishing group, 1996.
22. Fowkes FGR, Fulton PM. Critical-Appraisal of Published Research - Introductory Guidelines. Brit Med J 1991;**302**(6785):1136-40.
23. Genaidy AM, Lemasters GK, Lockey J, et al. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. Ergonomics 2007;**50**(6):920-60.
24. Hanke BA, Motschall E, Turp JC. Association between orthopedic and dental findings: What level of evidence is available? J Orofac Orthop 2007;**68**(2):91-107.
25. Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional, 2004.
26. Rothman K. *Chapter 5: Bias in Study Design. Epidemiology: An Introduction*. New York: Oxford Univeristy Press, 2002.
27. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerging themes in epidemiology 2008;**5**:23.
28. Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual 2014 Edition*, 2014.
29. Downes MJ, Dean RS, Stavisky JH, et al. Methods used to estimate the size of the owned cat and dog population: a systematic review. BMC veterinary research 2013;**9**:121.
30. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Practical assessment, research & evaluation 2007;**12**(10):1-8.
31. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 2004;**42**(1):15-29.
32. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. Bmj 2004;**328**(7430):39-41.
33. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj 2001;**323**(7303):42-6.
34. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama 2000;**283**(15):2008-12.
35. Wells G, Shea B, O'connell D, Peterson J, Welch V, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Secondary The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000. http://www.medicine.mcgill.ca/rtamblyn/Readings/The%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf.
36. Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011;**64**(1):79-89.
37. Fink A, Kosecoff J, Chassin M, et al. Consensus Methods - Characteristics and Guidelines for Use. Am J Public Health 1984;**74**(9):979-83.
38. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostat 2001;**2**:463-71.

39. Mckenna HP. The Delphi Technique - a Worthwhile Research Approach for Nursing. J Adv Nurs 1994;**19**(6):1221-25.
40. Linstone HA, Turoff M. *The Delphi method: Techniques and applications*: Addison-Wesley Reading, MA, 1975.
41. Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Brit Med J 2011;**343**.
42. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. Jama-J Am Med Assoc 1999;**282**(11):1054-60.

Supplementary material

Table 1. The 1st draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post review of the literature.

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Are the aims of the study clearly stated?			
Methods				
2	Is the type of study design appropriate for the stated aim?			
3	Is the sample size justified (based on pre-study considerations of statistical power)?			
4	Is the target or reference population clearly defined? (is it clear who the research was about)			
5	Is the sample frame taken from an appropriate population base so that it closely represents the overall population under investigation?			
6	In the selection process:			
	a. Were any inclusion/exclusion criteria used?			
	b. Was random selection used to obtain participants?			
7	Is the selection process likely to select subjects that were representative of the study population of interest?			
8	If appropriate, were measures undertaken to address and categorise non-responders?			
9	Do the variables measured, in the study, produce data that reflect the aims of the study? (Validity)			
	a. Are the outcomes of interest clearly measured?			
	b. Are the risk factors appropriately measured to be compared to the outcomes of interest?			
10	If appropriate, have the measurement instruments been trialled, piloted or published previously? (Reliability and reproducibility)			
11	Are the statistical methods clearly stated?			
12	If appropriate, is the means by which statistical significance is inferred stated?			
13	Are the methods sufficiently described to enable them to be repeated?			
Results				
14	Are the basic data adequately described?			
15	Is the response rate given, if appropriate?			
16	Is information about non-responders described, if appropriate?			
17	Are the results internally consistent?			
	a. Do the numbers add up?			
	b. Are any missing data acknowledged, or described?			
18	Are the results described objectively without author opinion?			
19	Are results pertaining to the study aim reported?			
20	If appropriate is the statistical significance level declared in the methods			

	adhered to?			
21	Are the results of all tests described in the methods presented?			
Discussion				
22	Are all results pertaining to the study aim discussed?			
23	Are the limitations of the study discussed?			
24	Is selection bias addressed?			
25	Is non-response addressed?			
26	Do the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?			
27	If appropriate are non-significant results discussed?			
	a. Do the authors consider issues around study design when interpreting non-significant results?			
	b. Do the authors consider issues around sample size when interpreting non-significant results?			
Conclusions				
28	Are the authors' conclusions justified by the results?			
Other				
29	Are any conflicts of interest/funding declared in the text?			
30	Was ethical aspect approval or consent of participants attained?			

Table 2. The 2nd draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post piloting with the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. This draft was used in the first round of the Delphi panel and the results of the consensus from the panel or each component are presented.

	Consensus
Introduction	
1. Is it clear what the aims of the study were?	94.12
Methods	
2. Was the type of study design appropriate for the stated aim?	94.12
3. Was the sample size justified (based on pre-study considerations of statistical power)?	76.47
4. Was the target or reference population clearly defined? (is it clear who the research was about?)	100.00
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	88.24
6.a. In the selection process: Were any inclusion/exclusion criteria used?	82.35
6.b. In the selection process: Was random selection used to obtain participants?	70.59
7. Was the selection process likely to select subjects that were representative of the study population of interest?	93.33
8. If appropriate, were measures undertaken to address and categorise non-responders?	87.50
9. Did the variables measured in the study, produce data that reflected the aims of the study? (Validity)	68.75
9.a. Were the outcomes of interest clearly measured?	86.67
9.b. Were the risk factors measured appropriate to the outcomes of interest?	66.67
10. If appropriate, had the measurement instruments been trialled, piloted or published previously?	93.33
11. Is it clear what statistical methods were used?	86.67
12. If appropriate is it possible to determine the means by which the statistical significance was inferred? (p-values, confidence intervals)	64.29
13. Were the methods sufficiently described to enable them to be repeated?	87.50
Results	
14. Were the basic data adequately described?	75.00
15. If appropriate, was the response rate sufficient to base conclusions on?	93.75
16. If appropriate, was information about non-responders described?	93.75
17. Were the results internally consistent?	57.14
17.a. Did the numbers add up?	56.25
17.b. Were any missing data acknowledged, or described, if appropriate?	93.75
18. Were the results described objectively without author opinion?	43.75
19. Were the results pertaining to the study aim reported?	81.25

20. If appropriate, was the statistical significance level declared in the methods adhered to?	68.75
21. Were the results of all tests described in the methods presented?	93.75
22. Were all results pertaining to the study aim discussed?	56.25
Discussion	
23. Were the limitations of the study discussed?	81.25
24. Was selection bias discussed appropriately?	62.50
25. Was non-response discussed appropriately?	68.75
26. Did the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?	53.33
27. If appropriate, were non-significant results discussed?	50.00
27.a. Did the authors consider issues around study design when interpreting non-significant results?	50.00
27.b. Did the authors consider issues around sample size when interpreting non-significant results?	50.00
28. Were the authors' conclusions justified by the results?	93.75
Other	
29. Were any conflicts of interest/funding declared in the text?	93.75
30. Was ethical approval or consent of participants attained?	81.25

Table 3. The 3rd draft of a CA tool created following round 1 of the Delphi study after comments and consensus was taken into account. Results on consensus for each question from the round 2 of a Delphi panel are presented.

	Consensus*
Introduction	
1. Is it clear what the aims of the study were?	
Methods	
2. Was the type of study design appropriate for the stated aim?	
3. If appropriate, was the sample size justified?	68.75
4. Was the target or reference population clearly defined? (Is it clear who the research was about?)	
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	
6. Was the selection process likely to select subjects that were representative of the study population of interest?	
7. Were measures undertaken to address and categorise non-responders?	81.25
8. Would the variables measured in the study produce data that reflected the aims of the study? (Validity)	62.5
9. Is it clear what statistical methods were used?	50
10. Is it clear how statistical significance was determined? (eg: p-values, confidence intervals)	62.5
11. Were the methods sufficiently described to enable them to be repeated?	
Results	
12. Were the basic data adequately described?	
13. If appropriate, was the response rate sufficient to enable sound conclusions to be drawn?	56.25
14. If appropriate, was information about non-responders described?	
15. Were the results internally consistent?	
16. Were all the results of the analyses described in the methods presented?	60
17. If a statistical significance level was declared in the methods, was it adhered to in the results?	31.25
Discussion	
18. Were the authors' discussions and conclusions justified by the results?	87.5
19. Were the limitations of the study discussed?	
Other	
20. Were there any funding sources or conflicts of interest that were likely to affect the authors' interpretation of the results?	75
21. Was ethical approval or consent of participants attained?	

**Where no consensus figure is given, consensus was reached on this question in the previous round.*

BMJ Open

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011458.R2
Article Type:	Research
Date Submitted by the Author:	05-Sep-2016
Complete List of Authors:	Downes, Martin; Griffith University School of Medicine, Centre for Applied Health Economics Brennan, Marnie; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science Williams, HC; University of Nottingham, Centre of Evidence Based Dermatology Dean, Rachel; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Medical education and training, Patient-centred medicine, Research methods
Keywords:	Cross sectional studies, Critical appraisal, Evidence-based Healthcare, Delphi

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Martin J. Downes[1¶] Marnie L Brennan[2¶], Hywel C Williams[3], Rachel S Dean[2*¶]

[1]Centre for Applied Health Economics, School of Medicine, Griffith University, Meadowbrook, Australia

[2]Centre for Evidence-based Veterinary Medicine, The University of Nottingham, Sutton Bonington Campus, United Kingdom.

[3]Centre of Evidence Based Dermatology, The University of Nottingham, Kings Meadow Campus, Lenton Lane, Nottingham, United Kingdom.

* Email: Rachel.Dean@nottingham.ac.uk; (RSD)

[¶]These authors contributed equally to this work

ABSTRACT

Objectives: The aim of this study was to develop a critical appraisal tool that addressed study design and reporting quality as well as the risk of bias in cross sectional studies. In addition, the aim was to produce a help document to guide the non-expert user through the tool.

Design: An initial scoping review of the published literature and key epidemiological texts was undertaken prior to the formation of a Delphi panel to establish key components for a critical appraisal tool for cross sectional studies. A consensus of 80% was required from the Delphi panel for any component to be included in the final tool.

Results: An initial list of 39 components was identified through examination of existing resources. An international Delphi panel of 18 medical and veterinary experts was established. After 3 rounds of the Delphi process the Appraisal tool for Cross-Sectional Studies (AXIS tool) was developed by consensus and consisted of 20 components. A detailed explanatory document was also developed with the tool, giving expanded explanation of each question and providing simple interpretations and examples of the epidemiological concepts being examined in each question to aid non-expert users.

Conclusion: Critical appraisal of the literature is a vital step in evidence synthesis and therefore evidence-based decision making in a number of different disciplines. The AXIS tool is therefore unique and was developed in a way that it can be used across disciplines to aid the inclusion of cross sectional studies in systematic reviews, guidelines and clinical decision making.

Strengths and Limitations of this study

Strengths of this study include:

- The development of a novel critical appraisal tool that can be used across disciplines
- A multimodal evidence-based approach was used to develop the tool
- Expertise was harnessed from a number of different disciplines

Limitations of this study:

- The Delphi panel was based on convenience and may not encompass all eventual users of the tool.
- A numerical scale to reflect quality was not included in the final tool, which may be perceived as a limitation.

INTRODUCTION

Critical appraisal (CA) is a skill central to undertaking evidence-based practice which is concerned with integrating the best external evidence with clinical care. This is because when reading any type of evidence, being critical of all aspects of the study design, execution and reporting is vital for assessing its quality before being applied to practice[1-3]. Systematic reviews have been used to develop guidelines and to answer important questions for evidence-based practice[3 4] and critical appraisal to assess the quality of studies that have been included is a crucial part of this process [5]. Teaching critical appraisal has become an important part of the curriculum in medical schools and plays a central role in the interpretation and dissemination of research for evidence-based practice[6-9].

Traditionally evidence-based practice has been about utilising systematic reviews of randomised control trials (RCTs) to inform the use of interventions[10]. However other types/designs of research studies are becoming increasingly important in evidence-based practice, such as diagnostic testing, risk factors for disease and prevalence studies [10], hence systematic reviews in this area have become necessary. Cross sectional studies are one of those study designs that are of increasing importance in evidence-based medicine. A cross sectional study has been defined as: 'An observational study whose outcome frequency measure is prevalence. The basis of a cross sectional study design is that a sample, or census, of subjects is obtained from the target population and the presence or the absence of the outcome is ascertained at a certain point' [11]. Various reporting guidelines are available for the creation of scientific manuscripts involving observational studies which provide guidance for authors reporting their findings.

In addition, well developed appraisal tools have been created for readers assessing the quality of cohort and case control studies[12 13]; however, there is currently a lack of an appraisal tool specifically aimed at cross sectional studies (CSSs). The Cochrane collaboration has developed a risk of bias tool for non-randomised studies (ROBINS-I)[14]; however, this is a generic tool for case control and cohort studies that does not facilitate a detailed and specific enough appraisal to be able to fully critique a cross sectional study, In addition it is only intended for use to assess risk of bias when making judgments about an intervention. Two systematic reviews failed to identify a standalone appraisal tool

specifically aimed at cross sectional studies[12 13]. Katrak et al. (2004) identified that critical appraisal tools had been formulated specifically for individual research questions but were not transferable to other cross sectional studies. We identified an appraisal tool, developed in Spanish, which specifically examined cross sectional studies[15]. Berra et al (2008) essentially converted each reporting item identified in the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) reporting guidelines and turned them into questions for their appraisal tool. Fundamentally the tool developed by Berra et al (2008) only appraises the quality of reporting of cross sectional studies and does not address risk of bias or other aspects of study quality[15 16]. Good quality of reporting of a study means that all aspects of the methods and the results are presented well and in line with international standards such as STROBE [17]; however, this is only one aspect of appraisal as a well reported study does not necessarily mean that the study is of high quality. Bias ('a systematic error, or deviation from the truth, in results or inferences'[5]) and study design are other areas that need to be considered when assessing the quality of included studies as these can be inherent even in a well reported study.

As the need for the inclusion of cross sectional studies in evidence synthesis grows, the importance of understanding the quality of reporting and assessment of bias of cross sectional studies becomes increasingly important. Therefore, a robust critical appraisal tool to address the quality of study design and reporting to enable the risk of bias to be identified is needed. Delphi methods and use of expert groups are increasingly being implemented to develop tools for reporting guidelines and appraisal tools[18 19].

The aim of this study was to develop a critical appraisal tool that was simple to use, that addressed study design quality (design and reporting) and risk of bias in cross sectional studies. A secondary aim was to produce a document to aid the use of the critical appraisal tool where appropriate.

METHODS

Development of the initial draft Critical Appraisal tool:

The authors completed a systematic search of the literature for critical appraisal tools of cross sectional studies (Supplementary material Table 1). A number of publications were identified in the review and a number of key epidemiological texts were also identified to assist in the development of the new tool[1 11 12 15 17 20-29].. MJD and MLB used these resources to subjectively identify areas that were to be included in the critical appraisal tool. These items were discussed with RSD and a first draft of the tool (Supplementary material Table 2) and accompanying help text was created using previously published critical appraisal tools for observational and other types of study designs, and other reference documents[1 11 12 15 17 20-29]. The help text was directed at general users and was developed in order to make the tool easy to use and understandable.

The first draft of the CA tool was piloted with colleagues within the Centre for Evidence-based Veterinary Medicine (CEVM) and the population health and welfare research group at the School of Veterinary Medicine and Science (SVMS), The University of Nottingham, and the Centre for Veterinary Epidemiology and Risk Analyses in University College Dublin (UCD). Colleagues used the tool to assess different research papers of varying quality that utilised cross sectional study design methodology during journal clubs and research meetings and provided feedback on their experience. The tool was used in the analysis of cross sectional studies for a published systematic review[30]. The tool was also trialled in a journal club and percentage agreement analysis was carried out and used to develop the tool further. The CA tool was also sent via email to nine individuals experienced with systematic reviews in veterinary medicine and/or study design for informal feedback. Feedback from the different groups was assessed and any changes to the CA tool were made accordingly. The analysis identified components that were to be included in a second draft of the CA tool of cross sectional studies (Supplementary material Table 3) which was used in the first round of the Delphi process.

The Delphi panel:

The purpose of the Delphi panel was to reach consensus on what components should be present in the critical appraisal tool and aid the development of the help text. Participants for the Delphi panel were sought from the fields of evidence-based medicine (EBM), evidence-based veterinary medicine (EVM), epidemiology, nursing and public health and

were required to be involved in university education in order to qualify for selection. Personal contacts of the authors and well known academics in the EBM/EVM fields were used as the initial contacts and potential members of the panel. Email was used to contact potential participants for enrolment in the Delphi study. These potential participants were also asked to provide additional recommendations for other potential participants. All potential participants were contacted a second time if no response was received from the first email; if no response was received after the second email the potential participant was not included any further in the study.

Participants were included if:

- they held a postgraduate qualification;
- they were recognised through publication and or key note presentations for their work in evidence-based medicine and veterinary medicine, epidemiology, or public health;
- taught at university level; and
- had authored in systematic reviews, reporting guidelines or critical appraisal.

The Delphi process:

Prior to conducting the Delphi process it was agreed that consensus for inclusion of each component in the tool would be set at 80%[31 32]. This meant that the Delphi process would continue until at least 80% of the panel agreed a component should be included in the final tool. Only if a component met the consensus criteria would it be included in the final tool, the steering committee did not change any component once it reached consensus or add any component that did not go through the Delphi panel. In each round, if a component had 80% consensus, it remained in the tool. If consensus was lower than 80% but greater than 50%, the component was considered for modification or was integrated into other components that were deemed to require reassessment for the next round of the Delphi. If consensus was ≤ 50%, components were removed from the tool.

The second draft (developed in phase one described above) of the critical appraisal tool (Supplementary material Table 3) was circulated in the first round of the Delphi process to the panel using an online questionnaire (SurveyGizmo©). Participants were asked: if each

component of the tool should be included or not; if any component required alteration or clarification; or if a further component should be added. Participants were asked to add any additional comments they had regarding each component. A hyperlink to the online questionnaire with the tool was distributed to the panel using email. Participants were given four weeks to complete their assessment of the tool using the questionnaire. Participants were reminded about the work required after one week, and again three days before the Delphi round was due to close. If participants failed to respond to a specific round they were still included in the following rounds of the Delphi process. The process was repeated, with a new draft of the critical appraisal tool circulated each time based on the findings and consensus of the previous round, until 80% consensus on all components of the tool was achieved.

On the third round of the Delphi process, a draft of the help text for the tool was also included in the questionnaire and consensus was sought as to whether the tool was suitable for the non-expert user, and participants were asked to comment on the text. The responses were compiled and analysed at the end of round 3. Consensus was sought for the suitability of the help text for the non-expert user and set at 80%. However if consensus was lower than 80% but greater than 50% the help text was considered for modification. If comments were given on the help text, these comments were integrated into the help text of the tool.

Ethical approval

The ethics committee at the School of Veterinary Medicine and Science, The University of Nottingham reviewed and approved this study. Written consent (via email) from the panellists was received by the authors following invitation to be included in the study.

RESULTS

The initial review of existing tools and texts identified 34 components that were deemed relevant for critical appraisal of cross sectional studies and were included in the 1st draft of the tool (Supplementary material Table 2). When piloted there was an overall percent agreement of 88.9%; however, 32.9% of the questions were unanswered. Post-feedback

modification after the pilot study identified 37 components to be included in 2nd draft of the CA tool (Supplementary material Table 3).

Twenty seven potential participants were contacted for the Delphi study. Eighteen experts (67%) agreed to participate in the Delphi panel. The most common reasons for not partaking was not enough time (n=5); of these, four were lecturers with research and clinical duties and one was a lecturer with research duties. Two contacts felt they were not suitably qualified for the Delphi panel (n=2); one was retired and the other was a lecturer with research and clinical duties. Two contacts did not respond to the emails; these were both lecturers with research duties. Of those that took part eight were involved in clinical, teaching and research duties and ten were involved in research and teaching, five of the participants were veterinary surgeons and six were medical clinicians. It was an international panel including ten participants from the United Kingdom, three from Australia, two from the United States of America, two from Canada, and one from Egypt. Participants were qualified a mean of 17.6 years (SD: 7.9) and the panel was made up of participants from varying disciplines (Table 1)

Table 1. The number of participants from each discipline enrolled in the Delphi panel for the development of the AXIS tool

Discipline	**N**
Epidemiology	4
Evidence-based medicine	9
Evidence-based veterinary medicine	2
Public Health	3

During round one (undertaken in February 2013) of the Delphi process, 20 components reached consensus, 13 components were assessed to require modification and it was deemed appropriate to remove four components from the tool. General comments mostly related to the tool having too many components.

> *" the tool needs to be succinct and easy and quick to use if possible - too many questions could have an impact."*
>
> *"List is too long at present and contains too many things that are general to all scientific studies"*
>
> Comments voiced included the discussion as part of the critical appraisal process being unnecessary and potentially misleading:
>
> *"the interpretation should, in my opinion, come from the methods and the results and not from what the author thinks it means"*
>
> *"I don't believe a Discussion section should be part of a critical appraisal."*

Therefore, in round one the tool was modified in an attempt to reduce its size and to encompass all comments. For round two (undertaken in May 2013) 11 components remained the same and did not require testing for consensus as this was established in round one; nine components that had previously reached consensus were incorporated with the 13 components that required modification to create 10 new components (Supplementary material Table 4).

In round two, consensus was reached on a further two components, six components were assessed to require modification and it was deemed appropriate to remove two components from the tool. Comments from the panel regarding the components of the tool that related to the discussion, suggested further reduction in these components due to their limited use as part of the critical appraisal process.

> *"the discussion could legitimately be highly speculative and not justified by the results provided that the authors don't present this as conclusions."*

With the reduction in the number of questions and modification of the wording, comments in round 2 reflected the positive nature to the usability of the tool.

> *"I like the fact that it is quite simple - not too overloaded with methodological questions."*

After round 2, the tool was further reduced in size and modified to create a 4th draft of the tool with 20 components incorporating 13 components with full consensus and seven modified components for circulation in round 3 of the Delphi process.

Following round 3 (undertaken in July 2013) of the Delphi process, there was consensus (81%) that all components of the tool were appropriate for use by non-expert users so no further rounds were necessary. The final critical appraisal tool for cross sectional studies (AXIS tool) consisting of 20 components is shown in Table 2. The comments from the panel regarding the help text were addressed and minor modifications to the text were made (Supplementary material 4). Seven (1, 4, 10, 11, 12, 16 and 18) of the final questions related to quality of reporting, seven (2, 3, 5, 8, 17, 19 and 20) of the questions related to study design quality, and six related to the possible introduction of biases in the study (6, 7, 9, 13, 14 and 15).

Table 2. The final AXIS tool following consensus on all components by the Delphi panel.

		Yes	No	Don't know/ Comment
Introduction				
1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			

13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results for the analyses described in the methods, presented?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

DISCUSSION

The AXIS Tool

A critical appraisal tool to assess the quality and risk of bias in cross-sectional studies (AXIS), along with supporting help text, was successfully developed by an expert panel using Delphi methodology. This is the first critical appraisal tool made available for assessing this type of evidence that can be incorporated in systematic reviews, guidelines and clinical decision-making.

Key Delphi Findings

One of the key items raised in comments from the experts was assessing quality of design versus quality of reporting. It is important to note that a well reported study may be of poor quality and conversely a poorly reported study could be a well conducted study[33 34]. It is also apparent that if a study is poorly reported it can be difficult to assess the quality of the study. Some information may be lacking due to poor reporting in studies making it difficult to assess the risk of biases and the quality of the study design. High quality and complete reporting of studies is a pre-requisite for judging quality[17 18 35]. For this reason, the AXIS tool incorporates some quality of reporting as well as quality of design and risk of biases to overcome these problems.

Using the tool

The tool was also reduced in size on each round of the Delphi process as commentators raised concerns around developing a tool with too many questions. The comments suggested that a long questionnaire would lead to the tool being cumbersome and difficult to use, and for this reason efforts were made to develop a much more concise tool.

The AXIS tool focuses mainly on the presented methods and results. It was the view of the Delphi group that the assessment as to whether the published findings of a study are credible and reliable should relate to the aims, methods and analysis of what is reported and not on the interpretation (e.g. discussion and conclusion) of the study. This view is also seen in other appraisal tools, is shared by other researchers and can be seen by the absence of questions relating to the discussion sections in critical appraisal tools for other types of studies[12 16 20 28 36].

As with all critical appraisal tools it is only possible for the reader to be able to critique what is reported. If an important aspect of a study is not in the manuscript, it is unclear to the reader whether it was done, and not reported, or not done at all. It is therefore the responsibility of the appraiser of the study to recognise omissions in reporting and consider how this affects the reliability of the results.

A comprehensive explanatory text is often used in appraisal tools for different types of study designs as it aids the reviewer when interpreting and analysing the outputs from the appraisal [12 17-20]. This approach was also used in the development of the AXIS tool where a reviewer can link each question to explanatory text to aid in answering and interpreting the questions.

Study strengths and limitations

The tool was developed through a rigorous process incorporating comprehensive review, testing and consultation via a Delphi panel. Using a similar process to other appraisal tools[37], we reviewed the relevant literature to develop a concise background on critical appraisal of cross sectional studies and to ensure no other relevant tools existed. While numerous tools exist for critical appraisal, we found a lack of tools for general use in cross sectional studies and this was consistent with what others have found previously[12 13]. In order to ensure quality and completeness of the tool we utilized recognised reporting guidelines, other appraisal tools and epidemiology design text in the development of the

initial tool which is similar to the development of appraisal tools of other types of studies[12].

The use of a multidisciplinary panel with experience in epidemiology and evidence-based medicine limits the effect of using a non-representative sample, and the use of the Delphi tool is well recognised for developing consensus in health care science[38]. The selection of a Delphi group is very important as it effects the results of the process[31]. As cross-sectional studies are used extensively in both human and veterinary research it was appropriate to use expertise from both of these fields. To ensure that the tool was developed to a high standard, a high level of consensus was required in order for the questions to be retained[31 32 39]. There was a high level of consensus between both veterinary and medical groups in this study, which adds to the rigour of the tool but also demonstrates how both healthcare areas can co-operate effectively to produce excellent outcomes.

The Delphi study was conducted using a carefully selected sample of experts and as such may not be representative of all possible users of the tool. However the purpose of a Delphi study is to purposely hand pick participants that have prior expertise in the area of interest[40]. The Delphi members came from a multidisciplinary network of professionals from medicine, nursing and veterinary medicine with experience in epidemiology and EBM/EVM and exposure to teaching and areas of EBM that were not just focused on systematic reviews of randomised controlled trials. The panel was restricted to those that were literate in the English language and may therefore not be representative of all nationalities. The interests and experiences of the panel will clearly have had an effect on the results of this study as this is common to all Delphi studies[31 41]. The majority of Delphi studies are conducted using between 15 and 20 participants[31], so a panel of 18 is consistent with other published Delphi panels. We aimed to recruit a minimum of 15 participants and as it was anticipated that not all participants contacted would be able to take part, more participants were contacted.

As the tool does not provide a numerical scale for assessing the quality of the study, a degree of subjective assessment is required. This has implications for interpretation after using the tool as there will be differences in individuals' judgements. However, it has been debated that quality numerical scales can be problematic as the outputs from assessment

checklists are not linear and as such are difficult to sum up or weight making them unpredictable at assessing study quality[39 42 43]. The AXIS tool has the benefit of providing the user the opportunity to assess each individual aspect of study design to give an overall assessment of the quality of the study. By providing this subjectivity, AXIS gives the user more flexibility in incorporating both quality of reporting and risk of bias when making judgments on the quality of a paper. This tool therefore provides an advantage over, Berra et al (2008) [15] which only allows the user to assess quality of reporting and tools such as the Cochrane risk of bias tool [5] which do not address poor reporting. Further studies would be needed to assess how practical this tool is when used by clinicians and if the critical appraisal of studies using AXIS is repeatable.

CONCLUSION

In conclusion, a unique tool (AXIS) for the critical appraisal of cross sectional studies was developed that can be used across disciplines e.g. health research groups and clinicians conducting systematic reviews, developing guidelines, undertaking journal clubs and private personal study. The components of the AXIS tool are based on a combination of evidence, epidemiological processes, experience of the researchers and Delphi participants. As with other evidence-based initiatives, the AXIS tool is intended to be an organic item that can change and be improved where required, with the validity of the tool to be measured and continuously assessed. We would invite any users of the tool to provide feedback so that the tool can be further developed if needed and can incorporate user experience to provide better usability.

ACKNOWLEDGMENTS:

The authors thank the following individuals who participated in the Delphi process: Peter Tugwell, Thomas McGinn, Kim Thomas, Mark Petticrew, Fiona Bath-Hextall, Amanda Burls, Sharon Mickan, Kevin Mackway Jones, Aiden Foster, Ian Lean, Simon More, Annette O'Connor, Jan Sargeant, Hannah Jones, Ahmed Elkhadem, Julian Higgins and Sinead Langan.

The authors would like thank those who piloted the tool in the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in

evidence-based veterinary medicine. The authors would also like thank Michelle Downes for designing the population diagram.

The Centre for Evidence-based Veterinary Medicine is supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Contributorship statement

MJD, MLB and RSD made substantial contribution to the conception and design of the work, as well as the acquisition, analysis, or interpretation of data. They were directly involved in the preparation and revision of the intellectual content of the manuscript submitted for publication. They will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

HCW made substantial contribution to the conception and design of the work, interpretation of data and the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

Competing interests

The authors have no competing interests.

Funding

The Centre for Evidence-based Veterinary Medicine and this project was supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Data sharing statement

The authors are willing to share any of the data collected during the Delphi process that is not in this published manuscript or the supplementary material.

REFERENCES

1. Cockcroft PD HM. *Handbook of Evidence-based Veterinary Medicine.* . 1st ed. Oxford, UK: Blackwell Publishing Ltd, 2003`.
2. McColl A, Smith H, White P, et al. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. Bmj 1998;**316**(7128):361-5

3. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. Bmj 1995;**310**(6987):1122-6
4. Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. Bmj 2001;**323**(7317):833-6
5. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*: Wiley Online Library, 2008.
6. Bennett KJ, Sackett DL, Haynes RB, et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. Jama 1987;**257**(18):2451-4
7. Mogford E, Gould L, Devoght A. Teaching critical health literacy in the US as a means to action on the social determinants of health. Health Promot Int 2011;**26**(1):4-13 doi: DOI 10.1093/heapro/daq049[published Online First: Epub Date]|.
8. Parkes J, Hyde C, Deeks J, et al. Teaching critical appraisal skills in health care settings. The Cochrane database of systematic reviews 2001(3):CD001270 doi: 10.1002/14651858.CD001270[published Online First: Epub Date]|.
9. Sackett DL. Teaching critical appraisal. Journal of general internal medicine 1990;**5**(3):272
10. Ho PM, Peterson PN, Masoudi FA. Evaluating the evidence: is there a rigid hierarchy? Circulation 2008;**118**(16):1675-84 doi: 10.1161/CIRCULATIONAHA.107.721357[published Online First: Epub Date]|.
11. Dohoo IR, Martin SW, Stryhn H. *Veterinary Epidemiologic Research*. Charlottetown, Canada: Ver Books, 2009.
12. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC medical research methodology 2004;**4**:22 doi: 10.1186/1471-2288-4-22[published Online First: Epub Date]|.
13. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International journal of epidemiology 2007;**36**(3):666-76 doi: 10.1093/ije/dym018[published Online First: Epub Date]|.
14. Sterne JAC HJ, Reeves BC on behalf of the development group for ROBINS-I. ROBINS-I: a tool for assessing Risk Of Bias In Non-randomized Studies of Interventions, Version 5. 2016
15. Berra S, Elorza-Ricart JM, Estrada M-D, et al. [A tool (corrected) for the critical appraisal of epidemiological cross-sectional studies]. Gac. Sanit. 2008;**22**(5):492-7
16. Chalmers TC, Smith H, Jr., Blackburn B, et al. A method for assessing the quality of a randomized control trial. Controlled clinical trials 1981;**2**(1):31-49
17. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. B World Health Organ 2007;**85**(11):867-72 doi: Doi 10.2471/Blt.07.045120[published Online First: Epub Date]|.
18. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999;**354**(9193):1896-900 doi: Doi 10.1016/S0140-6736(99)04149-5[published Online First: Epub Date]|.
19. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;**63**(8):834-40 doi: DOI 10.1016/j.jclinepi.2010.02.005[published Online First: Epub Date]|.
20. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology 2007;**7** doi: Artn 10

Doi 10.1186/1471-2288-7-10[published Online First: Epub Date]|.
21. BestBETs. BETs CA Worksheets. Secondary BETs CA Worksheets 2009. http://bestbets.org/links/BET-CA-worksheets.php.
22. Crombie I. *The Pocket Guide to Critical Appraisal. annotated edition. Wiley-Blackwell.* 1 ed. London, UK: BMJ Publishing group, 1996.

23. Fowkes FGR, Fulton PM. Critical-Appraisal of Published Research - Introductory Guidelines. Brit Med J 1991;**302**(6785):1136-40
24. Genaidy AM, Lemasters GK, Lockey J, et al. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. Ergonomics 2007;**50**(6):920-60 doi: 10.1080/00140130701237667[published Online First: Epub Date]|.
25. Hanke BA, Motschall E, Turp JC. Association between orthopedic and dental findings: What level of evidence is available? J. Orofac. Orthop. 2007;**68**(2):91-107 doi: 10.1007/s00056-007-0634-0[published Online First: Epub Date]|.
26. Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional, 2004.
27. Rothman K. *Chapter 5: Bias in Study Design. Epidemiology: An Introduction*. New York: Oxford Univeristy Press, 2002.
28. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerging themes in epidemiology 2008;**5**:23 doi: 10.1186/1742-7622-5-23[published Online First: Epub Date]|.
29. Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual 2014 Edition*, 2014.
30. Downes MJ, Dean RS, Stavisky JH, et al. Methods used to estimate the size of the owned cat and dog population: a systematic review. BMC veterinary research 2013;**9**:121 doi: 10.1186/1746-6148-9-121[published Online First: Epub Date]|.
31. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Practical assessment, research & evaluation 2007;**12**(10):1-8
32. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 2004;**42**(1):15-29 doi: http://dx.doi.org/10.1016/j.im.2003.11.002[published Online First: Epub Date]|.
33. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. Bmj 2004;**328**(7430):39-41 doi: 10.1136/bmj.328.7430.39[published Online First: Epub Date]|.
34. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj 2001;**323**(7303):42-6
35. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama 2000;**283**(15):2008-12
36. Wells G, Shea B, O'connell D, Peterson J, Welch V, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Secondary The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000. http://www.medicine.mcgill.ca/rtamblyn/Readings/The%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf.
37. Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011;**64**(1):79-89 doi: DOI 10.1016/j.jclinepi.2010.02.008[published Online First: Epub Date]|.
38. Fink A, Kosecoff J, Chassin M, et al. Consensus Methods - Characteristics and Guidelines for Use. Am J Public Health 1984;**74**(9):979-83 doi: Doi 10.2105/Ajph.74.9.979[published Online First: Epub Date]|.
39. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostat 2001;**2**:463-71 doi: 10.1093/biostatistics/2.4.463[published Online First: Epub Date]|.
40. Mckenna HP. The Delphi Technique - a Worthwhile Research Approach for Nursing. J Adv Nurs 1994;**19**(6):1221-25 doi: DOI 10.1111/j.1365-2648.1994.tb01207.x[published Online First: Epub Date]|.

41. Linstone HA, Turoff M. *The Delphi method: Techniques and applications*: Addison-Wesley Reading, MA, 1975.
42. Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Brit Med J 2011;**343** doi: Artn D5928

Doi 10.1136/Bmj.D5928[published Online First: Epub Date]|.
43. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. Jama-J Am Med Assoc 1999;**282**(11):1054-60 doi: DOI 10.1001/jama.282.11.1054[published Online First: Epub Date]|.

Supplementary material

Table 1: Search Terms Used to identify critical appraisal tools for cross sectional studies

Database	Number	
Medline 1948 to September Week 3 2011	71	Critical appraisal.mp. AND (exp Cross-Sectional Studies/ OR cross sectional.mp.)
CAB 1910 to 2011 Week 38	4	Critical appraisal.mp. AND cross sectional.mp.
Web of Science^{SM} (1899-present)	60	Topic=(Critical appraisal) AND Topic=(cross sectional)
BIOSIS Previews® (1969-present)	12	Topic=(Critical appraisal) AND Topic=(cross sectional)
Zoological Record® (1978-present)	0	Topic=(Critical appraisal) AND Topic=(cross sectional)
Embase 1974 to 2011 October 03	65	Critical appraisal.mp. AND (exp cross-sectional study/ OR cross sectional.mp.)
CINAHL® with Full Text	23	((MM "Cross Sectional Studies") OR "cross sectional") AND "Critical appraisal"
PsycINFO 1806 to September Week 4 2011	9	Critical appraisal.mp. AND cross sectional.mp.
Total	244	

Table 2. The 1st draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post review of the literature.

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Are the aims of the study clearly stated?			
Methods				
2	Is the type of study design appropriate for the stated aim?			
3	Is the sample size justified (based on pre-study considerations of statistical power)?			
4	Is the target or reference population clearly defined? (is it clear who the research was about)			
5	Is the sample frame taken from an appropriate population base so that it closely represents the overall population under investigation?			
6	In the selection process:			
	a. Were any inclusion/exclusion criteria used?			
	b. Was random selection used to obtain participants?			

7	Is the selection process likely to select subjects that were representative of the study population of interest?			
8	If appropriate, were measures undertaken to address and categorise non-responders?			
9	Do the variables measured, in the study, produce data that reflect the aims of the study? (Validity)			
	a. Are the outcomes of interest clearly measured?			
	b. Are the risk factors appropriately measured to be compared to the outcomes of interest?			
10	If appropriate, have the measurement instruments been trialled, piloted or published previously? (Reliability and reproducibility)			
11	Are the statistical methods clearly stated?			
12	If appropriate, is the means by which statistical significance is inferred stated?			
13	Are the methods sufficiently described to enable them to be repeated?			
Results				
14	Are the basic data adequately described?			
15	Is the response rate given, if appropriate?			
16	Is information about non-responders described, if appropriate?			
17	Are the results internally consistent?			
	a. Do the numbers add up?			
	b. Are any missing data acknowledged, or described?			
18	Are the results described objectively without author opinion?			
19	Are results pertaining to the study aim reported?			
20	If appropriate is the statistical significance level declared in the methods adhered to?			
21	Are the results of all tests described in the methods presented?			
Discussion				
22	Are all results pertaining to the study aim discussed?			
23	Are the limitations of the study discussed?			
24	Is selection bias addressed?			
25	Is non-response addressed?			
26	Do the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?			
27	If appropriate are non-significant results discussed?			
	a. Do the authors consider issues around study design when interpreting non-significant results?			
	b. Do the authors consider issues around sample size when interpreting non-significant results?			
Conclusions				
28	Are the authors' conclusions justified by the results?			
Other				
29	Are any conflicts of interest/funding declared in the text?			
30	Was ethical aspect approval or consent of participants attained?			

Table 3. The 2nd draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post piloting with the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. This draft was used in the first round of the Delphi panel and the results of the consensus from the panel or each component are presented.

	Consensus
Introduction	
1. Is it clear what the aims of the study were?	94.12
Methods	
2. Was the type of study design appropriate for the stated aim?	94.12
3. Was the sample size justified (based on pre-study considerations of statistical power)?	76.47
4. Was the target or reference population clearly defined? (is it clear who the research was about?)	100.00
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	88.24
6.a. In the selection process: Were any inclusion/exclusion criteria used?	82.35
6.b. In the selection process: Was random selection used to obtain participants?	70.59
7. Was the selection process likely to select subjects that were representative of the study population of interest?	93.33
8. If appropriate, were measures undertaken to address and categorise non-responders?	87.50
9. Did the variables measured in the study, produce data that reflected the aims of the study? (Validity)	68.75
9.a. Were the outcomes of interest clearly measured?	86.67
9.b. Were the risk factors measured appropriate to the outcomes of interest?	66.67
10. If appropriate, had the measurement instruments been trialled, piloted or published previously?	93.33
11. Is it clear what statistical methods were used?	86.67
12. If appropriate is it possible to determine the means by which the statistical significance was inferred? (p-values, confidence intervals)	64.29
13. Were the methods sufficiently described to enable them to be repeated?	87.50
Results	
14. Were the basic data adequately described?	75.00
15. If appropriate, was the response rate sufficient to base conclusions on?	93.75
16. If appropriate, was information about non-responders described?	93.75
17. Were the results internally consistent?	57.14
17.a. Did the numbers add up?	56.25
17.b. Were any missing data acknowledged, or described, if appropriate?	93.75

18. Were the results described objectively without author opinion?	43.75
19. Were the results pertaining to the study aim reported?	81.25
20. If appropriate, was the statistical significance level declared in the methods adhered to?	68.75
21. Were the results of all tests described in the methods presented?	93.75
22. Were all results pertaining to the study aim discussed?	56.25
Discussion	
23. Were the limitations of the study discussed?	81.25
24. Was selection bias discussed appropriately?	62.50
25. Was non-response discussed appropriately?	68.75
26. Did the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?	53.33
27. If appropriate, were non-significant results discussed?	50.00
27.a. Did the authors consider issues around study design when interpreting non-significant results?	50.00
27.b. Did the authors consider issues around sample size when interpreting non-significant results?	50.00
28. Were the authors' conclusions justified by the results?	93.75
Other	
29. Were any conflicts of interest/funding declared in the text?	93.75
30. Was ethical approval or consent of participants attained?	81.25

Table 4. The 3rd draft of a CA tool created following round 1 of the Delphi study after comments and consensus was taken into account. Results on consensus for each question from the round 2 of a Delphi panel are presented.

	Consensus*
Introduction	
1. Is it clear what the aims of the study were?	
Methods	
2. Was the type of study design appropriate for the stated aim?	
3. If appropriate, was the sample size justified?	68.75
4. Was the target or reference population clearly defined? (Is it clear who the research was about?)	
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	
6. Was the selection process likely to select subjects that were representative of the study population of interest?	
7. Were measures undertaken to address and categorise non-responders?	81.25
8. Would the variables measured in the study produce data that reflected the aims of the study? (Validity)	62.5
9. Is it clear what statistical methods were used?	50
10. Is it clear how statistical significance was determined? (eg: p-values, confidence intervals)	62.5
11. Were the methods sufficiently described to enable them to be repeated?	
Results	
12. Were the basic data adequately described?	
13. If appropriate, was the response rate sufficient to enable sound conclusions to be drawn?	56.25
14. If appropriate, was information about non-responders described?	
15. Were the results internally consistent?	
16. Were all the results of the analyses described in the methods presented?	60
17. If a statistical significance level was declared in the methods, was it adhered to in the results?	31.25
Discussion	
18. Were the authors' discussions and conclusions justified by the results?	87.5
19. Were the limitations of the study discussed?	
Other	
20. Were there any funding sources or conflicts of interest that were likely to affect the authors' interpretation of the results?	75
21. Was ethical approval or consent of participants attained?	

**Where no consensus figure is given, consensus was reached on this question in the previous round.*

Appraisal tool for Cross-Sectional Studies (AXIS)

Critical appraisal (CA) is used to systematically assess research papers and to judge the reliability of the study being presented in the paper. CA also helps in assessing the worth and relevance of the study [1]. There are many key areas to CA including assessing suitability of the study to answer the hypothesised question and the possibility of introducing bias into the study. Identifying these key areas in CA requires good reporting of the study, if the study is poorly reported the appraisal of suitability and bias becomes difficult.

The following appraisal tool was developed for use in appraising observational cross-sectional studies. It is designed to address issues that are often apparent in cross-sectional studies and to aid the reader when assessing the quality of the study that they are appraising. The questions on the following pages are presented in the order that they should generally appear in a paper. The aim of the tool is to aid systematic interpretation of a cross-sectional study and to inform decisions about the quality of the study being appraised.

The appraisal tool comes with an explanatory help text which gives some background knowledge and explanation as to what the questions are asking. The explanations are designed to inform why the questions are important. Clicking on a question will automatically take you to the relevant section in the help text. The appraisal tool has areas to record a “yes”, “no” or “don’t know” answer for each question and there is room for short comments as well.

Contents

Appraisal of Cross-sectional Studies

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			
13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results presented for all the analyses described in the methods?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

Introduction

The introduction serves to establish the context of the work that is about to be presented in the text of the paper. Relevant primary literature should be discussed and referenced throughout the introduction. The history and current understanding of the problem being researched should be presented. This should be concluded giving a rational as to why the current study is being presented and what the aims and/or hypothesis under investigated are [2,3].

Aims

The aim(s) of the study tells us if the study addresses an appropriate and clearly focused question. If the aim is not clearly stated or not stated at all, it will be difficult and in some cases impossible to assess the extent to which the study objectives were achieved. Ideally, an aim should be stated both at the beginning of the abstract and at the end of the introduction [3]. If the answer to question 1 is no, then it will make it difficult to assess some of the other questions in the critical appraisal process.

Methods

The methods section is used to present the experimental study design of the paper. The methods should be described clearly in easy to understand language and clearly identify measures, exposures and outcomes being used in the study [4]. More specific issues are addressed below.

Study Design

Question 2 is used to assess the appropriateness of using a cross-sectional study to achieve the aim(s) of the study. Cross-sectional studies are observational studies that provide a description of a population at a given time, and are useful in assessing prevalence and for testing for associations and differences between groups [5]. Examples of cross-sectional designs include point-in-time surveys, analysis of records and audits of practice [6]. The reader should try and decipher if a cross-sectional study design is appropriate for the questions being asked by the researcher.

Sample Size Justification

Sample size justification is crucial as sample size profoundly affects the significance of the outcomes of the study. If the sample size is too small then the conclusions drawn from the study will be under powered and may be inaccurate. This can occur by failing to detect an effect which truly exists (type II error) sometimes referred to as a "false negative". The probability of a type I error is also taken into account when determining sample size. A type I error is drawing significant conclusions when no real difference exists and is a function of the p-value (see Statistics section below) sometimes referred to as a "false positive".

Question 3 asks if sample size justification was reported, but it should also be clear what methods were used to determine the sample size. In some cases clustering of observations within groups can occur (e.g. patients within hospitals or livestock within herds) and this should be taken into account if sample size has been determined. It should be clear whether the inferences drawn actually relate to the attributes for which the sample size was calculated [7]. If sample size justification isn't given or restrictions make it difficult to reach the desired sample size then this should be declared in the text.

Target (Reference) Population

The target or reference population is the overall population that the research is directed towards. When doing a cross-sectional study, a target population is the overall population you are undertaking the study to make conclusions about or the population at risk of acquiring the condition being investigated [8–10] e.g. the total female population in the UK, or all dogs in the USA with cardiovascular disease. (See Figure 1) Question 4 asks if this is clearly defined in the study. It is important that this is understood both by the researcher and the reader; if it is not clearly defined then inferences made by the researcher may be inappropriate.

Sampling Frame

As a reader you need to determine if the sample frame being used is representative of the target population. The study population should be taken from the target population; units from this study population have information that is accessible and available which allows them to be placed in the study. The sampling frame is the list or source of the study population that the researcher has used when trying to recruit participants into the study (Figure 1). Ideally it should be exactly the same composition or structure as the target population. In practice it is generally much smaller, but should still be representative of the target population. Generally, for convenience, the sampling frame is a list of units that are within the target population e.g. list of

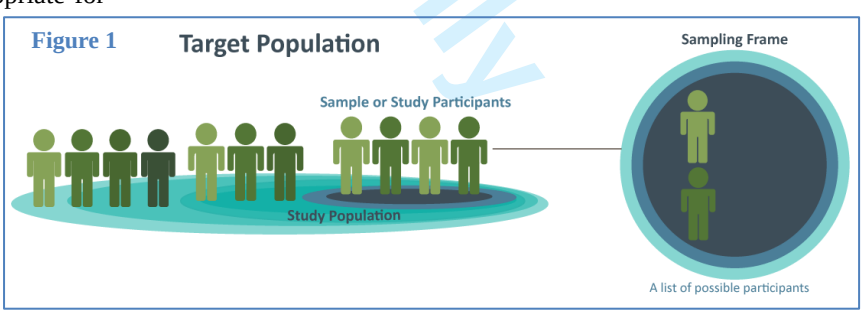

Figure 1

telephone owning households, computerised patient records etc. A sample of units is selected from the study population to take part in the study and is generally only a small proportion of the study population (see Sample Selection below) - this proportion ratio is known as the sampling fraction. It is very important that the sampling frame is representative of the target population as results from the study are going to be used to make assumptions about the target population [8–10].

Convenience sampling can be carried out in some situations and are used because the participants are easy to recruit. Convenience samples generally lead to non-representative or biased samples and therefore cannot be used to make assumptions about the characteristics of the target population [11]. Convenience samples are often used for pilot or analytical studies where the need for a representative sample is not required [12], however the authors should make this clear in the text.

Census

A census is where the target population and the study participants are the same at the time the census is taken. In theory questions 5, 6 and 7 don't apply to census studies. However even if a study is described as a census it should be very clearly stated where the study participants have been recruited from, and the reader should make the decision if the study truly is a census. A census may include all the population from the sample frame, but not all the target population; in this scenario questions 5 to 7 need to be addressed.

Sample Selection

Question 6 is used to establish how the researchers got from the sample frame to the participants in the study. It examines the potential for selection bias and how the researcher developed methods to deal with this. The sample selection process is important in determining to what extent the results of the study are generalizable to the target population. For question 6 we are looking in depth at how the sample (study participants) was selected from the sampling frame. It is important to know if there were any inclusion or exclusion criteria used, as inappropriate criteria can dramatically shift how representative the sample is of the target population [8,10,13].

Selection bias can occur if every unit in the sample frame doesn't have an equal chance of been included in the final study [11,14]. Randomisation is used to ensure that each participant in the sampling frame has an equal chance of being included in the sample. If methods of randomisation are not used, not described or are not truly random, this may lead to a non-representative sample being selected and hence affect the results of the study [10,11].

There are many other situational issues to take into account when determining if the population in the sample is likely to represent the target population. Often these issues are outside the control of the researcher, but sometimes are overlooked. One such issue is the healthy worker effect which is a well-known phenomenon in human cross-sectional studies [13]. An example of this is, a researcher trying to do a cross-sectional study to determine health factors in a factory population and decides to sample from workers at work on a particular day. Unfortunately there is a tendency to over select healthy workers as ill workers may tend to be at home on the day of selection. This will in turn lead to inferences been made about the health of the worker population but is only relevant to healthy workers and not ill workers. A veterinary example of this is a researcher trying to do a cross-sectional study to determine health factors in the general dog population and decides to sample from a local park. Unfortunately there is a tendency to over select healthy animals as sick animals will tend to be left at home and not taken for a walk. This will in turn lead to inference been made about the health of the dog population but is only relevant to healthy dogs and not sick dogs.

Self-selection is another example of selection bias that can be introduced and should be assessed [13]. For example, when using a postal questionnaire to examine eating habits and weight control, people who are overweight might read the survey and be less inclined to complete and return the survey than those with normal weight leading to over representation of people with normal weight. Similarly, if using a postal questionnaire to examine mastitis levels on cattle farms, farmers that have a high somatic cell counts (SCC) might be less inclined to complete the survey than those with normal or low SCC leading to over representation of farms with good SCC (see Non-responders below).

Non-responders

Non-response in cross-sectional studies is a difficult area to address. A non-responder is someone who does not respond either because they refuse to, cannot be contacted, or because their details cannot be documented. As a rule, if participants don't respond it is often difficult and sometimes impossible to gain any information about them. However other baseline statistics may exist that can be used as a comparator to assess how representative the sample is [14] e.g. age, sex, socio-economic classification. Methods used, if any, should be well described so that the results from the analyses can be interpreted. This is important as non-responders may be from a specific group, which can lead to a shift in the baseline data away from that group. This shift can lead to results that don't represent the target population. In some situations the sampling frame doesn't have a finite list or a fully defined baseline population. This also makes it difficult, and in some cases impossible, to quantify non-response and it may be inappropriate to do so in these situations. If the researchers are using non-defined populations this should also be declared clearly in the materials and methods section [15,16].

Measurement Validity & Reliability

Measurement validity is a gauge of how accurately the study measurements used assess the concepts that the researcher is attempting to explore. Measurement reliability is a gauge of the accuracy of the measurements taken or the procedures used during the study. Question 8 is used to address the concepts of measurement validity, and is specifically aimed to address the appropriateness of the measurements being used.

The importance of measurement validity is that it gives weight to applying the statistical inferences from the study to members of the target population. If inappropriate measures are used in the study it could lead to misclassification bias and it will be difficult to determine to what extent the study results are relevant to the target population [12,17].

Question 9 is an attempt to gauge the measurement reliability of the study measures. Measurements must be able to be reproduced and produce identical results if measured repeatedly, so that the measurements would be exactly the same if performed by another researcher. With this in mind, the measurements must be of international or globally accepted standards (e.g. IU standards) where possible and appropriate. If they are being used for the first time they must be trialled, or in the case of questionnaires, they should be piloted before being used.

Statistics

While interpretation of statistics can be quite difficult, a basic understanding of statistics can help you to assess the quality of the paper. Often many different methods can be used correctly to test the same data, but as there is such a wide range available, knowing what tests are most appropriate in particular situations can be hard to decipher. There is an expectation that the researcher has this understanding or has at least sought statistical assistance to ensure that the correct methods are used. Therefore for question 10 the emphasis for the reader is that the statistical methods, software packages used and the statistical significance levels are clearly stated even if the paper is just presenting descriptive statistics. The statistical significance level is usually described as a p-value. In most cases the p-value, at which the null hypothesis is rejected, is set at 0.05. The higher the p-value is set the greater the possibility of introducing a type I error. Confidence intervals should also be declared with p-values or instead of p-values as an indication of the precision of the estimates. It is usual to present a confidence interval of 95% which means that the researchers were 95 per cent confident that the true population value of the outcome lies between these intervals. This can be used to compare groups where an overlap would suggest no difference and a gap between confidence intervals would suggest a difference (Figure 2).

Figure 2

Overall Methods

Question 11 asks if the methods are sufficiently described to enable them to be repeated. If there are sections or even small pieces of information missing it could make a great difference for the reader when interpreting the results and the discussion as they may be unsure if the correct methods are being used.

Results

The results section of a paper is solely for the purpose of declaring the results of the data analysis and no opinion should be stated in this section. This gives the reader the opportunity to examine the results unhindered by the opinion of the researcher. It is important for the reader to form their own ideas or opinions about the results before progressing to the discussion stages.

Basic Data

Question 12 asks for a description of the basic data. Basic descriptive analysis aims to summarise the data, giving detailed information about the sample and the measurements taken in the study. The basic data gives an overview of the process of recruitment and if the sampling methods used to recruit individuals were successful in selecting a representative sample of the target population. If the sampling methods are unsuccessful in selecting a representative sample of the target population, those participants included in the study can often be different to the target population; this leads to inaccurate estimates of prevalence, incidence or risk factors for disease. Descriptive data of the measurements taken in the study give an overview of any differences between the groups, and may give insight into some of the reasons for statistical inferences that are made later in the paper.

Response Rate

As stated previously it can often be difficult to deal with non-responders. Question 13 requires that there is some attempt made to quantify the level of non-response by the researchers and asks the reader to interpret if the response rate is likely to lead to non-response bias. Question 14 is examining if any information on non-responders was available and if so were they comparable to those that did respond as this could help in answering question 13. Non-response bias occurs if the non-responders are substantially different to the rest of the population in the sample [15].

Internally Consistent Results

Question 15 is an exploration of the basic data and asks that the reader spends some time exploring the numbers given in the results; in the text, figures and tables. Information about the level of missing data should also be declared in the results. It is important to check that the numbers add up in the tables and the text. If the study has recruited 100 participants, the tables and the text should include data about 100 participants. If not, the missing data should be clearly declared and the reason for its non-appearance explained.

Comprehensive Description of Results

It is important to check that all the methods described previously lead to data in the results section (question 16). Sometimes the results from all analyses are not described. If this is noted it will be unclear whether the researcher found non-significant results or just didn't describe what was found. If there are results missing that you would expect to find, there is a concern that these missing results may not have been what the researcher wanted to see and hence the authors have omitted them. It is also important that the significance level declared in the methods is adhered to. As the reader, it is important to watch out for phrases such as "tended towards significance" in the text, and if these are used to pay close attention to the results.

Discussion

The discussion of a paper should summarise key results of the study objectives. It should give an overall interpretation of the results of the study keeping in mind the limitations and the external validity of the document. The discussion section should also address both significant and non-significant findings of the study and make comparisons with other research, citing their sources [2,4].

Justified Discussions and Conclusions

In question 17 there is an expectation that the researcher gives an overall summary of the main findings of the study and discusses these in detail. It is important that the reader considers the study as a whole when reading the researcher's conclusion. If the researcher's conclusion is different or is more definitive than the study suggests it should be, it can be an indication that the researcher has misunderstood their own study or has other motives or interests for coming to that conclusion.

It is up to the reader to explore the discussion fully in order to answer question 17. The following points should be taken into account:

Aim

In the discussion section the researcher should discuss all results that pertain to the overall aim of the study, even if they are not significant. If some results are overlooked in the discussion it could suggest that the researcher either doesn't believe the results, or doesn't want to draw attention to controversial discoveries from the study and may therefore be giving a biased overview of the research conducted.

Selection Bias

There is an expectation that the researcher discusses selection biases and takes these into account when interpreting the results of the study. This also gives a clear view of whether the researcher has an overall understanding of the study design. (See notes on selection bias in the methods section).

Non-response

Was there an interpretation of the results that included non-response? This is particularly important if the response rate was low, as non-responders may be a specific group, and lead to a shift in the baseline data (See notes on non-response in the methods section).

Confounding

Confounding is a major threat to the validity of practical inferences made from statistical analyses about cause and effect. Confounding occurs when the outcome of interest is associated with two different independent variables and one of those variables is closely associated with the outcome only because it is closely associated with the other variable (confounder). This can sometimes be accounted for using statistical methods however sometimes these associations are missed because the confounder isn't measured or isn't considered to be a confounder in the analyses. What then happens is an erroneous conclusion is made; that the variable might have a causal relationship with the outcome. The researcher should consider confounding both in the analyses and in the interpretation of the results [18]. An example would be where in a study on cancer a researcher concludes that increased alcohol intake causes lung cancer; however there was confounding in the sample that the researcher didn't discover. People in the study that were inclined to drink more alcohol were also inclined to smoke more (the confounder) and smoking was the cause of lung cancer not increased alcohol intake. Similarly, a study was undertaken to examine surgical deaths in cats. The researcher concluded that cats that had gaseous anaesthesia were more likely to die during surgery than those that had just injectable anaesthesia. There was confounding in the sample: cats that underwent surgery using gaseous anaesthesia were more likely to be ill or undergoing major surgical procedures (the confounders) and this was the cause for cats being more likely to die during surgery and not the use of gaseous anaesthetics.

Non-significant Results

Discussing non-significant results is as important as discussing significant results and should also be included in the discussion, especially if they have a direct association with the aim being investigated. Non-significant results can be influenced by factors associated with study design and

sample size. If there are biases introduced during the study design this can lead to non-significant results that in reality may be significant (this can work the other way around as well). If there are only small differences between groups, non-significant results may be apparent because the sample size is too small (see sample size justification). Again it is important that the researcher has a clear understanding of this and conveys that in the discussion.

Limitations

In question 18 we explore whether limitations are discussed. Unfortunately all forms of research have some limitations. The question here is whether the researcher has an understanding of the limitations involved in their study design. If this issue is not explored, this is cause for concern that the limitations don't stop at the design and that the researcher has a poor understanding of the study as a whole.

Other

Conflicts of Interest

It is very important that conflicts of interest or bodies involved in funding the study are declared in the text (question 19). This can give an impression as to background reasons for carrying out the study. Where studies are funded by a specific agency the researcher may unconsciously interpret in favour of the agencies' ideals; if the researcher has worked in a specific area their own ideas and beliefs may affect the interpretation of the results. It is up to the reader to identify these and come to the conclusion as to whether these conflicts of interest are relevant or not. This can be declared in different areas of the text and should be stated.

Ethical Approval

Question 20 deals with ethical approval and participant consent. It is important that these are sought before carrying out research on any animal or person.

References:

1 Young JM, Solomon MJ. How to critically appraise an article. *Nature Clinical Practice Gastroenterology & Hepatology* 2009;**6**:82–91. doi:10.1038/ncpgasthep1331

2 Lebrun J-L. *Scientific Writing 2.0: A Reader and Writers Guide*. World Scientific 2011.

3 Masic I. How to Search, Write, Prepare and Publish the Scientific Papers in the Biomedical Journals. *Acta Inform Med* 2011;**19**:68–79. doi:10.5455/aim.2011.19.68-79

4 Von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 2007;**370**:1453–7. doi:10.1016/S0140-6736(07)61602-X

5 Leeuw ED, Hox JJ, Dillman DA. Data Analysis. In: *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 317–86.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

6 Rothman KJ. Chapter 4: Types of Epidemiologic Study. In: *Epidemiology: An Introduction*. New York: : Oxford University Press 2002. 57–93.

7 Woodward M. Chapter 8: Sample Size Determination. In: *Epidemiology: Study Design and Data Analysis, Second Edition*. Chapman and Hall/CRC 2004. 381–426.

8 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 1: An Introduction to Survey Methodology. In: *Survey methodology*. Hoboken, NJ: 2004. 1–37.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

9 Lohr SL. Chapter 6: Coverage and Sampling. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 97–112.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

10 Woodward M. Chapter 1: Fundamental Issues. In: *Epidemiology: Study Design and Data Analysis, Second Edition*. Chapman and Hall/CRC 2004. 1–32.

11 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 3: Target Populations, Sampling Frames and Coverage Error. In: *Survey methodology*. Hoboken, NJ: 2004. 67–92.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

12 Dohoo IR, Martin SW, Stryhn H. Capther 2: Sampling. In: *Veterinary Epidemiologic Research*. Charlottetown, Canada: : Ver Books 2009. 27–52.

13 Rothman KJ. Chapter 5: Bias in Study Design. In: *Epidemiology: An Introduction*. New York: : Oxford University Press 2002. 94–112.

14 Bruce N, Pope D, Stanistreet D. *Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics*. 1st ed. Wiley-Blackwell 2008.

15 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 2: Inference and Error in Surveys. In: *Survey methodology*. Hoboken, NJ: 2004. 39–65.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

16 Lynn P. Chapter 3: The Problem of Nonresponse. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 35–96.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

17 Hox JJ. Chapter 20: Accommodating Measurement Errors. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

18 Rothman KJ. *Epidemiology: An Introduction*. New York: : Oxford University Press 2002.

BMJ Open

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011458.R3
Article Type:	Research
Date Submitted by the Author:	07-Nov-2016
Complete List of Authors:	Downes, Martin; Griffith University School of Medicine, Centre for Applied Health Economics Brennan, Marnie; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science Williams, HC; University of Nottingham, Centre of Evidence Based Dermatology Dean, Rachel; University of Nottingham, Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Medical education and training, Patient-centred medicine, Research methods
Keywords:	Cross sectional studies, Critical appraisal, Evidence-based Healthcare, Delphi

Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS)

Martin J. Downes[1¶] Marnie L Brennan[2¶], Hywel C Williams[3], Rachel S Dean[2*¶]

[1]Centre for Applied Health Economics, School of Medicine, Griffith University, Meadowbrook, Australia

[2]Centre for Evidence-based Veterinary Medicine, The University of Nottingham, Sutton Bonington Campus, United Kingdom.

[3]Centre of Evidence Based Dermatology, The University of Nottingham, Kings Meadow Campus, Lenton Lane, Nottingham, United Kingdom.

* Email: Rachel.Dean@nottingham.ac.uk; (RSD)

[¶]These authors contributed equally to this work

ABSTRACT

Objectives: The aim of this study was to develop a critical appraisal tool that addressed study design and reporting quality as well as the risk of bias in cross sectional studies. In addition, the aim was to produce a help document to guide the non-expert user through the tool.

Design: An initial scoping review of the published literature and key epidemiological texts was undertaken prior to the formation of a Delphi panel to establish key components for a critical appraisal tool for cross sectional studies. A consensus of 80% was required from the Delphi panel for any component to be included in the final tool.

Results: An initial list of 39 components was identified through examination of existing resources. An international Delphi panel of 18 medical and veterinary experts was established. After 3 rounds of the Delphi process the Appraisal tool for Cross-Sectional Studies (AXIS tool) was developed by consensus and consisted of 20 components. A detailed explanatory document was also developed with the tool, giving expanded explanation of each question and providing simple interpretations and examples of the epidemiological concepts being examined in each question to aid non-expert users.

Conclusion: Critical appraisal of the literature is a vital step in evidence synthesis and therefore evidence-based decision making in a number of different disciplines. The AXIS tool is therefore unique and was developed in a way that it can be used across disciplines to aid the inclusion of cross sectional studies in systematic reviews, guidelines and clinical decision making.

Strengths and Limitations of this study

Strengths of this study include:

- The development of a novel critical appraisal tool that can be used across disciplines
- A multimodal evidence-based approach was used to develop the tool
- Expertise was harnessed from a number of different disciplines

Limitations of this study:

- The Delphi panel was based on convenience and may not encompass all eventual users of the tool.
- A numerical scale to reflect quality was not included in the final tool, which may be perceived as a limitation.

INTRODUCTION

Critical appraisal (CA) is a skill central to undertaking evidence-based practice which is concerned with integrating the best external evidence with clinical care. This is because when reading any type of evidence, being critical of all aspects of the study design, execution and reporting is vital for assessing its quality before being applied to practice[1-3]. Systematic reviews have been used to develop guidelines and to answer important questions for evidence-based practice[3 4] and critical appraisal to assess the quality of studies that have been included is a crucial part of this process [5]. Teaching critical appraisal has become an important part of the curriculum in medical schools and plays a central role in the interpretation and dissemination of research for evidence-based practice[6-9].

Traditionally evidence-based practice has been about utilising systematic reviews of randomised control trials (RCTs) to inform the use of interventions[10]. However other types/designs of research studies are becoming increasingly important in evidence-based practice, such as diagnostic testing, risk factors for disease and prevalence studies [10], hence systematic reviews in this area have become necessary. Cross sectional studies are one of those study designs that are of increasing importance in evidence-based medicine. A cross sectional study has been defined as: 'An observational study whose outcome frequency measure is prevalence. The basis of a cross sectional study design is that a sample, or census, of subjects is obtained from the target population and the presence or the absence of the outcome is ascertained at a certain point' [11]. Various reporting guidelines are available for the creation of scientific manuscripts involving observational studies which provide guidance for authors reporting their findings.

In addition, well developed appraisal tools have been created for readers assessing the quality of cohort and case control studies[12 13]; however, there is currently a lack of an appraisal tool specifically aimed at cross sectional studies (CSSs). The Cochrane collaboration has developed a risk of bias tool for non-randomised studies (ROBINS-I)[14]; however, this is a generic tool for case control and cohort studies that does not facilitate a detailed and specific enough appraisal to be able to fully critique a cross sectional study, In addition it is only intended for use to assess risk of bias when making judgments about an intervention. Two systematic reviews failed to identify a standalone appraisal tool

specifically aimed at cross sectional studies[12 13]. Katrak et al. (2004) identified that critical appraisal tools had been formulated specifically for individual research questions but were not transferable to other cross sectional studies. We identified an appraisal tool, developed in Spanish, which specifically examined cross sectional studies[15]. Berra et al (2008) essentially converted each reporting item identified in the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) reporting guidelines and turned them into questions for their appraisal tool. Fundamentally the tool developed by Berra et al (2008) only appraises the quality of reporting of cross sectional studies and does not address risk of bias or other aspects of study quality[15 16]. Good quality of reporting of a study means that all aspects of the methods and the results are presented well and in line with international standards such as STROBE [17]; however, this is only one aspect of appraisal as a well reported study does not necessarily mean that the study is of high quality. Bias ('a systematic error, or deviation from the truth, in results or inferences'[5]) and study design are other areas that need to be considered when assessing the quality of included studies as these can be inherent even in a well reported study.

As the need for the inclusion of cross sectional studies in evidence synthesis grows, the importance of understanding the quality of reporting and assessment of bias of cross sectional studies becomes increasingly important. Therefore, a robust critical appraisal tool to address the quality of study design and reporting to enable the risk of bias to be identified is needed. Delphi methods and use of expert groups are increasingly being implemented to develop tools for reporting guidelines and appraisal tools[18 19].

The aim of this study was to develop a critical appraisal tool that was simple to use, that addressed study design quality (design and reporting) and risk of bias in cross sectional studies. A secondary aim was to produce a document to aid the use of the critical appraisal tool where appropriate.

METHODS

Development of the initial draft Critical Appraisal tool:

The authors completed a systematic search of the literature for critical appraisal tools of cross sectional studies (Supplementary material Table 1). A number of publications were identified in the review and a number of key epidemiological texts were also identified to assist in the development of the new tool[1 11 12 15 17 20-29].. MJD and MLB used these resources to subjectively identify areas that were to be included in the critical appraisal tool. These items were discussed with RSD (Rachel S Dean, author) and a first draft of the tool (Supplementary material Table 2) and accompanying help text was created using previously published critical appraisal tools for observational and other types of study designs, and other reference documents[1 11 12 15 17 20-29]. The help text was directed at general users and was developed in order to make the tool easy to use and understandable.

The first draft of the CA tool was piloted with colleagues within the Centre for Evidence-based Veterinary Medicine (CEVM) and the population health and welfare research group at the School of Veterinary Medicine and Science (SVMS), The University of Nottingham, and the Centre for Veterinary Epidemiology and Risk Analyses in University College Dublin (UCD). Colleagues used the tool to assess different research papers of varying quality that utilised cross sectional study design methodology during journal clubs and research meetings and provided feedback on their experience. The tool was used in the analysis of cross sectional studies for a published systematic review[30]. The tool was also trialled in a journal club and percentage agreement analysis was carried out and used to develop the tool further. The CA tool was also sent via email to nine individuals experienced with systematic reviews in veterinary medicine and/or study design for informal feedback. Feedback from the different groups was assessed and any changes to the CA tool were made accordingly. The analysis identified components that were to be included in a second draft of the CA tool of cross sectional studies (Supplementary material Table 3) which was used in the first round of the Delphi process.

The Delphi panel:

The purpose of the Delphi panel was to reach consensus on what components should be present in the critical appraisal tool and aid the development of the help text. Participants for the Delphi panel were sought from the fields of evidence-based medicine (EBM), evidence-based veterinary medicine (EVM), epidemiology, nursing and public health and

were required to be involved in university education in order to qualify for selection. Personal contacts of the authors and well known academics in the EBM/EVM fields were used as the initial contacts and potential members of the panel. Email was used to contact potential participants for enrolment in the Delphi study. These potential participants were also asked to provide additional recommendations for other potential participants. All potential participants were contacted a second time if no response was received from the first email; if no response was received after the second email the potential participant was not included any further in the study.

Participants were included if:

- they held a postgraduate qualification (e.g. PhD, MSc, European College Diploma in Veterinary Epidemiology and Public Health);
- they were recognised through publication and or key note presentations for their work in evidence-based medicine and veterinary medicine, epidemiology, or public health;
- taught at university level; and
- had authored in systematic reviews (in medicine or veterinary medicine), reporting guidelines or critical appraisal.

The Delphi process:

Prior to conducting the Delphi process it was agreed that consensus for inclusion of each component in the tool would be set at 80%[31 32]. This meant that the Delphi process would continue until at least 80% of the panel agreed a component should be included in the final tool. Only if a component met the consensus criteria would it be included in the final tool, the steering committee did not change any component once it reached consensus or add any component that did not go through the Delphi panel. In each round, if a component had 80% consensus, it remained in the tool. If consensus was lower than 80% but greater than 50%, the component was considered for modification or was integrated into other components that were deemed to require reassessment for the next round of the Delphi. If consensus was ≤ 50%, components were removed from the tool.

The second draft (developed in phase one described above) of the critical appraisal tool (Supplementary material Table 3) was circulated in the first round of the Delphi process to the panel using an online questionnaire (SurveyGizmo©). Participants were asked: if each component of the tool should be included or not; if any component required alteration or clarification; or if a further component should be added. Participants were asked to add any additional comments they had regarding each component. A hyperlink to the online questionnaire with the tool was distributed to the panel using email. Participants were given four weeks to complete their assessment of the tool using the questionnaire. Participants were reminded about the work required after one week, and again three days before the Delphi round was due to close. If participants failed to respond to a specific round they were still included in the following rounds of the Delphi process. The process was repeated, with a new draft of the critical appraisal tool circulated each time based on the findings and consensus of the previous round, until 80% consensus on all components of the tool was achieved.

On the third round of the Delphi process, a draft of the help text for the tool was also included in the questionnaire and consensus was sought as to whether the tool was suitable for the non-expert user, and participants were asked to comment on the text. The responses were compiled and analysed at the end of round 3. Consensus was sought for the suitability of the help text for the non-expert user and set at 80%. However if consensus was lower than 80% but greater than 50% the help text was considered for modification. If comments were given on the help text, these comments were integrated into the help text of the tool.

Ethical approval

The ethics committee at the School of Veterinary Medicine and Science, The University of Nottingham reviewed and approved this study. Written consent (via email) from the panellists was received by the authors following invitation to be included in the study.

RESULTS

The initial review of existing tools and texts identified 34 components that were deemed relevant for critical appraisal of cross sectional studies and were included in the 1st draft of the tool (Supplementary material Table 2). When piloted there was an overall percent agreement of 88.9%; however, 32.9% of the questions were unanswered. Post-feedback modification after the pilot study identified 37 components to be included in 2nd draft of the CA tool (Supplementary material Table 3).

Twenty seven potential participants were contacted for the Delphi study. Eighteen experts (67%) agreed to participate in the Delphi panel. The most common reasons for not partaking was not enough time (n=5); of these, four were lecturers with research and clinical duties and one was a lecturer with research duties. Two contacts felt they were not suitably qualified for the Delphi panel (n=2); one was retired and the other was a lecturer with research and clinical duties. Two contacts did not respond to the emails; these were both lecturers with research duties. Of those that took part eight were involved in clinical, teaching and research duties and ten were involved in research and teaching, five of the participants were veterinary surgeons and six were medical clinicians. It was an international panel including ten participants from the United Kingdom, three from Australia, two from the United States of America, two from Canada, and one from Egypt. Participants were qualified a mean of 17.6 years (SD: 7.9) and the panel was made up of participants from varying disciplines (Table 1)

Table 1. The number of participants from each discipline enrolled in the Delphi panel for the development of the AXIS tool

Discipline	**N**
Epidemiology	4
Evidence-based medicine	9
Evidence-based veterinary medicine	2
Public Health	3

During round one (undertaken in February 2013) of the Delphi process, 20 components reached consensus, 13 components were assessed to require modification and it was deemed appropriate to remove four components from the tool. General comments mostly related to the tool having too many components.

> *" the tool needs to be succinct and easy and quick to use if possible - too many questions could have an impact."*
>
> *"List is too long at present and contains too many things that are general to all scientific studies"*
>
> Comments voiced included the discussion as part of the critical appraisal process being unnecessary and potentially misleading:
>
> *"the interpretation should, in my opinion, come from the methods and the results and not from what the author thinks it means"*
>
> *"I don't believe a Discussion section should be part of a critical appraisal."*

Therefore, in round one the tool was modified in an attempt to reduce its size and to encompass all comments. For round two (undertaken in May 2013) 11 components remained the same and did not require testing for consensus as this was established in round one; nine components that had previously reached consensus were incorporated with the 13 components that required modification to create 10 new components (Supplementary material Table 4).

In round two, consensus was reached on a further two components, six components were assessed to require modification and it was deemed appropriate to remove two components from the tool. Comments from the panel regarding the components of the tool that related to the discussion, suggested further reduction in these components due to their limited use as part of the critical appraisal process.

> *"the discussion could legitimately be highly speculative and not justified by the results provided that the authors don't present this as conclusions."*

With the reduction in the number of questions and modification of the wording, comments in round 2 reflected the positive nature to the usability of the tool.

> *"I like the fact that it is quite simple - not too overloaded with methodological questions."*

After round 2, the tool was further reduced in size and modified to create a 4th draft of the tool with 20 components incorporating 13 components with full consensus and seven modified components for circulation in round 3 of the Delphi process.

Following round 3 (undertaken in July 2013) of the Delphi process, there was consensus (81%) that all components of the tool were appropriate for use by non-expert users so no further rounds were necessary. The final critical appraisal tool for cross sectional studies (AXIS tool) consisting of 20 components is shown in Table 2. The comments from the panel regarding the help text were addressed and minor modifications to the text were made (Supplementary material 4). Seven (1, 4, 10, 11, 12, 16 and 18) of the final questions related to quality of reporting, seven (2, 3, 5, 8, 17, 19 and 20) of the questions related to study design quality, and six related to the possible introduction of biases in the study (6, 7, 9, 13, 14 and 15).

Table 2. The final AXIS tool following consensus on all components by the Delphi panel.

		Yes	No	Don't know/ Comment
Introduction				
1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			

13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results for the analyses described in the methods, presented?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

DISCUSSION

The AXIS Tool

A critical appraisal tool to assess the quality and risk of bias in cross-sectional studies (AXIS), along with supporting help text, was successfully developed by an expert panel using Delphi methodology. This is the first critical appraisal tool made available for assessing this type of evidence that can be incorporated in systematic reviews, guidelines and clinical decision-making.

Key Delphi Findings

One of the key items raised in comments from the experts was assessing quality of design versus quality of reporting. It is important to note that a well reported study may be of poor quality and conversely a poorly reported study could be a well conducted study[33 34]. It is also apparent that if a study is poorly reported it can be difficult to assess the quality of the study. Some information may be lacking due to poor reporting in studies making it difficult to assess the risk of biases and the quality of the study design. High quality and complete reporting of studies is a pre-requisite for judging quality[17 18 35]. For this reason, the AXIS tool incorporates some quality of reporting as well as quality of design and risk of biases to overcome these problems.

Using the tool

The tool was also reduced in size on each round of the Delphi process as commentators raised concerns around developing a tool with too many questions. The comments suggested that a long questionnaire would lead to the tool being cumbersome and difficult to use, and for this reason efforts were made to develop a much more concise tool.

The AXIS tool focuses mainly on the presented methods and results. It was the view of the Delphi group that the assessment as to whether the published findings of a study are credible and reliable should relate to the aims, methods and analysis of what is reported and not on the interpretation (e.g. discussion and conclusion) of the study. This view is also seen in other appraisal tools, is shared by other researchers and can be seen by the absence of questions relating to the discussion sections in critical appraisal tools for other types of studies[12 16 20 28 36].

As with all critical appraisal tools it is only possible for the reader to be able to critique what is reported. If an important aspect of a study is not in the manuscript, it is unclear to the reader whether it was done, and not reported, or not done at all. It is therefore the responsibility of the appraiser of the study to recognise omissions in reporting and consider how this affects the reliability of the results.

A comprehensive explanatory text is often used in appraisal tools for different types of study designs as it aids the reviewer when interpreting and analysing the outputs from the appraisal [12 17-20]. This approach was also used in the development of the AXIS tool where a reviewer can link each question to explanatory text to aid in answering and interpreting the questions.

Study strengths and limitations

The tool was developed through a rigorous process incorporating comprehensive review, testing and consultation via a Delphi panel. Using a similar process to other appraisal tools[37], we reviewed the relevant literature to develop a concise background on critical appraisal of cross sectional studies and to ensure no other relevant tools existed. While numerous tools exist for critical appraisal, we found a lack of tools for general use in cross sectional studies and this was consistent with what others have found previously[12 13]. In order to ensure quality and completeness of the tool we utilized recognised reporting guidelines, other appraisal tools and epidemiology design text in the development of the

initial tool which is similar to the development of appraisal tools of other types of studies[12].

The use of a multidisciplinary panel with experience in epidemiology and evidence-based medicine limits the effect of using a non-representative sample, and the use of the Delphi tool is well recognised for developing consensus in health care science[38]. The selection of a Delphi group is very important as it effects the results of the process[31]. As cross-sectional studies are used extensively in both human and veterinary research it was appropriate to use expertise from both of these fields. To ensure that the tool was developed to a high standard, a high level of consensus was required in order for the questions to be retained[31 32 39]. There was a high level of consensus between both veterinary and medical groups in this study, which adds to the rigour of the tool but also demonstrates how both healthcare areas can co-operate effectively to produce excellent outcomes.

The Delphi study was conducted using a carefully selected sample of experts and as such may not be representative of all possible users of the tool. However the purpose of a Delphi study is to purposely hand pick participants that have prior expertise in the area of interest[40]. The Delphi members came from a multidisciplinary network of professionals from medicine, nursing and veterinary medicine with experience in epidemiology and EBM/EVM and exposure to teaching and areas of EBM that were not just focused on systematic reviews of randomised controlled trials. The panel was restricted to those that were literate in the English language and may therefore not be representative of all nationalities. The interests and experiences of the panel will clearly have had an effect on the results of this study as this is common to all Delphi studies[31 41]. The majority of Delphi studies are conducted using between 15 and 20 participants[31], so a panel of 18 is consistent with other published Delphi panels. We aimed to recruit a minimum of 15 participants and as it was anticipated that not all participants contacted would be able to take part, more participants were contacted.

As the tool does not provide a numerical scale for assessing the quality of the study, a degree of subjective assessment is required. This has implications for interpretation after using the tool as there will be differences in individuals' judgements. However, it has been debated that quality numerical scales can be problematic as the outputs from assessment

checklists are not linear and as such are difficult to sum up or weight making them unpredictable at assessing study quality[39 42 43]. The AXIS tool has the benefit of providing the user the opportunity to assess each individual aspect of study design to give an overall assessment of the quality of the study. By providing this subjectivity, AXIS gives the user more flexibility in incorporating both quality of reporting and risk of bias when making judgments on the quality of a paper. This tool therefore provides an advantage over, Berra et al (2008) [15] which only allows the user to assess quality of reporting and tools such as the Cochrane risk of bias tool [5] which do not address poor reporting. Further studies would be needed to assess how practical this tool is when used by clinicians and if the critical appraisal of studies using AXIS is repeatable.

CONCLUSION

In conclusion, a unique tool (AXIS) for the critical appraisal of cross sectional studies was developed that can be used across disciplines e.g. health research groups and clinicians conducting systematic reviews, developing guidelines, undertaking journal clubs and private personal study. The components of the AXIS tool are based on a combination of evidence, epidemiological processes, experience of the researchers and Delphi participants. As with other evidence-based initiatives, the AXIS tool is intended to be an organic item that can change and be improved where required, with the validity of the tool to be measured and continuously assessed. We would invite any users of the tool to provide feedback so that the tool can be further developed if needed and can incorporate user experience to provide better usability.

ACKNOWLEDGMENTS:

The authors thank the following individuals who participated in the Delphi process: Peter Tugwell, Thomas McGinn, Kim Thomas, Mark Petticrew, Fiona Bath-Hextall, Amanda Burls, Sharon Mickan, Kevin Mackway Jones, Aiden Foster, Ian Lean, Simon More, Annette O'Connor, Jan Sargeant, Hannah Jones, Ahmed Elkhadem, Julian Higgins and Sinead Langan.

The authors would like thank those who piloted the tool in the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in

evidence-based veterinary medicine. The authors would also like thank Michelle Downes for designing the population diagram.

The Centre for Evidence-based Veterinary Medicine is supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Contributorship statement

MJD, MLB and RSD made substantial contribution to the conception and design of the work, as well as the acquisition, analysis, or interpretation of data. They were directly involved in the preparation and revision of the intellectual content of the manuscript submitted for publication. They will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

HCW made substantial contribution to the conception and design of the work, interpretation of data and the preparation and revision of the intellectual content of the manuscript submitted for publication. He will be involved in the final approval of the manuscript prior to publication and agrees to be accountable for all aspect of the work.

Competing interests

The authors have no competing interests.

Funding

The Centre for Evidence-based Veterinary Medicine and this project was supported by an unrestrictive grant from Elanco Animal Health and The University of Nottingham.

Data sharing statement

The authors are willing to share any of the data collected during the Delphi process that is not in this published manuscript or the supplementary material.

REFERENCES

1. Cockcroft PD HM. *Handbook of Evidence-based Veterinary Medicine.* . 1st ed. Oxford, UK: Blackwell Publishing Ltd, 2003`.
2. McColl A, Smith H, White P, et al. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. Bmj 1998;**316**(7128):361-5

3. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. Bmj 1995;**310**(6987):1122-6
4. Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. Bmj 2001;**323**(7317):833-6
5. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*: Wiley Online Library, 2008.
6. Bennett KJ, Sackett DL, Haynes RB, et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. Jama 1987;**257**(18):2451-4
7. Mogford E, Gould L, Devoght A. Teaching critical health literacy in the US as a means to action on the social determinants of health. Health Promot Int 2011;**26**(1):4-13 doi: DOI 10.1093/heapro/daq049[published Online First: Epub Date]|.
8. Parkes J, Hyde C, Deeks J, et al. Teaching critical appraisal skills in health care settings. The Cochrane database of systematic reviews 2001(3):CD001270 doi: 10.1002/14651858.CD001270[published Online First: Epub Date]|.
9. Sackett DL. Teaching critical appraisal. Journal of general internal medicine 1990;**5**(3):272
10. Ho PM, Peterson PN, Masoudi FA. Evaluating the evidence: is there a rigid hierarchy? Circulation 2008;**118**(16):1675-84 doi: 10.1161/CIRCULATIONAHA.107.721357[published Online First: Epub Date]|.
11. Dohoo IR, Martin SW, Stryhn H. *Veterinary Epidemiologic Research*. Charlottetown, Canada: Ver Books, 2009.
12. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC medical research methodology 2004;**4**:22 doi: 10.1186/1471-2288-4-22[published Online First: Epub Date]|.
13. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International journal of epidemiology 2007;**36**(3):666-76 doi: 10.1093/ije/dym018[published Online First: Epub Date]|.
14. Sterne JAC HJ, Reeves BC on behalf of the development group for ROBINS-I. ROBINS-I: a tool for assessing Risk Of Bias In Non-randomized Studies of Interventions, Version 5. 2016
15. Berra S, Elorza-Ricart JM, Estrada M-D, et al. [A tool (corrected) for the critical appraisal of epidemiological cross-sectional studies]. Gac. Sanit. 2008;**22**(5):492-7
16. Chalmers TC, Smith H, Jr., Blackburn B, et al. A method for assessing the quality of a randomized control trial. Controlled clinical trials 1981;**2**(1):31-49
17. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. B World Health Organ 2007;**85**(11):867-72 doi: Doi 10.2471/Blt.07.045120[published Online First: Epub Date]|.
18. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999;**354**(9193):1896-900 doi: Doi 10.1016/S0140-6736(99)04149-5[published Online First: Epub Date]|.
19. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;**63**(8):834-40 doi: DOI 10.1016/j.jclinepi.2010.02.005[published Online First: Epub Date]|.
20. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology 2007;**7** doi: Artn 10

Doi 10.1186/1471-2288-7-10[published Online First: Epub Date]|.
21. BestBETs. BETs CA Worksheets. Secondary BETs CA Worksheets 2009. http://bestbets.org/links/BET-CA-worksheets.php.
22. Crombie I. *The Pocket Guide to Critical Appraisal. annotated edition. Wiley-Blackwell.* 1 ed. London, UK: BMJ Publishing group, 1996.

23. Fowkes FGR, Fulton PM. Critical-Appraisal of Published Research - Introductory Guidelines. Brit Med J 1991;**302**(6785):1136-40
24. Genaidy AM, Lemasters GK, Lockey J, et al. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. Ergonomics 2007;**50**(6):920-60 doi: 10.1080/00140130701237667[published Online First: Epub Date]|.
25. Hanke BA, Motschall E, Turp JC. Association between orthopedic and dental findings: What level of evidence is available? J. Orofac. Orthop. 2007;**68**(2):91-107 doi: 10.1007/s00056-007-0634-0[published Online First: Epub Date]|.
26. Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional, 2004.
27. Rothman K. *Chapter 5: Bias in Study Design. Epidemiology: An Introduction*. New York: Oxford Univeristy Press, 2002.
28. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerging themes in epidemiology 2008;**5**:23 doi: 10.1186/1742-7622-5-23[published Online First: Epub Date]|.
29. Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual 2014 Edition*, 2014.
30. Downes MJ, Dean RS, Stavisky JH, et al. Methods used to estimate the size of the owned cat and dog population: a systematic review. BMC veterinary research 2013;**9**:121 doi: 10.1186/1746-6148-9-121[published Online First: Epub Date]|.
31. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Practical assessment, research & evaluation 2007;**12**(10):1-8
32. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 2004;**42**(1):15-29 doi: http://dx.doi.org/10.1016/j.im.2003.11.002[published Online First: Epub Date]|.
33. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. Bmj 2004;**328**(7430):39-41 doi: 10.1136/bmj.328.7430.39[published Online First: Epub Date]|.
34. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj 2001;**323**(7303):42-6
35. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama 2000;**283**(15):2008-12
36. Wells G, Shea B, O'connell D, Peterson J, Welch V, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Secondary The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000. http://www.medicine.mcgill.ca/rtamblyn/Readings/The%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf.
37. Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011;**64**(1):79-89 doi: DOI 10.1016/j.jclinepi.2010.02.008[published Online First: Epub Date]|.
38. Fink A, Kosecoff J, Chassin M, et al. Consensus Methods - Characteristics and Guidelines for Use. Am J Public Health 1984;**74**(9):979-83 doi: Doi 10.2105/Ajph.74.9.979[published Online First: Epub Date]|.
39. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostat 2001;**2**:463-71 doi: 10.1093/biostatistics/2.4.463[published Online First: Epub Date]|.
40. Mckenna HP. The Delphi Technique - a Worthwhile Research Approach for Nursing. J Adv Nurs 1994;**19**(6):1221-25 doi: DOI 10.1111/j.1365-2648.1994.tb01207.x[published Online First: Epub Date]|.

41. Linstone HA, Turoff M. *The Delphi method: Techniques and applications*: Addison-Wesley Reading, MA, 1975.
42. Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Brit Med J 2011;**343** doi: Artn D5928

Doi 10.1136/Bmj.D5928[published Online First: Epub Date]|.
43. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. Jama-J Am Med Assoc 1999;**282**(11):1054-60 doi: DOI 10.1001/jama.282.11.1054[published Online First: Epub Date]|.

Supplementary material

Table 1: Search Terms Used to identify critical appraisal tools for cross sectional studies

Database	Number	
Medline 1948 to September Week 3 2011	71	Critical appraisal.mp. AND (exp Cross-Sectional Studies/ OR cross sectional.mp.)
CAB 1910 to 2011 Week 38	4	Critical appraisal.mp. AND cross sectional.mp.
Web of Science^{SM} (1899-present)	60	Topic=(Critical appraisal) AND Topic=(cross sectional)
BIOSIS Previews® (1969-present)	12	Topic=(Critical appraisal) AND Topic=(cross sectional)
Zoological Record® (1978-present)	0	Topic=(Critical appraisal) AND Topic=(cross sectional)
Embase 1974 to 2011 October 03	65	Critical appraisal.mp. AND (exp cross-sectional study/ OR cross sectional.mp.)
CINAHL® with Full Text	23	((MM "Cross Sectional Studies") OR "cross sectional") AND "Critical appraisal"
PsycINFO 1806 to September Week 4 2011	9	Critical appraisal.mp. AND cross sectional.mp.
Total	244	

Table 2. The 1st draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post review of the literature.

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Are the aims of the study clearly stated?			
Methods				
2	Is the type of study design appropriate for the stated aim?			
3	Is the sample size justified (based on pre-study considerations of statistical power)?			
4	Is the target or reference population clearly defined? (is it clear who the research was about)			
5	Is the sample frame taken from an appropriate population base so that it closely represents the overall population under investigation?			
6	In the selection process:			
	a. Were any inclusion/exclusion criteria used?			
	b. Was random selection used to obtain participants?			

7	Is the selection process likely to select subjects that were representative of the study population of interest?			
8	If appropriate, were measures undertaken to address and categorise non-responders?			
9	Do the variables measured, in the study, produce data that reflect the aims of the study? (Validity)			
	a. Are the outcomes of interest clearly measured?			
	b. Are the risk factors appropriately measured to be compared to the outcomes of interest?			
10	If appropriate, have the measurement instruments been trialled, piloted or published previously? (Reliability and reproducibility)			
11	Are the statistical methods clearly stated?			
12	If appropriate, is the means by which statistical significance is inferred stated?			
13	Are the methods sufficiently described to enable them to be repeated?			
Results				
14	Are the basic data adequately described?			
15	Is the response rate given, if appropriate?			
16	Is information about non-responders described, if appropriate?			
17	Are the results internally consistent?			
	a. Do the numbers add up?			
	b. Are any missing data acknowledged, or described?			
18	Are the results described objectively without author opinion?			
19	Are results pertaining to the study aim reported?			
20	If appropriate is the statistical significance level declared in the methods adhered to?			
21	Are the results of all tests described in the methods presented?			
Discussion				
22	Are all results pertaining to the study aim discussed?			
23	Are the limitations of the study discussed?			
24	Is selection bias addressed?			
25	Is non-response addressed?			
26	Do the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?			
27	If appropriate are non-significant results discussed?			
	a. Do the authors consider issues around study design when interpreting non-significant results?			
	b. Do the authors consider issues around sample size when interpreting non-significant results?			
Conclusions				
28	Are the authors' conclusions justified by the results?			
Other				
29	Are any conflicts of interest/funding declared in the text?			
30	Was ethical aspect approval or consent of participants attained?			

Table 3. The 2nd draft of a CA tool including components that were identified as relevant to critical appraisal of cross sectional studies post piloting with the Centre for Evidence-based Veterinary Medicine (UoN), the Population Health and Welfare group (UoN), the Centre for Veterinary Epidemiology and Risk Analyses (UCD) and the online forum of experts in evidence based veterinary medicine. This draft was used in the first round of the Delphi panel and the results of the consensus from the panel or each component are presented.

	Consensus
Introduction	
1. Is it clear what the aims of the study were?	94.12
Methods	
2. Was the type of study design appropriate for the stated aim?	94.12
3. Was the sample size justified (based on pre-study considerations of statistical power)?	76.47
4. Was the target or reference population clearly defined? (is it clear who the research was about?)	100.00
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	88.24
6.a. In the selection process: Were any inclusion/exclusion criteria used?	82.35
6.b. In the selection process: Was random selection used to obtain participants?	70.59
7. Was the selection process likely to select subjects that were representative of the study population of interest?	93.33
8. If appropriate, were measures undertaken to address and categorise non-responders?	87.50
9. Did the variables measured in the study, produce data that reflected the aims of the study? (Validity)	68.75
9.a. Were the outcomes of interest clearly measured?	86.67
9.b. Were the risk factors measured appropriate to the outcomes of interest?	66.67
10. If appropriate, had the measurement instruments been trialled, piloted or published previously?	93.33
11. Is it clear what statistical methods were used?	86.67
12. If appropriate is it possible to determine the means by which the statistical significance was inferred? (p-values, confidence intervals)	64.29
13. Were the methods sufficiently described to enable them to be repeated?	87.50
Results	
14. Were the basic data adequately described?	75.00
15. If appropriate, was the response rate sufficient to base conclusions on?	93.75
16. If appropriate, was information about non-responders described?	93.75
17. Were the results internally consistent?	57.14
17.a. Did the numbers add up?	56.25
17.b. Were any missing data acknowledged, or described, if appropriate?	93.75

18. Were the results described objectively without author opinion?	43.75
19. Were the results pertaining to the study aim reported?	81.25
20. If appropriate, was the statistical significance level declared in the methods adhered to?	68.75
21. Were the results of all tests described in the methods presented?	93.75
22. Were all results pertaining to the study aim discussed?	56.25
Discussion	
23. Were the limitations of the study discussed?	81.25
24. Was selection bias discussed appropriately?	62.50
25. Was non-response discussed appropriately?	68.75
26. Did the authors address any relevant reasons for their findings, other than the tested hypothesis (Confounding)?	53.33
27. If appropriate, were non-significant results discussed?	50.00
27.a. Did the authors consider issues around study design when interpreting non-significant results?	50.00
27.b. Did the authors consider issues around sample size when interpreting non-significant results?	50.00
28. Were the authors' conclusions justified by the results?	93.75
Other	
29. Were any conflicts of interest/funding declared in the text?	93.75
30. Was ethical approval or consent of participants attained?	81.25

Table 4. The 3rd draft of a CA tool created following round 1 of the Delphi study after comments and consensus was taken into account. Results on consensus for each question from the round 2 of a Delphi panel are presented.

	Consensus*
Introduction	
1. Is it clear what the aims of the study were?	
Methods	
2. Was the type of study design appropriate for the stated aim?	
3. If appropriate, was the sample size justified?	68.75
4. Was the target or reference population clearly defined? (Is it clear who the research was about?)	
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	
6. Was the selection process likely to select subjects that were representative of the study population of interest?	
7. Were measures undertaken to address and categorise non-responders?	81.25
8. Would the variables measured in the study produce data that reflected the aims of the study? (Validity)	62.5
9. Is it clear what statistical methods were used?	50
10. Is it clear how statistical significance was determined? (eg: p-values, confidence intervals)	62.5
11. Were the methods sufficiently described to enable them to be repeated?	
Results	
12. Were the basic data adequately described?	
13. If appropriate, was the response rate sufficient to enable sound conclusions to be drawn?	56.25
14. If appropriate, was information about non-responders described?	
15. Were the results internally consistent?	
16. Were all the results of the analyses described in the methods presented?	60
17. If a statistical significance level was declared in the methods, was it adhered to in the results?	31.25
Discussion	
18. Were the authors' discussions and conclusions justified by the results?	87.5
19. Were the limitations of the study discussed?	
Other	
20. Were there any funding sources or conflicts of interest that were likely to affect the authors' interpretation of the results?	75
21. Was ethical approval or consent of participants attained?	

**Where no consensus figure is given, consensus was reached on this question in the previous round.*

Appraisal tool for Cross-Sectional Studies (AXIS)

Critical appraisal (CA) is used to systematically assess research papers and to judge the reliability of the study being presented in the paper. CA also helps in assessing the worth and relevance of the study [1]. There are many key areas to CA including assessing suitability of the study to answer the hypothesised question and the possibility of introducing bias into the study. Identifying these key areas in CA requires good reporting of the study, if the study is poorly reported the appraisal of suitability and bias becomes difficult.

The following appraisal tool was developed for use in appraising observational cross-sectional studies. It is designed to address issues that are often apparent in cross-sectional studies and to aid the reader when assessing the quality of the study that they are appraising. The questions on the following pages are presented in the order that they should generally appear in a paper. The aim of the tool is to aid systematic interpretation of a cross-sectional study and to inform decisions about the quality of the study being appraised.

The appraisal tool comes with an explanatory help text which gives some background knowledge and explanation as to what the questions are asking. The explanations are designed to inform why the questions are important. Clicking on a question will automatically take you to the relevant section in the help text. The appraisal tool has areas to record a “yes”, “no” or “don’t know” answer for each question and there is room for short comments as well.

Contents

Appraisal of Cross-sectional Studies

	Question	Yes	No	Don't know/ Comment
Introduction				
1	Were the aims/objectives of the study clear?			
Methods				
2	Was the study design appropriate for the stated aim(s)?			
3	Was the sample size justified?			
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)			
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?			
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?			
7	Were measures undertaken to address and categorise non-responders?			
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?			
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?			
10	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)			
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?			
Results				
12	Were the basic data adequately described?			
13	Does the response rate raise concerns about non-response bias?			
14	If appropriate, was information about non-responders described?			
15	Were the results internally consistent?			
16	Were the results presented for all the analyses described in the methods?			
Discussion				
17	Were the authors' discussions and conclusions justified by the results?			
18	Were the limitations of the study discussed?			
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?			
20	Was ethical approval or consent of participants attained?			

Introduction

The introduction serves to establish the context of the work that is about to be presented in the text of the paper. Relevant primary literature should be discussed and referenced throughout the introduction. The history and current understanding of the problem being researched should be presented. This should be concluded giving a rational as to why the current study is being presented and what the aims and/or hypothesis under investigated are [2,3].

Aims

The aim(s) of the study tells us if the study addresses an appropriate and clearly focused question. If the aim is not clearly stated or not stated at all, it will be difficult and in some cases impossible to assess the extent to which the study objectives were achieved. Ideally, an aim should be stated both at the beginning of the abstract and at the end of the introduction [3]. If the answer to question 1 is no, then it will make it difficult to assess some of the other questions in the critical appraisal process.

Methods

The methods section is used to present the experimental study design of the paper. The methods should be described clearly in easy to understand language and clearly identify measures, exposures and outcomes being used in the study [4]. More specific issues are addressed below.

Study Design

Question 2 is used to assess the appropriateness of using a cross-sectional study to achieve the aim(s) of the study. Cross-sectional studies are observational studies that provide a description of a population at a given time, and are useful in assessing prevalence and for testing for associations and differences between groups [5]. Examples of cross-sectional designs include point-in-time surveys, analysis of records and audits of practice [6]. The reader should try and decipher if a cross-sectional study design is appropriate for the questions being asked by the researcher.

Sample Size Justification

Sample size justification is crucial as sample size profoundly affects the significance of the outcomes of the study. If the sample size is too small then the conclusions drawn from the study will be under powered and may be inaccurate. This can occur by failing to detect an effect which truly exists (type II error) sometimes referred to as a "false negative". The probability of a type I error is also taken into account when determining sample size. A type I error is drawing significant conclusions when no real difference exists and is a function of the p-value (see Statistics section below) sometimes referred to as a "false positive".

Question 3 asks if sample size justification was reported, but it should also be clear what methods were used to determine the sample size. In some cases clustering of observations within groups can occur (e.g. patients within hospitals or livestock within herds) and this should be taken into account if sample size has been determined. It should be clear whether the inferences drawn actually relate to the attributes for which the sample size was calculated [7]. If sample size justification isn't given or restrictions make it difficult to reach the desired sample size then this should be declared in the text.

Target (Reference) Population

The target or reference population is the overall population that the research is directed towards. When doing a cross-sectional study, a target population is the overall population you are undertaking the study to make conclusions about or the population at risk of acquiring the condition being investigated [8–10] e.g. the total female population in the UK, or all dogs in the USA with cardiovascular disease. (See Figure 1) Question 4 asks if this is clearly defined in the study. It is important that this is understood both by the researcher and the reader; if it is not clearly defined then inferences made by the researcher may be inappropriate.

Sampling Frame

As a reader you need to determine if the sample frame being used is representative of the target population. The study population should be taken from the target population; units from this study population have information that is accessible and available which allows them to be placed in the study. The sampling frame is the list or source of the study population that the researcher has used when trying to recruit participants into the study (Figure 1). Ideally it should be exactly the same composition or structure as the target population. In practice it is generally much smaller, but should still be representative of the target population. Generally, for convenience, the sampling frame is a list of units that are within the target population e.g. list of

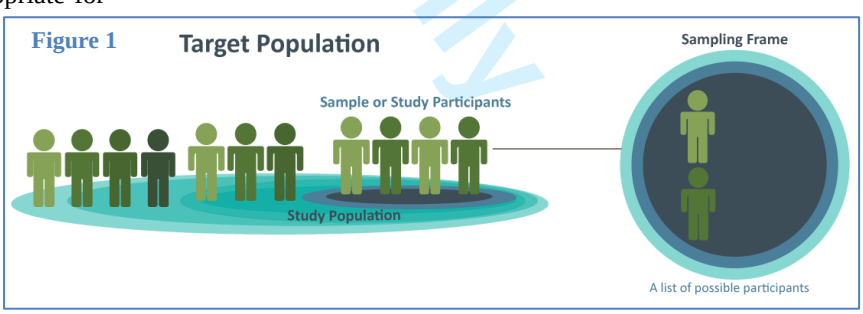

Figure 1

telephone owning households, computerised patient records etc. A sample of units is selected from the study population to take part in the study and is generally only a small proportion of the study population (see Sample Selection below) - this proportion ratio is known as the sampling fraction. It is very important that the sampling frame is representative of the target population as results from the study are going to be used to make assumptions about the target population [8–10].

Convenience sampling can be carried out in some situations and are used because the participants are easy to recruit. Convenience samples generally lead to non-representative or biased samples and therefore cannot be used to make assumptions about the characteristics of the target population [11]. Convenience samples are often used for pilot or analytical studies where the need for a representative sample is not required [12], however the authors should make this clear in the text.

Census

A census is where the target population and the study participants are the same at the time the census is taken. In theory questions 5, 6 and 7 don't apply to census studies. However even if a study is described as a census it should be very clearly stated where the study participants have been recruited from, and the reader should make the decision if the study truly is a census. A census may include all the population from the sample frame, but not all the target population; in this scenario questions 5 to 7 need to be addressed.

Sample Selection

Question 6 is used to establish how the researchers got from the sample frame to the participants in the study. It examines the potential for selection bias and how the researcher developed methods to deal with this. The sample selection process is important in determining to what extent the results of the study are generalizable to the target population. For question 6 we are looking in depth at how the sample (study participants) was selected from the sampling frame. It is important to know if there were any inclusion or exclusion criteria used, as inappropriate criteria can dramatically shift how representative the sample is of the target population [8,10,13].

Selection bias can occur if every unit in the sample frame doesn't have an equal chance of been included in the final study [11,14]. Randomisation is used to ensure that each participant in the sampling frame has an equal chance of being included in the sample. If methods of randomisation are not used, not described or are not truly random, this may lead to a non-representative sample being selected and hence affect the results of the study [10,11].

There are many other situational issues to take into account when determining if the population in the sample is likely to represent the target population. Often these issues are outside the control of the researcher, but sometimes are overlooked. One such issue is the healthy worker effect which is a well-known phenomenon in human cross-sectional studies [13]. An example of this is, a researcher trying to do a cross-sectional study to determine health factors in a factory population and decides to sample from workers at work on a particular day. Unfortunately there is a tendency to over select healthy workers as ill workers may tend to be at home on the day of selection. This will in turn lead to inferences been made about the health of the worker population but is only relevant to healthy workers and not ill workers. A veterinary example of this is a researcher trying to do a cross-sectional study to determine health factors in the general dog population and decides to sample from a local park. Unfortunately there is a tendency to over select healthy animals as sick animals will tend to be left at home and not taken for a walk. This will in turn lead to inference been made about the health of the dog population but is only relevant to healthy dogs and not sick dogs.

Self-selection is another example of selection bias that can be introduced and should be assessed [13]. For example, when using a postal questionnaire to examine eating habits and weight control, people who are overweight might read the survey and be less inclined to complete and return the survey than those with normal weight leading to over representation of people with normal weight. Similarly, if using a postal questionnaire to examine mastitis levels on cattle farms, farmers that have a high somatic cell counts (SCC) might be less inclined to complete the survey than those with normal or low SCC leading to over representation of farms with good SCC (see Non-responders below).

Non-responders

Non-response in cross-sectional studies is a difficult area to address. A non-responder is someone who does not respond either because they refuse to, cannot be contacted, or because their details cannot be documented. As a rule, if participants don't respond it is often difficult and sometimes impossible to gain any information about them. However other baseline statistics may exist that can be used as a comparator to assess how representative the sample is [14] e.g. age, sex, socio-economic classification. Methods used, if any, should be well described so that the results from the analyses can be interpreted. This is important as non-responders may be from a specific group, which can lead to a shift in the baseline data away from that group. This shift can lead to results that don't represent the target population. In some situations the sampling frame doesn't have a finite list or a fully defined baseline population. This also makes it difficult, and in some cases impossible, to quantify non-response and it may be inappropriate to do so in these situations. If the researchers are using non-defined populations this should also be declared clearly in the materials and methods section [15,16].

Measurement Validity & Reliability

Measurement validity is a gauge of how accurately the study measurements used assess the concepts that the researcher is attempting to explore. Measurement reliability is a gauge of the accuracy of the measurements taken or the procedures used during the study. Question 8 is used to address the concepts of measurement validity, and is specifically aimed to address the appropriateness of the measurements being used.

The importance of measurement validity is that it gives weight to applying the statistical inferences from the study to members of the target population. If inappropriate measures are used in the study it could lead to misclassification bias and it will be difficult to determine to what extent the study results are relevant to the target population [12,17].

Question 9 is an attempt to gauge the measurement reliability of the study measures. Measurements must be able to be reproduced and produce identical results if measured repeatedly, so that the measurements would be exactly the same if performed by another researcher. With this in mind, the measurements must be of international or globally accepted standards (e.g. IU standards) where possible and appropriate. If they are being used for the first time they must be trialled, or in the case of questionnaires, they should be piloted before being used.

Statistics

While interpretation of statistics can be quite difficult, a basic understanding of statistics can help you to assess the quality of the paper. Often many different methods can be used correctly to test the same data, but as there is such a wide range available, knowing what tests are most appropriate in particular situations can be hard to decipher. There is an expectation that the researcher has this understanding or has at least sought statistical assistance to ensure that the correct methods are used. Therefore for question 10 the emphasis for the reader is that the statistical methods, software packages used and the statistical significance levels are clearly stated even if the paper is just presenting descriptive statistics. The statistical significance level is usually described as a p-value. In most cases the p-value, at which the null hypothesis is rejected, is set at 0.05. The higher the p-value is set the greater the possibility of introducing a type I error. Confidence intervals should also be declared with p-values or instead of p-values as an indication of the precision of the estimates. It is usual to present a confidence interval of 95% which means that the researchers were 95 per cent confident that the true population value of the outcome lies between these intervals. This can be used to compare groups where an overlap would suggest no difference and a gap between confidence intervals would suggest a difference (Figure 2).

Figure 2

Overall Methods

Question 11 asks if the methods are sufficiently described to enable them to be repeated. If there are sections or even small pieces of information missing it could make a great difference for the reader when interpreting the results and the discussion as they may be unsure if the correct methods are being used.

Results

The results section of a paper is solely for the purpose of declaring the results of the data analysis and no opinion should be stated in this section. This gives the reader the opportunity to examine the results unhindered by the opinion of the researcher. It is important for the reader to form their own ideas or opinions about the results before progressing to the discussion stages.

Basic Data

Question 12 asks for a description of the basic data. Basic descriptive analysis aims to summarise the data, giving detailed information about the sample and the measurements taken in the study. The basic data gives an overview of the process of recruitment and if the sampling methods used to recruit individuals were successful in selecting a representative sample of the target population. If the sampling methods are unsuccessful in selecting a representative sample of the target population, those participants included in the study can often be different to the target population; this leads to inaccurate estimates of prevalence, incidence or risk factors for disease. Descriptive data of the measurements taken in the study give an overview of any differences between the groups, and may give insight into some of the reasons for statistical inferences that are made later in the paper.

Response Rate

As stated previously it can often be difficult to deal with non-responders. Question 13 requires that there is some attempt made to quantify the level of non-response by the researchers and asks the reader to interpret if the response rate is likely to lead to non-response bias. Question 14 is examining if any information on non-responders was available and if so were they comparable to those that did respond as this could help in answering question 13. Non-response bias occurs if the non-responders are substantially different to the rest of the population in the sample [15].

Internally Consistent Results

Question 15 is an exploration of the basic data and asks that the reader spends some time exploring the numbers given in the results; in the text, figures and tables. Information about the level of missing data should also be declared in the results. It is important to check that the numbers add up in the tables and the text. If the study has recruited 100 participants, the tables and the text should include data about 100 participants. If not, the missing data should be clearly declared and the reason for its non-appearance explained.

Comprehensive Description of Results

It is important to check that all the methods described previously lead to data in the results section (question 16). Sometimes the results from all analyses are not described. If this is noted it will be unclear whether the researcher found non-significant results or just didn't describe what was found. If there are results missing that you would expect to find, there is a concern that these missing results may not have been what the researcher wanted to see and hence the authors have omitted them. It is also important that the significance level declared in the methods is adhered to. As the reader, it is important to watch out for phrases such as "tended towards significance" in the text, and if these are used to pay close attention to the results.

Discussion

The discussion of a paper should summarise key results of the study objectives. It should give an overall interpretation of the results of the study keeping in mind the limitations and the external validity of the document. The discussion section should also address both significant and non-significant findings of the study and make comparisons with other research, citing their sources [2,4].

Justified Discussions and Conclusions

In question 17 there is an expectation that the researcher gives an overall summary of the main findings of the study and discusses these in detail. It is important that the reader considers the study as a whole when reading the researcher's conclusion. If the researcher's conclusion is different or is more definitive than the study suggests it should be, it can be an indication that the researcher has misunderstood their own study or has other motives or interests for coming to that conclusion.

It is up to the reader to explore the discussion fully in order to answer question 17. The following points should be taken into account:

Aim

In the discussion section the researcher should discuss all results that pertain to the overall aim of the study, even if they are not significant. If some results are overlooked in the discussion it could suggest that the researcher either doesn't believe the results, or doesn't want to draw attention to controversial discoveries from the study and may therefore be giving a biased overview of the research conducted.

Selection Bias

There is an expectation that the researcher discusses selection biases and takes these into account when interpreting the results of the study. This also gives a clear view of whether the researcher has an overall understanding of the study design. (See notes on selection bias in the methods section).

Non-response

Was there an interpretation of the results that included non-response? This is particularly important if the response rate was low, as non-responders may be a specific group, and lead to a shift in the baseline data (See notes on non-response in the methods section).

Confounding

Confounding is a major threat to the validity of practical inferences made from statistical analyses about cause and effect. Confounding occurs when the outcome of interest is associated with two different independent variables and one of those variables is closely associated with the outcome only because it is closely associated with the other variable (confounder). This can sometimes be accounted for using statistical methods however sometimes these associations are missed because the confounder isn't measured or isn't considered to be a confounder in the analyses. What then happens is an erroneous conclusion is made; that the variable might have a causal relationship with the outcome. The researcher should consider confounding both in the analyses and in the interpretation of the results [18]. An example would be where in a study on cancer a researcher concludes that increased alcohol intake causes lung cancer; however there was confounding in the sample that the researcher didn't discover. People in the study that were inclined to drink more alcohol were also inclined to smoke more (the confounder) and smoking was the cause of lung cancer not increased alcohol intake. Similarly, a study was undertaken to examine surgical deaths in cats. The researcher concluded that cats that had gaseous anaesthesia were more likely to die during surgery than those that had just injectable anaesthesia. There was confounding in the sample: cats that underwent surgery using gaseous anaesthesia were more likely to be ill or undergoing major surgical procedures (the confounders) and this was the cause for cats being more likely to die during surgery and not the use of gaseous anaesthetics.

Non-significant Results

Discussing non-significant results is as important as discussing significant results and should also be included in the discussion, especially if they have a direct association with the aim being investigated. Non-significant results can be influenced by factors associated with study design and

sample size. If there are biases introduced during the study design this can lead to non-significant results that in reality may be significant (this can work the other way around as well). If there are only small differences between groups, non-significant results may be apparent because the sample size is too small (see sample size justification). Again it is important that the researcher has a clear understanding of this and conveys that in the discussion.

Limitations

In question 18 we explore whether limitations are discussed. Unfortunately all forms of research have some limitations. The question here is whether the researcher has an understanding of the limitations involved in their study design. If this issue is not explored, this is cause for concern that the limitations don't stop at the design and that the researcher has a poor understanding of the study as a whole.

Other

Conflicts of Interest

It is very important that conflicts of interest or bodies involved in funding the study are declared in the text (question 19). This can give an impression as to background reasons for carrying out the study. Where studies are funded by a specific agency the researcher may unconsciously interpret in favour of the agencies' ideals; if the researcher has worked in a specific area their own ideas and beliefs may affect the interpretation of the results. It is up to the reader to identify these and come to the conclusion as to whether these conflicts of interest are relevant or not. This can be declared in different areas of the text and should be stated.

Ethical Approval

Question 20 deals with ethical approval and participant consent. It is important that these are sought before carrying out research on any animal or person.

References:

1 Young JM, Solomon MJ. How to critically appraise an article. *Nature Clinical Practice Gastroenterology & Hepatology* 2009;**6**:82–91. doi:10.1038/ncpgasthep1331

2 Lebrun J-L. *Scientific Writing 2.0: A Reader and Writers Guide*. World Scientific 2011.

3 Masic I. How to Search, Write, Prepare and Publish the Scientific Papers in the Biomedical Journals. *Acta Inform Med* 2011;**19**:68–79. doi:10.5455/aim.2011.19.68-79

4 Von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 2007;**370**:1453–7. doi:10.1016/S0140-6736(07)61602-X

5 Leeuw ED, Hox JJ, Dillman DA. Data Analysis. In: *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 317–86.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

6 Rothman KJ. Chapter 4: Types of Epidemiologic Study. In: *Epidemiology: An Introduction*. New York: : Oxford University Press 2002. 57–93.

7 Woodward M. Chapter 8: Sample Size Determination. In: *Epidemiology: Study Design and Data Analysis, Second Edition*. Chapman and Hall/CRC 2004. 381–426.

8 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 1: An Introduction to Survey Methodology. In: *Survey methodology*. Hoboken, NJ: 2004. 1–37.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

9 Lohr SL. Chapter 6: Coverage and Sampling. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 97–112.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

10 Woodward M. Chapter 1: Fundamental Issues. In: *Epidemiology: Study Design and Data Analysis, Second Edition*. Chapman and Hall/CRC 2004. 1–32.

11 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 3: Target Populations, Sampling Frames and Coverage Error. In: *Survey methodology*. Hoboken, NJ: 2004. 67–92.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

12 Dohoo IR, Martin SW, Stryhn H. Capther 2: Sampling. In: *Veterinary Epidemiologic Research*. Charlottetown, Canada: : Ver Books 2009. 27–52.

13 Rothman KJ. Chapter 5: Bias in Study Design. In: *Epidemiology: An Introduction*. New York: : Oxford University Press 2002. 94–112.

14 Bruce N, Pope D, Stanistreet D. *Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics*. 1st ed. Wiley-Blackwell 2008.

15 Groves RM, Fowler FJJ, Couper MP, *et al.* Chapter 2: Inference and Error in Surveys. In: *Survey methodology*. Hoboken, NJ: 2004. 39–65.http://www.loc.gov/catdir/toc/wiley041/2004044064.html http://www.loc.gov/catdir/bios/wiley047/2004044064.html http://www.loc.gov/catdir/description/wiley041/2004044064.html

16 Lynn P. Chapter 3: The Problem of Nonresponse. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. 35–96.http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

17 Hox JJ. Chapter 20: Accommodating Measurement Errors. In: Leeuw ED, Hox JJ, Dillman DA, eds. *International handbook of survey methodology*. New York, NY [etc.]: Erlbaum 2008. http://library.wur.nl/WebQuery/clc/1877609 (accessed 26 Nov2012).

18 Rothman KJ. *Epidemiology: An Introduction*. New York: : Oxford University Press 2002.