

BMJ Open 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 Dean²

To cite: Downes MJ, Brennan ML, Williams HC, *et al.* Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open* 2016;**6**:e011458. doi:10.1136/bmjopen-2016-011458

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-011458>).

MJD, MLB and RSD contributed equally to this work.

Received 15 February 2016
Revised 7 November 2016
Accepted 8 November 2016



CrossMark

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

²Centre for Evidence-based Veterinary Medicine, The University of Nottingham, Loughborough, UK

³Centre of Evidence Based Dermatology, The University of Nottingham, Nottingham, UK

Correspondence to

Dr Rachel S Dean;
rachel.dean@nottingham.ac.uk

ABSTRACT

Objectives: The aim of this study was to develop a critical appraisal (CA) tool that addressed study design and reporting quality as well as the risk of bias in cross-sectional studies (CSSs). 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 CA tool for CSSs. 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.

Conclusions: CA 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 CSSs in systematic reviews, guidelines and clinical decision-making.

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 CA to assess the quality of studies that have been

Strengths and limitations of this study

- 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.
- 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.

included is a crucial part of this process.⁵ Teaching CA 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 using systematic reviews of randomised controlled 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 (CSSs) are one of those study designs that are of increasing importance in evidence-based medicine (EBM). A CSS 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 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 do not facilitate a detailed and specific enough appraisal to be able to fully critique a CSS. In addition, it is only intended for use to assess risk of bias when making judgements about an intervention. Two systematic reviews failed to identify a standalone appraisal tool specifically aimed at CSSs.^{12 13} Katrak *et al* identified that CA tools had been formulated specifically for individual research questions but were not transferable to other CSSs. We identified an appraisal tool, developed in Spanish, which specifically examined CSSs.¹⁵ Berra *et al* 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*¹⁵ only appraises the quality of reporting of CSSs and does not address risk of bias or other aspects of study quality.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 CSSs in evidence synthesis grows, the importance of understanding the quality of reporting and assessment of bias of CSSs becomes increasingly important. Therefore, a robust CA 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 CA tool that was simple to use, that addressed study design quality (design and reporting) and risk of bias in CSSs. A secondary aim was to produce a document to aid the use of the CA tool where appropriate.

METHODS

Development of the initial draft CA tool

The authors completed a systematic search of the literature for CA tools of CSSs (see online supplementary table S1). 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 CA tool. These items were discussed with RSD and a first draft of the tool (see online supplementary table S2) and accompanying help text was created using previously published CA 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 used CSS design methodology during journal clubs and research meetings and provided feedback on their experience. The tool was used in the analysis of CSSs 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 CSSs (see online supplementary table S3) 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 CA tool and aid the development of the help text. Participants for the Delphi panel were sought from the fields of 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 (eg, PhD, MSc, European College Diploma in Veterinary Public Health);

- ▶ they were recognised through publication and/or key note presentations for their work in EBM 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 CA.

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 >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 I described above) of the CA tool (see online supplementary table S3) 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 4 weeks to complete their assessment of the tool using the questionnaire. Participants were reminded about the work required after 1 week, and again 3 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 CA 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 >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.

RESULTS

The initial review of existing tools and texts identified 34 components that were deemed relevant for CA of CSSs and were included in the first draft of the tool (see online supplementary table S2). When piloted, there was an overall per cent agreement of 88.9%; however, 32.9% of the questions were unanswered. Postfeedback modification after the pilot study identified 37 components to be included in the second draft of the CA tool (see online supplementary table S3).

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 were 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, 8 were involved in clinical, teaching and research duties and 10 were involved in research and teaching, 5 of the participants were veterinary surgeons and 6 were medical clinicians. It was an international panel, including 10 participants from the UK, 3 from Australia, 2 from the USA, 2 from Canada and 1 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).

During round 1 (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 4 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 CA 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.

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

I don't believe a Discussion section should be part of a critical appraisal.

Therefore, in round 1, the tool was modified in an attempt to reduce its size and to encompass all comments. For round 2 (undertaken in May 2013), 11 components remained the same and did not require testing for consensus as this was established in round 1; 9 components that had previously reached consensus were incorporated with the 13 components that required modification to create 10 new components (see online supplementary table S4).

In round 2, 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 CA 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 fourth draft of the tool with 20 components incorporating 13 components with full consensus and 7 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 CA tool for CSSs (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 (see online supplementary material 4). Seven (1, 4, 10, 11, 12, 16 and 18) of the final questions related to quality of

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

	Do not know/	
	Yes	No comment
<i>Introduction</i>		
1		
<i>Methods</i>		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
<i>Results</i>		
12		
13		
14		
15		
16		
<i>Discussion</i>		
17		
18		
<i>Other</i>		
19		
20		

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).

DISCUSSION

The AXIS tool

A CA tool to assess the quality and risk of bias in CSSs (AXIS), along with supporting help text, was successfully developed by an expert panel using Delphi methodology. This is the first CA 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 prerequisite 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 (eg, 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 CA tools for other types of studies.^{12 16 20 28 36}

As with all CA 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 performed, and not reported, or not performed 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 CA of CSSs and to ensure no other relevant tools existed. While numerous tools exist for CA, we found a lack of tools for general use in CSSs and this was consistent with what others have found previously.^{12 13} In order to ensure quality and completeness of the tool, we utilised 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 EBM limits the effect of using a non-representative sample, and the use of the Delphi tool is well recognised for developing consensus in healthcare science.³⁸ The selection of a Delphi group is very important as it effects the results of the process.³¹ As CSSs are used extensively in 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 veterinary and medical groups in this study, which adds to the rigour of the tool but also demonstrates how both healthcare areas can cooperate 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 RCTs. 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 quality of reporting and risk of bias when making judgements on the quality of a paper. This tool therefore provides an advantage over Berra *et al*¹⁵ which only allows the user to assess quality of reporting and tools such as the Cochrane risk of bias tool⁵ 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 CA of studies using AXIS is repeatable.

CONCLUSION

In conclusion, a unique tool (AXIS) for the CA of CSSs was developed that can be used across disciplines, for example, 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.

Twitter Follow Rachel Dean at @CEVetM

Acknowledgements 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 to 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 to thank Michelle Downes for designing the population diagram. The Centre for Evidence-based Veterinary Medicine is supported by an unrestricted grant from Elanco Animal Health and The University of Nottingham.

Contributors 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 agree 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.

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

Competing interests None declared.

Ethics 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.

Provenance and peer review Not commissioned; externally peer reviewed.

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

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Cockcroft PD, Holmes MA. *Handbook of evidence-based veterinary medicine*. 1st edn. Oxford, UK: Blackwell Publishing Ltd, 2003.
- McColl A, Smith H, White P, *et al*. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ* 1998;316:361–5.
- Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* 1995;310:1122–6.
- Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. *BMJ* 2001;323:833–6.
- Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*. Wiley Online Library, 2008.
- Bennett KJ, Sackett DL, Haynes RB, *et al*. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA* 1987;257:2451–4.
- 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;26:4–13.
- Parkes J, Hyde C, Deeks J, *et al*. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev* 2001;(3):CD001270.
- Sackett DL. Teaching critical appraisal. *J Gen Intern Med* 1990;5:272.
- Ho PM, Peterson PN, Masoudi FA. Evaluating the evidence: is there a rigid hierarchy? *Circulation* 2008;118:1675–84.
- Dohoo IR, Martin SW, Stryhn H. *Veterinary epidemiologic research*. Charlottetown, Canada: Ver Books, 2009.
- Katrak P, Bialocerkowski AE, Massy-Westropp N, *et al*. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 2004;4:22.
- 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. *Int J Epidemiol* 2007;36:666–76.
- 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-randomised studies of interventions. 2016. *BMJ* 2016;355:i4919
- Berra S, Eiorza-Ricart JM, Estrada MD, *et al*. [A tool (corrected) for the critical appraisal of epidemiological cross-sectional studies]. *Gac Sanit* 2008;22:492–7.
- Chalmers TC, Smith H, Jr., Blackburn B, *et al*. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;2:31–49.

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. *Bull World Health Organ* 2007;85:867–72.
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;354:1896–900.
19. Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:834–40.
20. Shea BJ, Grimshaw JM, Wells GA, *et al*. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
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 edn. Wiley-Blackwell. 1 edn. London, UK: BMJ Publishing Group, 1996.
23. Fowkes FGR, Fulton PM. Critical-appraisal of published research—introductory guidelines. *BMJ* 1991;302:1136–40.
24. Genaidy AM, Lemasters GK, Lockey J, *et al*. An epidemiological appraisal instrument—a tool for evaluation of epidemiological studies. *Ergonomics* 2007;50:290–60.
25. Hanke BA, Motschall E, Türp JC. Association between orthopedic and dental findings: what level of evidence is available? *J Orofac Orthop* 2007;68:91–107.
26. Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional, 2004. <http://heb.wcf.ac.uk/projectmethod/appendix8.htm>
27. Rothman K. Chapter 5: Bias in study design. In: *Epidemiology: an introduction*. New York: Oxford University 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. *Emerg Themes Epidemiol* 2008;5:23.
29. Joanna Briggs Institute. *Joanna Briggs Institute reviewers' manual 2014 Edition*. 2014. <http://joannabriggs.org/assets/docs/sumari/ReviewersManual-2014.pdf>
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 Vet Res* 2013;9:121.
31. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007;12:1–8.
32. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inform Manage* 2004;42:15–29.
33. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;328:39–41.
34. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
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;283:2008–12.
36. Wells G, Shea B, O'connell D, *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;64:79–89.
38. Fink A, Kosecoff J, Chassin M, *et al*. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–83.
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.
40. Mckenna HP. The Delphi technique—a worthwhile research approach for nursing. *J Adv Nurs* 1994;19:1221–5.
41. Linstone HA, Turoff M. *The Delphi method: techniques and applications*. MA: Addison-Wesley Reading, 1975.
42. Higgins JPT, Altman DG, Gotzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Br Med J* 2011;343:d5928.
43. Juni P, Witschi A, Bloch R, *et al*. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.