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".

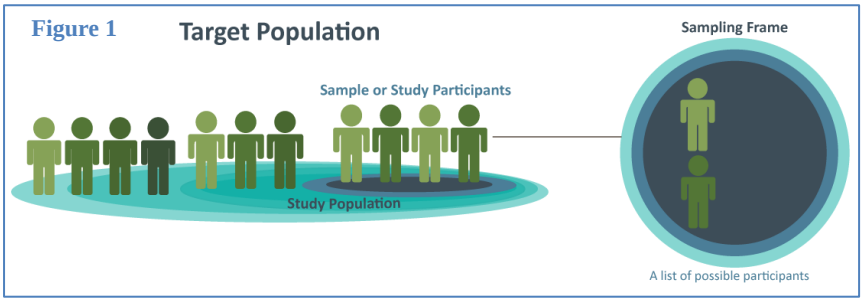
Question 3 asks if sample size justification was reported, but it should also be clear what methods were used to determine the sample size. In some cases clustering of observations within groups can occur (e.g. patients within hospitals or livestock within herds) and this should be taken into account 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 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.