



The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence

Zachary Munn*, Sandeep Moola, Dagmara Riitano, Karolina Lisy

Abstract

Background: Recently there has been a significant increase in the number of systematic reviews addressing questions of prevalence. Key features of a systematic review include the creation of an a priori protocol, clear inclusion criteria, a structured and systematic search process, critical appraisal of studies, and a formal process of data extraction followed by methods to synthesize, or combine, this data. Currently there exists no standard method for conducting critical appraisal of studies in systematic reviews of prevalence data.

Methods: A working group was created to assess current critical appraisal tools for studies reporting prevalence data and develop a new tool for these studies in systematic reviews of prevalence. Following the development of this tool it was piloted amongst an experienced group of sixteen healthcare researchers.

Results: The results of the pilot found that this tool was a valid approach to assessing the methodological quality of studies reporting prevalence data to be included in systematic reviews. Participants found the tool acceptable and easy to use. Some comments were provided which helped refine the criteria.

Conclusion: The results of this pilot study found that this tool was well-accepted by users and further refinements have been made to the tool based on their feedback. We now put forward this tool for use by authors conducting prevalence systematic reviews.

Keywords: Prevalence, Survey, Critical Appraisal, Systematic Review

Copyright: © 2014 by Kerman University of Medical Sciences

Citation: Munn Z, Moola S, Riitano D, Lisy K. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *Int J Health Policy Manag* 2014; 3: 123–128. doi: 10.15171/ijhpm.2014.71

Article History:

Received: 18 June 2014

Accepted: 11 August 2014

ePublished: 13 August 2014

*Correspondence to:

Zachary Munn

Email: zachary.munn@adelaide.edu.au

Introduction

The prevalence of a disease indicates the number of people in a population that have the disease at a given point in time (1). The accurate measurement of disease burden among populations, whether at a local, national, or global level, is of critical importance for governments, policy-makers, health professionals and the general population to inform the development, delivery and use of health services. For example, accurate information regarding measures of disease can assist in planning management of disease services (by ensuring sufficient resources are available to cope with the burden of disease), set priorities regarding public health initiatives, and evaluate changes and trends in diseases over time. However, policy-makers are often faced with conflicting reports of disease prevalence in the literature.

The systematic review of evidence has been proposed and is now well-accepted as the ideal method to summarize the literature relating to a certain social or healthcare topic (2,3). The systematic review can provide a reliable summary of the literature to inform decision-makers in a timely fashion. Key features of a systematic review include the creation of an a priori protocol, clear inclusion criteria, a comprehensive and systematic search process, the critical appraisal of studies, and a formal process of data extraction followed by methods to synthesize, or combine, this data (4). In this way, systematic reviews extend beyond the subjective, narrative reporting characteristics of a traditional literature review to provide a

comprehensive, rigorous, and transparent synthesis of the literature on a certain topic.

Historically, systematic reviews have predominantly focused on the synthesis of quantitative evidence to establish the effects of interventions. In the last five years, there has been a substantial increase in the number of systematic reviews addressing questions of prevalence (Figure 1). However, currently there does not appear to be any formal guidance for authors wishing to conduct a review of prevalence. Consequently, there is significant variability in the methods used to conduct these reviews.

The Joanna Briggs Institute (JBI) and the Cochrane Collaboration are evidence-based organizations that were formed to develop methodologies and guidance on the process of conducting systematic reviews (2,5–8). In 2012, a working group was formed within the Joanna Briggs Institute to evaluate systematic reviews of prevalence and develop guidance for researchers wishing to conduct such reviews. The group identified that the major area where prevalence reviews were disparate was in their conduct of critical appraisal or quality assessment of included studies. For example, whilst some reviews used instruments that were appropriate for reviews of prevalence data (9,10), others used instruments or criteria not designed to critically appraise studies reporting prevalence (such as reporting guidelines, study design specific tools, or self-developed criteria for their review question) (11–13), or refrained from conducting a

formal quality assessment altogether (14,15). Therefore, the working group sought to address this gap by developing and testing a critical appraisal form that could be used for studies included in systematic reviews of prevalence data.

Materials and methods

Developing the Tool

The working party began by conducting a search for systematic reviews of prevalence data to determine how the methodological quality of studies included in these reviews were assessed. The group then searched for critical appraisal tools that have been used to assess studies reporting on prevalence data. A number of tools were identified including the Joanna Briggs Institute's Descriptive/Case series critical appraisal tool. A non-exhaustive list is shown in Table 1. Critical appraisal tools from the Cochrane Collaboration and the Critical Appraisal Skills Program (CASP) were also identified.

Although many of these checklists identified important criteria, it was felt by the group that none of these tools were complete and ideal for use during assessment of quality during the systematic review process. Based on a review of these criteria and our own knowledge and research we developed a tool specifically for use in systematic reviews of prevalence data. This tool was initially trialed by the working group and refined until it was deemed ready for further

external review. Details on how to answer each question in the tool are available in Appendix 1 and the Joanna Briggs Institute guidance on conducting prevalence and incidence reviews (Table 2) (21).

Pilot testing

A pilot of the tool was conducted during the 2013 Joanna Briggs Institute convention in Adelaide during October of that year. A workshop was held on systematic reviews of prevalence and incidence where attendees were given a cross-sectional study (22) to appraise with the new tool, along with a short survey that was developed to establish the face validity, ease of use, acceptability and timeliness (i.e. time taken to complete) of the tool, and feedback on areas for improvement (23). The questions asked and how they were measured is reported in Table 3.

Results

Sixteen workshop participants completed the critical appraisal task and survey. Of the 16, 13 participants stated they had an academic/research background, 2 said they had a health background, and one said they had both. The average time spent working in research was 11 years, with the minimum being 2.5 years and the maximum experience being 30 years. For ease of use of the critical appraisal tool, the mean score on the 5-point Likert Scale was 3.63, with the majority (75%) of

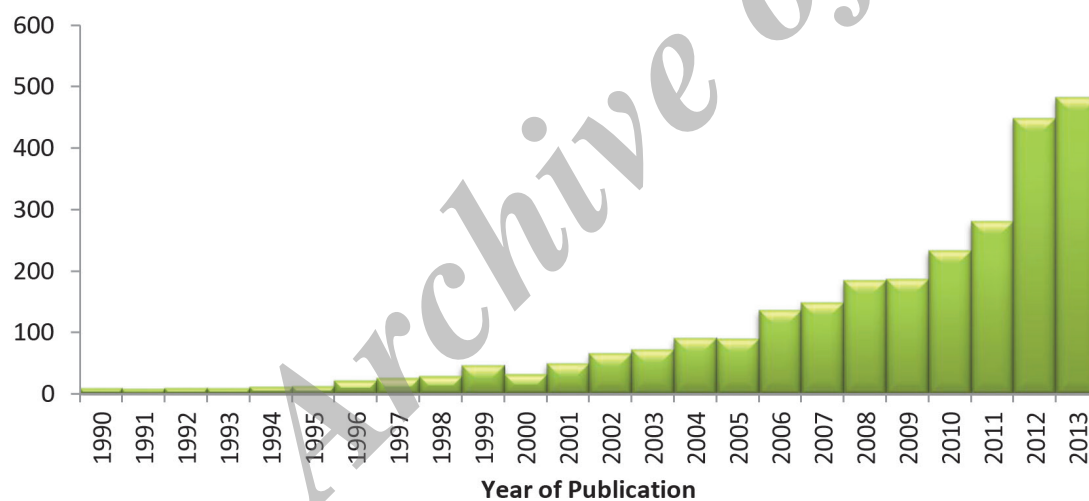


Figure 1. Number of systematic reviews of prevalence by year of publication identified in a PubMed search

Table 1. Existing critical appraisal tools

Name	Number of Criteria	Comments
JBIC (3) Descriptive/Case series studies appraisal tool	9	Targeted towards specific study designs
Centre for Evidence-Based Management Critical Appraisal of a Survey (16)	12	Targeted towards a specific study design
Loney <i>et al.</i> Critical Appraisal tool for prevalence (17)	8	Designed specifically for studies assessing prevalence
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist (18)	22	Targeted towards reporting
National Collaborating Centre for Environmental Health Critical Appraisal of Cross-Sectional Studies (19)	11	Addresses external and internal validity as well as reporting standards
Hoy <i>et al.</i> 's risk of bias tool (20)	10	Addresses external and internal validity

Table 2. The Joanna Briggs Institute Prevalence Critical Appraisal Tool

Criteria	Yes	No	Unclear	Not applicable
1. Was the sample representative of the target population?				
2. Were study participants recruited in an appropriate way?				
3. Was the sample size adequate?				
4. Were the study subjects and the setting described in detail?				
5. Was the data analysis conducted with sufficient coverage of the identified sample?				
6. Were objective, standard criteria used for the measurement of the condition?				
7. Was the condition measured reliably?				
8. Was there appropriate statistical analysis?				
9. Are all important confounding factors/subgroups/differences identified and accounted for?				
10. Were subpopulations identified using objective criteria?				

Table 3. Survey pilot tool

Question	Measurement
Ease of use of the tool	5-point Likert scale (1 very difficult, 5 very easy)
Is this a valid tool for prevalence data?	Yes/No
Timeliness	5-point Likert Scale (1 very unacceptable, 5 very acceptable)
Acceptability	5-point Likert Scale (1 very unacceptable, 5 very acceptable)
Redundant questions	Free text
Comments for improvement	Free text

participants providing a rating of 4, corresponding to 'easy'. For the acceptability of the tool, the mean score was 4.33, with all participants giving either a ranking of 4 (acceptable) or 5 (very acceptable). For timeliness, the mean score was 3.94, with 88% providing a score of 4 (acceptable). Out of all the participants, all except 1 viewed the tool as a valid quality appraisal checklist for prevalence data (Table 4). There were a number of suggestions provided for refinement and improvement of the tool. These comments resulted in some changes in the order of the questions of the tool and the supporting information used to assist in judging criteria, although no changes were made to the individual questions.

Discussion

Systematic reviews of prevalence and incidence data are becoming increasingly important as decision makers realize their usefulness in informing policy and practice. These reviews have the potential to better support healthcare professionals, policy-makers, and consumers in making evidence-based decisions that effectively target and address

burden of disease issues both now and in to the future.

The conduct of a systematic review is a scientific exercise that produces results which may influence healthcare decisions. As such, reviews are required to have the same rigor expected of all research. The quality of a review, and any recommendations for practice that may arise, depends on the extent to which scientific review methods are followed to minimize the risk of error and bias. The explicit and rigorous methods of the process distinguish systematic reviews from traditional reviews of the literature (2).

Systematic reviews normally rely on the use of critical appraisal checklists that are tailored to assess the quality of a particular study design. For example, there may be separate checklists used to appraise randomized controlled trials, cohort studies, cross-sectional studies and so on. Prevalence data can be sourced from various study designs, even randomized controlled trials (11); however, critical appraisal tools directed at assessing the risk of bias of randomized controlled trials are aimed at assessing biases related to causal effects and hence are not appropriate for reviews examining the prevalence

Table 4. Results from the survey

	Experience	Ease of tool use	Acceptability	Time
Number of cases	15	16	15	16
Minimum	2.50	2.00	4.00	2.00
Maximum	30.00	4.00	5.00	5.00
Arithmetic mean	11.37	3.63	4.33	3.94
95.00% lower confidence limit	6.94	3.24	4.06	3.63
95.00% upper confidence limit	15.80	4.00	4.60	4.24
Standard deviation	8.00	0.72	0.49	0.57

of a condition. For example, criteria regarding the use of an intention-to-treat analysis as often seen in critical appraisal checklists for randomized controlled trials are not a true quality indicator for questions of prevalence.

Due to this, a new tool assessing validity and quality indicators specific to issues of prevalence has been developed. This checklist addresses critical issues of internal and external validity that must be considered when assessing validity of prevalence data that can be used across different study designs (not just cross-sectional studies but all studies that might report prevalence data). The criteria address the following issues:

- Ensuring a representative sample.
- Ensuring appropriate recruitment.
- Ensuring an adequate sample size.
- Ensuring appropriate description and reporting of study subjects and setting.
- Ensuring data coverage of the identified sample is adequate.
- Ensuring the condition was measured reliably and objectively.
- Ensuring appropriate statistical analysis.
- Ensuring confounding factors/subgroups/differences are identified and accounted for.

A pilot test of this tool amongst a group of experienced healthcare professionals and researchers found that this tool had face validity and high levels of acceptability, ease of use and timeliness to complete. The initial results of this pilot testing are encouraging. This tool now needs to be tested further in a larger scale study to assess its other clinimetric properties, particularly its construct validity and inter-rater reliability. We have developed this tool as we did not feel that any of the current checklists identified from our search sufficiently addressed important quality issues in prevalence studies. Some of the tools [most notably the tool refined by Hoy *et al.* (20)] contain similar questions to our tool but there are important differences. For example, we provide a criteria regarding sample size which is not included in the Hoy *et al.* checklist. Our tool also has the advantage of being simple, easy and quick as shown during the pilot testing. This tool will now be incorporated into the next version of the Joanna Briggs Institute's systematic review package.

Conclusion

Critical appraisal is a pivotal step in the process of systematic reviews. As reviews of questions addressing prevalence become more well-known, critical appraisal tools addressing studies reporting prevalence data are needed. Following a search of the literature a new tool has been proposed that can be used for studies reporting prevalence data, developed by a working party within the Joanna Briggs Institute. The results of this pilot study found that this tool was well-accepted by users and further refinements have been made to the tool based on their feedback. We now put forward this tool for use by authors conducting prevalence systematic reviews.

Acknowledgements

The authors would like to acknowledge the participants of the

workshop and the wider staff of the Joanna Briggs Institute and its collaboration for their feedback.

Ethical issues

No ethical issues are raised.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZM lead the methodological group and drafted the paper. SM, DR, and KL were members of the working group and provided substantial input regarding its development and testing.

References

1. Webb P, Bain C, Pirozzo S. *Essential epidemiology: an introduction for students and health professionals*. New York: Cambridge University Press; 2005.
2. Munn Z, Tufanaru C, Aromataris E. Data extraction and synthesis in systematic reviews. *Am J Nurs* 2014; 114: 49–54. doi: [10.1097/01.naj.0000451683.66447.89](https://doi.org/10.1097/01.naj.0000451683.66447.89)
3. The Joanna Briggs Institute. *Reviewer's Manual*. Australia: The Joanna Briggs Institute; 2014.
4. Pearson A, Robertson-Malt S, Rittenmeyer L. *Synthesizing Qualitative Evidence*. Philadelphia: Lippincott Williams & Wilkins; 2011.
5. Noyes J, Popay J, Pearson A, Hannes K, Booth A. Qualitative research and Cochrane reviews. In: Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
6. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBIR Reports* 2004; 2: 45–64. doi: [10.1111/j.1479-6988.2004.00008.x](https://doi.org/10.1111/j.1479-6988.2004.00008.x)
7. Pearson A, Jordan Z, Munn Z. Translational science and evidence-based healthcare: a clarification and reconceptualization of how knowledge is generated and used in healthcare. *Nursing Research and Practice* 2012; 2012: 792519. doi: [10.1155/2012/792519](https://doi.org/10.1155/2012/792519)
8. The Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual: 2011 edition*. Adelaide: The Joanna Briggs Institute; 2011.
9. Sawyer A, Ayers S, Smith H. Pre- and postnatal psychological wellbeing in Africa: a systematic review. *J Affect Disord* 2010; 123: 17–29. doi: [10.1016/j.jad.2009.06.027](https://doi.org/10.1016/j.jad.2009.06.027)
10. Mirza I, Jenkins R. Risk factors, prevalence, and treatment of anxiety and depressive disorders in Pakistan: systematic review. *BMJ* 2004; 328: 794. doi: [10.1136/bmj.328.7443.794](https://doi.org/10.1136/bmj.328.7443.794)
11. Van Lancker A, Velghe A, Van Hecke A, Verbrugghe M, Van Den Noortgate N, Grypdonck M, *et al.* Prevalence of symptoms in older cancer patients receiving palliative care: a systematic review and meta-analysis. *J Pain Symptom Manage* 2014; 47: 90–104. doi: [10.1016/j.jpainsymman.2013.02.016](https://doi.org/10.1016/j.jpainsymman.2013.02.016)
12. Klaassen KM, Dulak MG, van de Kerkhof PC, Pasch MC. The prevalence of onychomycosis in psoriatic patients: a systematic review. *J Eur Acad Dermatol Venereol* 2014; 28: 533–41. doi: [10.1111/jdv.12239](https://doi.org/10.1111/jdv.12239)
13. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* 2004; 2:13. doi: [10.1186/1741-7015-2-13](https://doi.org/10.1186/1741-7015-2-13)
14. Goto A, Goto M, Noda M, Tsugane S. Incidence of type 2 diabetes in Japan: a systematic review and meta-analysis. *PLoS*

- One 2013; 8: e74699. doi: [10.1371/journal.pone.0074699](https://doi.org/10.1371/journal.pone.0074699)
15. Bahekar AA, Singh S, Saha S, Molnar J, Arora R. The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. *Am Heart J* 2007; 154: 830–7. doi: [10.1016/j.ahj.2007.06.037](https://doi.org/10.1016/j.ahj.2007.06.037)
 16. Centre for Evidence-Based Management. Critical appraisal of a survey. [updated 2014 June 5]. Available from: <http://www.cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-Survey.pdf>
 17. Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: prevalence or incidence of a health problem. *Chronic Dis Can* 1998; 19: 170–6.
 18. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4: e297. doi: [10.1371/journal.pmed.0040297](https://doi.org/10.1371/journal.pmed.0040297)
 19. National Collaborating Centre for Environmental Health (NCCEH). A Primer for Evaluating the Quality of Studies on Environmental Health Critical Appraisal of Cross-Sectional Studies. [updated 2014 June 5]. Available from: http://www.nccch.ca/sites/default/files/Critical_Appraisal_Cross-Sectional_Studies_Aug_2011.pdf
 20. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, *et al.* Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol* 2012; 65: 934–9. doi: [10.1016/j.jclinepi.2011.11.014](https://doi.org/10.1016/j.jclinepi.2011.11.014)
 21. Munn Z, Moola S, Lisy K, Riitano D. *The Systematic Review of Prevalence and Incidence Data The Joanna Briggs Institute Reviewer's Manual 2014*. Australia: The Joanna Briggs Institute; 2014.
 22. Lim ES, Ko YK, Ban KO. Prevalence and risk factors of metabolic syndrome in the Korean population—Korean National Health Insurance Corporation Survey 2008. *J Adv Nurs* 2013; 69: 1549–61. doi: [10.1111/jan.12013](https://doi.org/10.1111/jan.12013)
 23. Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001; 54: 651–4. doi: [10.1016/s0895-4356\(00\)00360-7](https://doi.org/10.1016/s0895-4356(00)00360-7)
 24. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. *Archives of Orofacial Sciences* 2006; 1: 9–14.
 25. Daniel WW. *Biostatistics: A Foundation for Analysis in the Health Sciences*. 7th ed. New York: John Wiley & Sons; 1999.

Key Messages

Implications for policy makers

- Until now there has been substantial variability in how studies reporting prevalence data are critically appraised. The tool proposed within this paper can be considered as a valid option for researchers and policy-makers when conducting systematic reviews of prevalence.
- This tool guides the assessment of internal and external validity of studies reporting prevalence data.

Implications for public

Systematic reviews are of great importance to provide a critical summary of the research and inform evidence-based practice. Prevalence systematic reviews are becoming increasingly popular within the research community. This article proposes a new tool that can be used during the conduct of these types of systematic reviews to critically appraise studies to ensure that their results are valid.

Appendix 1

Prevalence Critical Appraisal Instrument

The 10 criteria used to assess the methodological quality of studies reporting prevalence data and an explanation are described below. These questions can be answered either with a yes, no, unclear, or not applicable.

Answers: Yes, No, Unclear or Not/Applicable

1. Was the sample representative of the target population?

This question relies upon knowledge of the broader characteristics of the population of interest. If the study is of women with breast cancer, knowledge of at least the characteristics, demographics, and medical history is needed. The term “target population” should not be taken to infer every individual from everywhere or with similar disease or exposure characteristics. Instead, give consideration to specific population characteristics in the study, including age range, gender, morbidities, medications, and other potentially influential factors. For example, a sample may not be representative of the target population if a certain group has been used (such as those working for one organisation, or one profession) and the results then inferred to the target

population (i.e. working adults).

2. Were study participants recruited in an appropriate way?

Recruitment is the calling or advertising strategy for gaining interest in the study, and is not the same as sampling. Studies may report random sampling from a population, and the methods section should report how sampling was performed. What source of data were study participants recruited from? Was the sampling frame appropriate? For example, census data is a good example of appropriate recruitment as a good census will identify everybody. Was everybody included who should have been included? Were any groups of persons excluded? Was the whole population of interest surveyed? If not, was random sampling from a defined subset of the population employed? Was stratified random sampling with eligibility criteria used to ensure the sample was representative of the population that the researchers were generalizing to?

3. Was the sample size adequate?

An adequate sample size is important to ensure good precision of the final estimate. Ideally we are looking for evidence that the authors conducted a sample size calculation

to determine an adequate sample size. This will estimate how many subjects are needed to produce a reliable estimate of the measure(s) of interest. For conditions with a low prevalence, a larger sample size is needed. Also consider sample sizes for subgroup (or characteristics) analyses, and whether these are appropriate. Sometimes, the study will be large enough (as in large national surveys) whereby a sample size calculation is not required. In these cases, sample size can be considered adequate.

When there is no sample size calculation and it is not a large national survey, the reviewers may consider conducting their own sample size analysis using the following formula (24,25):

$$n = \frac{Z^2 P(1-P)}{d^2}$$

Where:

n= sample size

Z= Z statistic for a level of confidence

P= Expected prevalence or proportion (in proportion of one; if 20%, P= 0.2)

d= precision (in proportion of one; if 5%, d= 0.05)

4. Were the study subjects and setting described in detail?

Certain diseases or conditions vary in prevalence across different geographic regions and populations (e.g. women vs. men, socio-demographic variables between countries). Has the study sample been described in sufficient detail so that other researchers can determine if it is comparable to the population of interest to them?

5. Is the data analysis conducted with sufficient coverage of the identified sample?

A large number of dropouts, refusals or “not founds” amongst selected subjects may diminish a study’s validity, as can low response rates for survey studies.

- Did the authors describe the reasons for non-response and compare persons in the study to those not in the study, particularly with regards to their socio-demographic characteristics?
- Could the not-responders have led to an underestimate of prevalence of the disease or condition under investigation?
- If reasons for non-response appear to be unrelated to the outcome measured and the characteristics of non-responders are comparable to those in the study, the researchers may be able to justify a more modest response rate.
- Did the means of assessment or measurement negatively affect the response rate (measurement should be easily accessible, conveniently timed for participants, acceptable in length, and suitable in content).

6. Were objective, standard criteria used for measurement of the condition?

Here we are looking for measurement or classification

bias. Many health problems are not easily diagnosed or defined and some measures may not be capable of including or excluding appropriate levels or stages of the health problem. If the outcomes were assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If the outcomes were assessed using observer reported, or self-reported scales, the risk of over- or under-reporting is increased, and objectivity is compromised. Importantly, determine if the measurement tools used were validated instruments as this has a significant impact on outcome assessment validity.

7. Was the condition measured reliably?

Considerable judgment is required to determine the presence of some health outcomes. Having established the objectivity of the outcome measurement instrument (see item 6 of this scale), it is important to establish how the measurement was conducted. Were those involved in collecting data trained or educated in the use of the instrument/s? If there was more than one data collector, were they similar in terms of level of education, clinical or research experience, or level of responsibility in the piece of research being appraised?

- Has the researcher justified the methods chosen?
- Has the researcher made the methods explicit? (For interview method, how were interviews conducted?)

8. Was there appropriate statistical analysis?

As with any consideration of statistical analysis, consideration should be given to whether there was a more appropriate alternate statistical method that could have been used. The methods section should be detailed enough for reviewers to identify the analytical technique used and how specific variables were measured. Additionally, it is also important to assess the appropriateness of the analytical strategy in terms of the assumptions associated with the approach as differing methods of analysis are based on differing assumptions about the data and how it will respond. Prevalence rates found in studies only provide estimates of the true prevalence of a problem in the larger population. Since some subgroups are very small, 95% confidence intervals are usually given.

9. Are all important confounding factors/ subgroups/differences identified and accounted for?

Incidence and prevalence studies often draw or report findings regarding the differences between groups. It is important that authors of these studies identify all important confounding factors, subgroups and differences and account for these.

10. Were subpopulations identified using objective criteria?

Objective criteria should also be used where possible to identify subgroups (refer to question 6).