Supplementary File 3: Critical appraisal of the included studies

JBI Critical Appraisal Checklist for Studies Reporting Prevalence Data

Study	Was the sample frame appropriate to address the target population?[1]	Were study participants sampled in an appropriate way?[2]	Was the sample size adequate?[3]	Were the study subjects and the setting described in detail?[4]	Was the data analysis conducted with sufficient coverage of the identified sample?[5]	Were valid Methods used for de identification of the condition?[6]	Was the condition measured in a standard, reliable way for all participants?[7]	Was there appropriate statistical analysis?[8]	Was the response rate adequate, and if not, was the low response rate managed appropriately?[9]
Bean et al. (2006)	✓	✓	✓	✓	✓	X	✓	✓	✓
Hodes et al. (2008)	X	✓	~	✓	✓	✓	✓	✓	✓
Jensen et al. (2015)	✓	X	~	✓	~	✓	✓	✓	~
Oppedal & Isdoe (2012)	✓	~	✓	✓	~	✓	✓	✓	X
Vervliet et al. (2014b)	✓	X	~	✓	✓	✓	X	✓	✓

✓= yes X= no ~ = unclear

Explanations of the criteria used to perform the critical appraisal of studies reporting prevalence data

(cited from Munn Z, Moola S, Lisy K et al. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and incidence data. *Int J Evid Based Healthc* 2015;13:147–153; adaptations applied for this review marked red)

[1] This question relies upon knowledge of the broader characteristics of the population of interest and the geographical area. 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. […]

[2] Studies may report random sampling from a population, and the methods section should report how sampling was performed. Random probabilistic sampling from a defined subset of the population (sample frame) should be employed in most cases, however, random probabilistic sampling is not needed when everyone in the sampling frame will be included/analysed. For example, reporting on all the data from a good census is appropriate as a good census will identify everybody. When using cluster sampling, such as a random sample of villages within a region, the methods need to be clearly stated as the precision of the final prevalence estimate incorporates the clustering effect. Convenience samples, such as a street survey or interviewing lots of people at a public gatherings are not considered to provide a representative sample of the base population.

[3] The larger the sample, the narrower will be the confidence interval around the prevalence estimate, making the results more precise. 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 rate the adequateness of the sample size as 'unclear'.

[4] 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] Coverage bias can occur when not all subgroups of the identified sample respond at the same rate. For instance, you may have a very high response rate overall for your study, but the response rate for a certain subgroup (i.e. older adults) may be quite low.

[6] 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] 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? When there was more than one observer or collector, was there comparison of results from across the observers? Was the condition measured in the same way for all participants?

[8] Importantly, the numerator and denominator should be clearly reported, and percentages should be given with confidence intervals. 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.

[9] A large number of dropouts, refusals or “not founds” amongst selected subjects may diminish a study’s validity, as can a low response rates for survey studies. The authors should clearly discuss the response rate and any reasons for non-response and compare persons in the study to those not in the study, particularly with regards to their socio-demographic characteristics. If reasons for non-response appear to be unrelated to the outcome measured and the characteristics of non-responders are comparable to those who do respond in the study (addressed in question 5, coverage bias), the researchers may be able to justify a more modest response rate.

JBI Critical Appraisal Checklist for Analytical Cross Sectional Studies

Study	Were the criteria for inclusion in the sample clearly defined?[a]	Were the study subjects and the setting described in detail?[b]	Was the exposure measured in a valid and reliable way?[c]	Were objective, standard criteria used for the measurement of the condition?[d]	Were confounding factors identified?[e]	Were strategies to with confounding factors stated?[f]	Were the outcomes measured in a valid and reliable way?[g]	Was appropriate statistical analysis uses?[h]
Keles et al. (2015)	✓	✓	✓	✓	✓	~	X	✓
Reijneveld et al. (2005)	✓	✓	✓	✓	✓	✓	X	✓
Seglem et al. (2015)	✓	X	✓	X	✓	~	X	✓

✓= yes X= no ~ = unclear

Explanations of the criteria used to perform the critical appraisal of analytical cross sectional studies

(cited from The Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2016 edition. Australia: The Joanna Briggs Institute 2016.)

[a] The authors should provide clear inclusion and exclusion criteria that they developed prior to recruitment of the study participants. The inclusion/exclusion criteria should be specified (e.g., risk, stage of disease progression) with sufficient detail and all the necessary information critical to the study.

[b] The study sample should be described in sufficient detail so that other researchers can determine if it is comparable to the population of interest to them. The authors should provide a clear description of the population from which the study participants were selected or recruited, including demographics, location, and time period.

[c] The study should clearly describe the method of measurement of exposure. Assessing validity requires that a 'gold standard' is available to which the measure can be compared. The validity of exposure measurement usually relates to whether a current measure is appropriate or whether a measure of past exposure is needed. Reliability refers to the processes included in an epidemiological study to check repeatability of measurements of the exposures. These usually include intra-observer reliability and inter-observer reliability.

[d] It is useful to determine if patients were included in the study based on either a specified diagnosis or definition. This is more likely to decrease the risk of bias. Characteristics are another useful approach to matching groups, and studies that did not use specified diagnostic methods or definitions should provide evidence on matching by key characteristics.

[e] Confounding has occurred where the estimated intervention exposure effect is biased by the presence of some difference between the comparison groups (apart from the exposure investigated/of interest). Typical confounders include baseline characteristics, prognostic factors, or concomitant exposures (e.g. smoking). A confounder is a difference between the comparison groups and it influences the direction of the study results. A high quality study at the level of cohort design will identify the potential confounders and measure them (where possible). This is difficult for studies where behavioral, attitudinal or lifestyle factors may impact on the results.

[f] Strategies to deal with effects of confounding factors may be dealt within the study design or in data analysis. By matching or stratifying sampling of participants, effects of confounding factors can be adjusted for. When dealing adjustment in data analysis, assess the statistics used in the study. Most will be some form of multivariate regression analysis to account for the confounding factors measured.

[g] Read the methods section of the paper. If for e.g. lung cancer is assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If lung cancer is 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. Having established the objectivity of the outcome measurement (e.g. lung cancer) instrument,

it's important to establish how the measurement was conducted. Were those involved in collecting data trained or educated in the use of the instrument/s? (e.g. radiographers). 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?

[h] 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 of cohort studies should be detailed enough for reviewers to identify which analytical techniques were used (in particular, regression or stratification) and how specific confounders were measured. For studies utilising regression analysis, it is useful to identify if the study identified which variables were included and how they related to the outcome. If stratification was the analytical approach used, were the strata of analysis defined by the specified variables? 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.

JBI Critical Appraisal Checklist for Cohort Studies

Study	Were the groups similar and recruited from the same population?[A]	Were the exposures measured similarly to assign people to both exposed and unexposed groups?[B]	Was the exposure measured in a valid and reliable way?[C]	Were confounding factors identified?[D]	Were strategies to deal with confounding factors stated?[E]	Were the groups / participants free of the outcome at the start of the study (or at the moment of exposure)?[F]	Were the outcomes measured in a valid and reliable way?[G]	Was the follow up time reported and sufficient to belong enough for outcome to occur?[H]	Was the follow-up complete, and if not, were the reasons to loss to follow-up described and explored?[I]	Were strategies to address incomplete follow-up utilized?[J]	Was appropriate statistical analysis used?[K]
Jensen et al. (2014)	O	O	O	X	X	O	X	✓	✓	X	✓
Smid et al. (2011)	O	O	O	✓	✓	O	X	✓	✓	X	✓
Vervliet et al. (2014a)	O	O	O	X	X	O	X	✓	✓	X	✓

✓= yes X= no ~ = unclear O= not applicable

Explanations of the criteria used to perform the critical appraisal of longitudinal studies

(cited from The Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2016 edition. Australia: The Joanna Briggs Institute 2016.)

[A] Check the paper carefully for descriptions of participants to determine if patients within and across groups have similar characteristics in relation to exposure (e.g. risk factor under investigation). The two groups selected for comparison should be as similar as possible in all characteristics except for their exposure status, relevant to the study in question. The authors should provide clear inclusion and exclusion criteria that they developed prior to recruitment of the study participants.

[B] A high quality study at the level of cohort design should mention or describe how the exposures were measured. The exposure measures should be clearly defined and described in detail. This will enable reviewers to assess whether or not the participants received the exposure of interest.

[C] The study should clearly describe the method of measurement of exposure. Assessing validity requires that a 'gold standard' is available to which the measure can be compared. The validity of exposure measurement usually relates to whether a current measure is appropriate or whether a measure of past exposure is needed. Reliability refers to the processes included in an epidemiological study to check repeatability of measurements of the exposures. These usually include intra---observer reliability and inter---observer reliability.

[D] Confounding has occurred where the estimated intervention exposure effect is biased by the presence of some difference between the comparison groups (apart from the exposure investigated/of interest). Typical confounders include baseline characteristics, prognostic factors, or concomitant exposures (e.g. smoking). A confounder is a difference between the comparison groups and it influences the direction of the study results. A high quality study at the level of cohort design will identify the potential confounders and measure them (where possible). This is difficult for studies where behavioral, attitudinal or lifestyle factors may impact on the results.

[E] Strategies to deal with effects of confounding factors may be dealt within the study design or in data analysis. By matching or stratifying sampling of participants, effects of confounding factors can be adjusted for. When dealing adjustment in data analysis, assess the statistics used in the study. Most will be some form of multivariate regression analysis to account for the confounding factors measured. Look out for description of statistical methods as regression methods such as logistic regression are usually employed to deal with confounding factors variables of interest.

[F] The participants should be free of the outcomes of interest at the start of the study. Refer to the ‘methods’ section in the paper for this information, which is usually found in descriptions of participant/sample recruitment, definitions of variables, and/or inclusion/exclusion criteria.

[G] Read the methods section of the paper. If for e.g. lung cancer is assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If lung cancer is 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. Having established the objectivity of the outcome measurement (e.g. lung cancer) instrument,

it’s important to establish how the measurement was conducted. Were those involved in collecting data trained or educated in the use of the instrument/s? (e.g. radiographers). 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?

[H] The appropriate length of time for follow---up will vary with the nature and characteristics of the population of interest and/or the intervention, disease or exposure. To estimate an appropriate duration of follow---up, read across multiple papers and take note of the range for duration of follow---up. The opinions of experts in clinical practice or clinical research may also assist in determining an appropriate duration of follow---up. For example, a longer timeframe may be needed to examine the association between occupational exposure to asbestos and the risk of lung cancer. It is important, particularly in cohort studies that follow up is long enough to enable the outcomes. However, it should be remembered that the research question and outcomes being examined would probably dictate the follow up time.

[I] It is important in a cohort study that a greater percentage of people are followed up. As a general guideline, at least 80% of patients should be followed up. Generally a dropout rate of 5% or less is considered insignificant. A rate of 20% or greater is considered to significantly impact on the validity of the study. However, in observational studies conducted over a lengthy period of time a higher dropout rate is to be expected. A decision on whether to include or exclude a study because of a high dropout rate is a matter of judgement based on the reasons why people dropped out, and whether dropout rates were comparable in the exposed and unexposed groups. Reporting of efforts to follow up participants that dropped out may be regarded as an indicator of a well conducted study. Look for clear and justifiable description of why people were left out, excluded, dropped out etc. If there is no clear description or a statement in this regards, this will be a 'no'.

[J] Some people may withdraw due to change employment or some may die; however, it is important that their outcomes are assessed. Selection bias may occur as a result of incomplete follow---up. Therefore, participants with unequal follow---up periods must be taken into account in the analysis, which should be adjusted to allow for differences in length of follow---up periods. This is usually done by calculating rates which use person---years at risk, i.e. considering time in the denominator.

[K] 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 of cohort studies should be detailed enough for reviewers to identify which analytical techniques were used (in particular, regression or stratification) and how specific confounders were measured. For studies utilising regression analysis, it is useful to identify if the study identified which variables were included and how they related to the outcome. If stratification was the analytical approach used, were the strata of analysis defined by the specified variables? 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.