Appendix C. Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures

This item bank is intended to evaluate the quality of studies examining the outcomes of interventions, treatments, or exposures. Eligible study designs include observational studies (cohort studies, case-control, case-series, and cross-sectional studies). It is not intended to rate the quality of studies concerning the accuracy of diagnostic tests. Abstractors can use the empty text box included with each question to document an explanation of their rating for later review. This may be particularly helpful in relation to a "cannot determine" response choice.

Study Definitions

Case series

Description: A study that describes a group of patients with a similar diagnosis and/or treatment. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of a condition through a collection of individual case reports.

Design features:

1. There is no comparison between groups to assess the effect/association of an intervention/exposure and an outcome.
2. There is no comparison with the same group over time.

Cross-sectional study

A study in which both the exposure and the outcome status in a target population are assessed concurrently that is, at the same point in time or during a brief period of time. The temporal sequence of cause and effect cannot necessarily be determined. They are most commonly used to assess prevalence. A common method for data collection is a survey.

Case control study

A study in which participants are selected based on the known outcome(s) of interest (e.g., disease, injury). Exposure status is then collected based on the participants' past experiences. Exposure status is compared between the two (or more) groups: those who have the outcome of interest and those who do not have the outcome of interest (controls). This is a retrospective study that collects data on events that have already occurred.

Cohort studies

A study in which individuals in the group without the outcome(s) of interest (e.g., disease) are classified according to exposure status (exposed or unexposed) and then are followed over time

to determine if the development of the outcome of interest is different in the exposed and unexposed groups.

Q1: Do the inclusion/exclusion criteria vary across the comparison groups of the study? *[PI: Drop question if not relevant to all included studies. To use this question for studies with one group, the focus of the question on comparison groups and related response categories would need to be changed to individuals.]*

PI:

Yes, varies ..... ☐

Partially: some, but not all criteria, applied to all groups or not clearly stated if some criteria are applied to all groups ..... ☐

No, does not vary ..... ☐

Cannot determine: article does not specify ..... ☐

Not applicable: study has only one group and so does not include comparison groups ..... ☐

Explanation for rating:

Q2: Does the strategy for recruiting participants into the study differ across groups? *[PIs: Drop question if not relevant to all included studies. To use this question for studies with one group, the focus of the question on comparison groups and related response categories would need to be changed to individuals.]*

PI:

Yes, differs ..... ☐

No, does not differ ..... ☐

Cannot determine ..... ☐

Not applicable: one study group ..... ☐

Explanation for rating:

Q3: Is the selection of the comparison group inappropriate, after taking into account feasibility and ethical considerations? *[PI: Provide instruction to the abstractor based on the type of study. Interventions with community components are likely to have contamination if all groups are drawn from the same community. Interventions without community components should select groups from the same source (e.g., community or hospital) to reduce baseline differences across groups. For case-control studies, controls should represent the population from which cases arose; that is, controls should have met the case definition if they had the outcome.]*

PI:

Yes, inappropriate.....	☐	***Explanation for rating:***
No, not inappropriate	☐	
Cannot determine or no description of the derivation of the comparison group.....	☐	
Not applicable: study does not include a comparison group (case series, one study group)	☐	

Q4: Does the study fail to account for important variations in the execution of the study from the proposed protocol? *[PI: Consider intensity, duration, frequency, route, setting, and timing of intervention/exposures. Drop if not relevant for body of literature.]*

PI:

Yes, fails to account	☐	***Explanation for rating:***
Partially, fails to account	☐	
No, does not fail to account	☐	
Cannot determine.....	☐	
Not applicable: not an intervention study or no variations	☐	

Q5: Was the outcome assessor not blinded to the intervention or exposure status of participants? *[PI: There may be circumstances where clinical evaluators cannot be blinded to exposure status. Drop if not relevant to the body of literature.]*

PI:

Yes, not blinded..... ☐

No, blinded..... ☐

Not applicable: assessor cannot be blinded ..... ☐

Explanation for rating:

Q6: Were valid and reliable measures, implemented consistently across all study participants used to assess inclusion/exclusion criteria, intervention/exposure outcomes, participant health benefits and harms, and confounding? *[PI: Important measures should be identified for abstractors and if there is more than one, they should be listed separately. PI may need to establish a threshold for what would constitute acceptable measures based on study topic. When subjective or objective measures could be collected, subjective measures based on self-report may be considered as being less reliable and valid than objective measures such as clinical reports and lab findings. Some characteristics may require that sources for establishing their validity and/or reliability be described or referenced. If so, provide instruction to abstractors.]*

PI:

Yes, valid and reliable measure used..... ☐

No, valid and reliable measure not used ..... ☐

Cannot determine or measurement approach not reported ..... ☐

Explanation for rating:

Q7: Was the length of follow-up different across study groups? *[Abstractor: When follow-up was the same for all study participants, the answer is no. If different lengths of follow-up were adjusted by statistical techniques, (e.g., survival analysis), the answer is no. Studies in which differences in follow-up were ignored should be answered yes.]*

PI:

Yes, different or cannot determine ..... ☐

No, not different or remedied through analysis ... ☐

Not applicable: cross-sectional or only one group followed over time..... ☐

Explanation for rating:

Q8: In cases of high loss to follow-up (or differential loss to follow-up), was the impact assessed (e.g., through sensitivity analysis or other adjustment method)? [PI: Attrition is measured in relation to the time between baseline (allocation in some instances) and outcome measurement for both retrospective and prospective studies and could include data loss from switching. Attrition rates may vary by outcome and time of measurement. Specify the criterion to meet relevant standards for the topic. Specify measurement period of interest, if repeated measures. Cochrane standard for attrition is 20 percent for shorter term (<1 year) and 30 percent for longer term (≥ 1 year).]

PI:

Yes, impact assessed ..... ☐

No, impact not assessed..... ☐

Cannot determine..... ☐

Not applicable: no loss to follow-up or loss to follow-up was not considered to be high, cross-sectional study, or case-control study selected on outcome..... ☐

Explanation for rating:

Q9: Are any important primary outcomes missing from the results? *[PI: Identify all primary outcomes that one would expect to be reported in the study, including timing of measurement.]*

PI:

Yes, important outcome(s) missing	☐	**Explanation for rating:**
No important outcome(s) missing.....	☐	
Cannot determine.....	☐	

Q10: Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results? *[PI: Identify all important harms that one would expect be reported in the study, including timing of measurement. Drop if not relevant to body of literature.]*

PI:

Yes, important outcomes missing.....	☐	**Explanation for rating:**
No important outcomes missing.....	☐	
Assessment of harms not applicable to this study.....	☐	

Q11: Are results believable taking study limitations into consideration? *[Abstractor: This question is intended to capture the overall quality of the study. Consider issues that may limit your ability to interpret the results of the study. Review responses to earlier questions for specific criteria.]*

PI:

Yes, believable ..... ☐

No, not believable ..... ☐

Explanation for rating:

Questions to Assess Confounding (Q6, Q12-13)

Q12: Any attempt to balance the allocation between the groups or match groups (e.g., through stratification, matching, propensity scores). *[PI: Drop if not relevant to the body of evidence.]*

PI:

Yes or study accounts for imbalance between groups through a post hoc approach such as multivariate analysis..... ☐

No or cannot determine ..... ☐

Not applicable: study does not include a comparison group (case series or one study group) ..... ☐

Explanation for rating:

Q13: Were important confounding variables not taken into account in the design and/or analysis (e.g., through matching, stratification, interaction terms, multivariate analysis, or other statistical adjustment such as instrumental variables)? [PI: Provide instruction to abstractors on known confounding variables and inadequate adjustment for confounding for each outcome.]

PI:

Yes, not accounted for or not identified..... ☐	**Explanation for rating:**
Partially: some variables taken into account or adjustment achieved to some extent..... ☐	
No: taken into account… ..... ☐	
Cannot determine..... ☐	

Modified from: Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol. 2012 Feb; 65(2):163-78. PMID: 21959223.