Last updated: 8 May 2021

CEECAT: Collaboration for Environmental Evidence Critical Appraisal Tool Version 0.2 (Prototype)

Contents

Part A: General Description of the Tool

We are currently developing a critical appraisal tool for evaluating 'risk of bias' (or internal validity) of primary studies assessing effectiveness of interventions or impacts of exposures in environmental management, and here we provide a second draft of the tool. The tool is still under development and requires initial testing, but it may help environmental evidence synthesists conduct critical appraisal. Application of the tool involves: (a) answering the checklist questions within individual risk-of-bias criteria (see below); (b) judging the risk of bias within individual risk-of-bias criteria; and (c) making an overall judgement about risk of bias for the study findings (estimate of effectiveness of intervention or impact of exposure).

Understanding Risk of Bias

Bias or inaccuracy is referred to as deviation from the truth [1] or a systematic error in study's findings (including inference that is made in the study) [2], and thus random errors (imprecisions) in results or inferences of impact or effectiveness will not be addressed in this tool. Having mentioned that however, systematic errors in real-world data cannot often be quantified, and thus distinguishing between systematic errors and random errors will not be possible unless true values are known [1]. This is why the concept of risk of bias (measure of internal validity) will be used to evaluate how susceptible studies are to bias [3]. Study findings will always have a risk of bias since it will not be possible to prove that there is no bias in the findings. Having a risk of bias does not mean that the findings are biased but there is always a possibility for the findings being biased to some extent. The spectrum of risk of bias we provide is categorical (low, medium, high) and thus risk of bias will not be quantified. Our critical appraisal tool enables qualitative assessment of risk of bias (internal validity).

Subject Scope

The tool is designed for environmental management research such as applied ecology, biodiversity conservation and conservation genetics, soil, water and air pollution, agriculture, park and protected area management, environmental epidemiology and pathogen control, species invasions, river and wetland management, exploitation of natural resources and fisheries, waste management, sustainable energy and consumption, and broader contexts of environmental sustainability may also be relevant if outcomes of interest are measured quantitatively.

In general, 'medical research involving human subjects' is beyond the scope of this tool. Such research often accords with the Declaration of Helsinki (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/) and provides ethics statement (name of ethics committee, date of approval and project ID). Although implications for environmental management may be provided in such research, they are beyond the scope of this tool. For example, the impact of exposure to pesticides on human urinary biomarkers may be of interest of U.S. Environmental Protection Agency. In such cases, we refer assessors to the risk of bias assessment tools developed in the healthcare sector (e.g., 'RoB 2' [4] or 'ROBINS-I' [5]).

Purely laboratory-based biological research (*in vivo* and *in vitro* experiments) is also beyond the scope of this tool. The checklist below may be used to check if the use of this tool is appropriate for a planned environmental evidence review (**Figure A1**). If any of the items in 'beyond the scope' applies, the planned review is not within the scope even when all of the items in 'within the scope' apply.

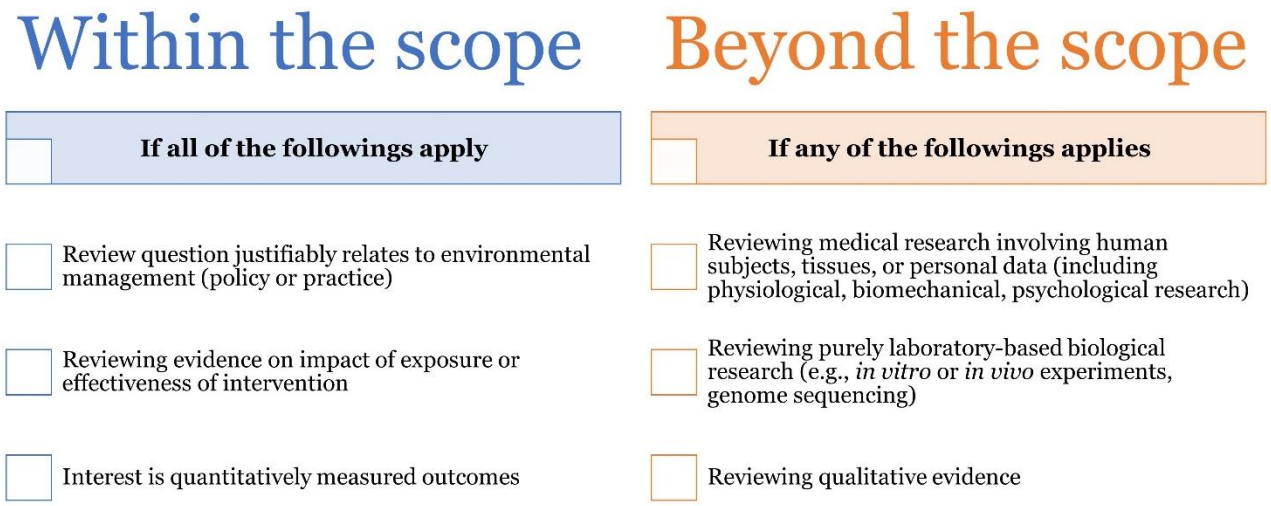

Figure A1. *Checklist for the scope of the tool. Note: If any of the conditions in 'beyond the scope' applies, a planned review is not within the scope even when all of the conditions in 'within the scope' apply.*

Individual Risk-of-Bias Criteria

There are eight risk-of-bias criteria in our tool:

1. Risk of confounding (bias due to uncontrolled variable (confounder or third variable) that influences both the intervention/exposure and the outcome)
2. Risk of selection bias (bias arising from systematic differences in the selection of subjects or areas into the study or analysis after intervention or exposure)
3. Risk of misclassification bias (bias due to misclassification of intervention, exposure and/or comparator of interest)
4. Risk of performance bias (bias due to altered treatment procedure of interest)
5. Risk of detection bias (bias arising from systematic differences in measurement of outcomes of interest)
6. Risk of attrition bias (bias due to systematic differences in missing data between intervention/exposure and comparator groups)
7. Risk of outcome reporting bias (bias in reporting of study findings)
8. Risk of analysis bias (bias due to error in applied statistical methods)

Criterion 1: Risk of Confounding

This criterion is concerned with bias due to uncontrolled variable (often referred to as confounder or third variable) that influences both the intervention/exposure and the outcome. An example of confounding is that if the effect of population density of salmonids on their survival to be studied, presence of predators may be a confounding factor as it may influence both survival and population density. Controlling for the presence of predators thus may reduce the risk of confounding in this instance [6]. Assessment of risk of confounding requires subject knowledge for determining potential confounding factors of the addressed causal relationship. Catalogue of Bias provides examples of confounding in medical sciences: https://catalogofbias.org/biases/confounding/.

Criterion 2: Risk of Selection Bias

This criterion is concerned with bias arising from systematic differences in the selection of subjects or areas into the study or analysis after intervention or exposure. An example of selection bias is that if the effect of insect herbivory on losses of woody plant foliage to be studied, haphazard leaf selection may be more likely to result in systematic differences in the selection of leaves compared to random or systematic selection, and thus may affect the estimate of effect [7]. Selection bias is sometimes categorised as a source of confounding if selection is made before intervention or exposure and there is no alteration of the selection after intervention or exposure. Confounding due to selection of subjects or areas can be dealt before intervention or exposure (e.g. stratified sampling in which the population of inference is divided into subpopulations [8]), and thus it needs

to be assessed in **Criterion 1** above. This criterion is thus designed to assess risk of selection bias after intervention or exposure and alteration of selection after intervention or exposure, and this borderline between confounding and selection bias is consistent with a widely applied risk of bias tool in the healthcare sector [5]. Catalogue of Bias provides broader descriptions of selection bias in medical sciences: https://catalogofbias.org/biases/selection-bias/.

Criterion 3: Risk of Misclassification Bias

This criterion is concerned with bias due to misclassification of intervention, exposure and/or comparator of interest. Definitions of intervention/exposure and comparator are necessary for replication and for avoiding misclassification [9]. For example, if the effect of pesticides on the mortality of honey bees (*Apis mellifera*) to be studied, defining what pesticides (compounds) are relevant and what count as exposure (e.g. compounds found in pollen, beeswax or honey bees) may be necessary. When certain individuals are classified in 'no exposure' group as a comparator, evidence of 'no exposure' and descriptions of what compounds are screened for in what samples (pollen, beeswax or honey bees) may be necessary because, for example, as the number of screened compounds increases, the number of detected compounds may increase [10]. If screened compounds are not extensive in 'no exposure' comparator group and researchers failed to detect pesticides residue in the 'no exposure' group (when there is residue), it will lead to misclassification and the effect may be biased towards null (no effect) in this instance as this comparison is actually 'exposure to pesticides' vs. 'exposure to pesticides'. However, note that if researchers are interested in specific compounds only, and there is supporting information of a correctly classified comparison, then the existence of other pesticides in samples, that are not of the focus, will be of issue of confounding (i.e. the influence of the existence of other pesticides will be a third variable) and thus this should be dealt in **Criterion 1** above. Catalogue of Bias provides examples of misclassification bias in medical sciences: https://catalogofbias.org/biases/misclassification-bias/.

Criterion 4: Risk of Performance Bias

This criterion is concerned with bias due to altered treatment procedure of interest or deviated initiation, implementation or discontinuation either by persons who apply experimental treatments or persons who receive the treatments if human subjects are to be studied. This criterion is only applicable to studies that apply experimental treatments. If no experimental treatments are applied (e.g. observational site comparison studies), this criterion will be 'not applicable'. In this tool, 'treatments' are defined as different procedures to be compared for measuring effects [9]. The term 'control' is often used loosely to denote 'no treatment' in scientific literature, however, it is technically a treatment (a procedure to be compared against) [9], and thus both intervention/exposure and control need to be assessed in this criterion. Performance bias may arise if treatments are altered either intentionally or unintentionally after the procedures have taken place (this is not necessarily initiation of intervention/exposure as the time lag between the start of procedure and the start of intervention may differ and the start of intervention may be recorded incorrectly) and if the alterations are not taken into account (e.g., altered procedures are not reflected on data collection sheet, or areas that have been applied altered procedures are not excluded). For example, if nitrogen fertiliser is applied more than initially planned and this deviated implementation is not recoded, a difference in crop yield (measure of effect) may be overestimated. In medical randomised controlled trials, it is recommended to apply double-blinding (both participants and healthcare provider) to avoid performance bias because knowledge of healthcare intervention or allocation may affect the procedures [5]. However, if there are no deviations from procedures of interest (or there is supporting information of 'no alterations') and procedures are successfully implemented for all subjects or areas without any influence of awareness, then awareness of the procedures alone will not affect an estimate of an effect from the view of bias due to altered treatment procedure (note: outcome measurers' awareness of details of study is addressed below in **Criterion 5**). Catalogue of Bias provides examples of performance bias in medical sciences: https://catalogofbias.org/biases/performance-bias/.

Criterion 5: Risk of Detection Bias

This criterion is concerned with bias arising from systematic differences in measurement of outcomes of interest. Systematic errors in measurement of outcomes may occur if outcome data are determined differently between groups, either intentionally (e.g. influence of desire to obtain a certain direction of effect) or unintentionally (e.g. due to cognitive bias or human errors). For

example, if different extractants (say water extractant and calcium chloride extractant) are used to obtain soil organic carbon concentration data (as outcome of interest) in different groups, the effect may be over- or underestimated due to the systematic difference in the process of measuring the outcomes [11]. When studying complex systems, and especially when many steps are involved in measuring outcomes, each calibration method or applied instrument may need to be the same between groups because, for example, devices used in a specific step may be biased and thus differently affect outcome data between groups [8]. Note if the same biased device is used to measure outcomes in both groups, the bias will cancel out, and it will not affect estimate of effectiveness or impact. For example, if true outcome measurements of dissolved organic carbon concentrations are 10 mg/L for both groups, and the biased outcome measurements are 20 mg/L for both groups, the differences between the groups are the same (0 mg/L), and hence there is no bias in the estimate of the effect in this case. Catalogue of Bias provides examples of detection bias in medical sciences: https://catalogofbias.org/biases/detection-bias/.

Criterion 6: Risk of Attrition Bias

This criterion is concerned with bias due to systematic differences in missing data between intervention/exposure and comparator groups. Attrition bias may occur if subjects or areas are missed unequally between groups, and missed subjects or areas have characteristics that are associated with intervention, exposure and/or outcome of interest. For example, if the effect of different light intensities on leaf lifespan of evergreen woody plants to be studied, missing data on individuals with certain characteristics, for example certain maturity or age of plants (i.e. with certain qualitative or quantitative values that describe how mature they are or how old they are), for one group may result in a systematic difference in characteristics of groups which may in turn result in over- or underestimation of the effect as life stage or age may be an interacting factor [12]. However, if percentages of missing data are the same (or nearly the same) and the collected data still allow a valid comparison (e.g. immature plants are missed for both groups but data on all mature plants are collected successfully for both groups), the missing data will not affect the estimate of effect from the view of bias due to systematic differences in missing data. 'Attrition bias' is sometimes categorised as a type of 'selection bias' [13], however in our tool, 'attrition' means not only loss of subjects or areas (equivalent of 'participants' in medical trials), but also loss of relevant data since unequal loss of data may occur even after data are successfully collected from all subjects or areas included in a study (e.g. during copying data from the original record to a form for data analysis) [8,9]. This division is consistent with a widely applied risk-of-bias tool in the healthcare sector [5]. Catalogue of Bias provides examples of attrition bias in medical sciences: https://catalogofbias.org/biases/attrition-bias/.

Criterion 7: Risk of Outcome Reporting Bias

This criterion is concerned with bias in reporting of study findings. Outcome reporting bias may arise if findings are selectively disclosed in reporting that does not reflect actual findings. Selective disclosure may appear at three different levels [4]:

- Selective disclosure of findings from multiple measurements
- Selective disclosure of findings from multiple subgroups or subpopulations
- Selective disclosure of findings from multiple analyses

In every way, it is theoretically possible that findings are selectively disclosed to report statistically significant results and/or interesting or desired results of the parties or individuals who involved in the research. Catalogue of Bias provides examples of outcome reporting bias in medical sciences: https://catalogofbias.org/biases/outcome-reporting-bias/.

Criterion 8: Risk of Analysis Bias

This criterion is concerned with bias due to error in applied statistical methods. There is currently no such criterion (domain) in widely applied risk-of-bias assessment tools in the healthcare sector (RoB 2 [4] and ROBINS-I [5]). Steenland et al. argued that there is no dedicated criterion for assessing the ability of obtaining unbiased results of inferential statistics, including appropriateness of chosen statistical methods, in four risk-of-bias assessment tools in health sciences. They thus suggested that this kind of criterion should preferably be added in the risk-of-

bias assessment tools [14]. We agree with the suggestion and so established a dedicated criterion in our tool and call this criterion 'risk of analysis bias'. The James Lind Library describes analysis bias as 'biases can be introduced when knowledge of the results of studies influences analysis and reporting decisions, for example, when studies stop earlier than planned, or with biased selection of the treatment outcomes measured' (https://www.jameslindlibrary.org/research-topics/biases/analysis-bias/ accessed on 4 May 2021). The description is rather broad, and thus we refer analysis bias more narrowly in our tool as bias due to errors in applied analyses, and applied analyses specifically mean applied statistical methods in our tool. The checklist questions ask about three specific issues:

- Errors in applied descriptive statistics (e.g. miscalculations of sample sizes, means)
- Errors in applied inferential statistics (including null hypothesis testing, estimation, coding)
- Violation of assumption for the applied inferential statistics and appropriateness of the applied statistical methods (e.g. criteria for normality and equal variances are not satisfied, inappropriate choice of statistical tests)

It is possible that all the other criteria (1–7) are considered to be low risks of bias but there are errors in applied descriptive or inferential statistics. If there is any error at this stage, it is very likely that the outcome of a study (measure of effectiveness or impact, or inference) will be changed (no matter how trivial or substantial it is), and thus the algorithm suggests high risk of bias when there is any error to reflect that concern. There may also be a case where no inferential statistics (e.g. hypothesis testing) is conducted in a study for specific findings of assessors' interest. For example, a meta-analysis review is being carried out and the assessors are only interested in specific comparable raw data that are only a part of the study (and thus there is no dedicated hypothesis testing for the findings of interest). In such case, assessors may select 'not applicable' for the second and third bullet points above, and when there is no error in descriptive statistics, the default algorithm suggests medium risk of bias to indicate the concern of no formal conduct of inferential statistics (e.g. null hypothesis testing, estimation of effect).

Checklist Questions in Risk-of-Bias Criteria

The tool provides multiple checklist questions which are designed to help judgement about risk of bias within each risk-of-bias criterion. Assessors are required to answer all checklist questions in 'general' category, as well as in 'conditional' category if conditions are met. There is also an 'optional' category for each risk-of-bias criterion to allow assessors to communicate the magnitude and direction of potential bias and/or to consider results of quantitative assessment of risk of bias (e.g. through simulations) if conducted. If assessors feel that optional checklist questions are worth answering, they can answer them or otherwise 'skip' them. The response options for the majority of checklist questions are fixed. These are:

- Yes (Y);
- Seemingly yes (SY);
- Seemingly no (SN);
- No (N);
- Unclear; and
- Not applicable (NA)

When 'unclear' is selected, the default algorithm suggests assessors a certain direction (Y/SY or N/SN). In this default algorithm, selecting an 'unclear' response equates to selecting a response that suggests higher risk of bias (i.e. higher risk of bias is favoured when 'unclear' is selected). When conditions are not met, assessors can select 'not applicable'. Responses to all checklist questions will be needed for judging risk of bias in the eight individual risk-of-bias criteria, as well as for judging overall risk of bias for the study findings, so please make sure to record your responses as you go.

Risk-of-Bias Judgement within Risk-of-Bias Criteria

Once assessors have responded to all checklist questions within a risk-of-bias criterion, they will have to judge a risk of bias for the criterion. The levels of risk of bias can be selected from the following:

- Low risk of bias (Low)
- Medium risk of bias (Med)
- High risk of bias (High)

The process of making a judgement about risk of bias is rather straightforward with the default setting. We will provide a roadmap-like diagram (algorithm) for suggested judgement in each risk-of-bias criterion. If an optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias may be upgraded (e.g. from high to medium) or downgraded (e.g. medium to high) from the suggested judgement, depending on result of quantitative assessment. We suggest that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias.

Overall Risk of Bias Judgement

Once assessors have judged all risk-of-bias criteria for a study's findings, they can make an overall judgement about risk of bias. The same levels of risk of bias as above will be used as follows:

- Overall low risk of bias: a study is considered to have low risk of bias for all risk-of-bias criteria for the findings
- Overall medium risk of bias: a study is considered to have medium risk of bias in at least one risk-of-bias criterion, but not to have high risk of bias for any risk-of-bias-criteria for the findings
- Overall high risk of bias: a study is considered to have high risk of bias in at least one risk-of-bias criterion for the findings

Please note this three-level classification of overall risk of bias may be too simple for some evidence syntheses because it does not take into account the frequencies of low and medium risk of bias within an overall medium risk of bias. For example, one study's findings are judged to have low risk of bias for seven criteria and medium risk of bias for one criterion (this will result in an overall medium risk of bias) will be the same as another study's findings that are judged to have medium risk of bias for the eight risk-of-bias criteria (this will also result in an overall medium risk of bias). Thus, recording and communicating the extent of risk of bias within an overall medium risk of bias may be useful.

If some or all of the optional checklist questions are answered through quantitative assessment of risk of bias (e.g. through simulations), some adjustment of overall judgement of risk of bias may be applied. In such case, we recommend assessors to note detailed rationale and justification for drawing an overall conclusion of risk of bias for the study findings so that transparent communication of risk of bias for the study findings will be ensured.

Customisation (Optional)

Users may wish to customise the tool based on detailed rationale and justification, so we provide some potential customisation.

Regarding the process of making a judgement about risk of bias within each risk-of-bias criterion, the default algorithm is the recommended setting. However, assessors can change the algorithm if they so wish. For example, in the default setting, selecting a 'unclear' response equates to selecting a response that directs to higher risk of bias (i.e. higher risk of bias is favoured when 'unclear' is selected). If assessors do not feel this setting is right, they can customise.

Regarding making an overall judgement about risk of bias, some levels of risk of bias can be modified or added. For example, it is possible for assessors to divide overall medium risk of bias

into multiple levels using the frequencies of low and medium risk of bias for better communicating the risk of bias (e.g. splitting into medium-low and medium-high). The same logic may apply to an overall high risk of bias if assessors would like to investigate the high risk of bias further (e.g. dividing an overall high risk of bias into multiple levels).

Tutorial

In the current prototype version, assessors will normally need one copy of the Excel or portable document format (PDF) file (**Part B** of this document) for each study's estimate of effectiveness of intervention or impact of exposure. For example, if a review team has 10 separate estimates of effectiveness of intervention or impact of exposure, they will need 10 copies of the Excel or **Part B** of this PDF file to record their responses.

The PDF File

It should be straightforward to apply the tool so please follow the instructions provided below in **Part B** of this PDF file. How to record assessors' decisions is entirely up to them. They can work with digital or hard copies. They may use an Excel sheet called 'All_decisions' described below to record or merge all decisions.

The Excel File

The checklist questions for individual risk-of-bias criteria are provided in each of the 'Criterion' sheets. In the individual Criterion sheets (i.e. from 'Criterion1' to 'Criterion8'), there is a drop-down list (shaded in grey) for each checklist question so assessors can just select a response (**Figure A2**).

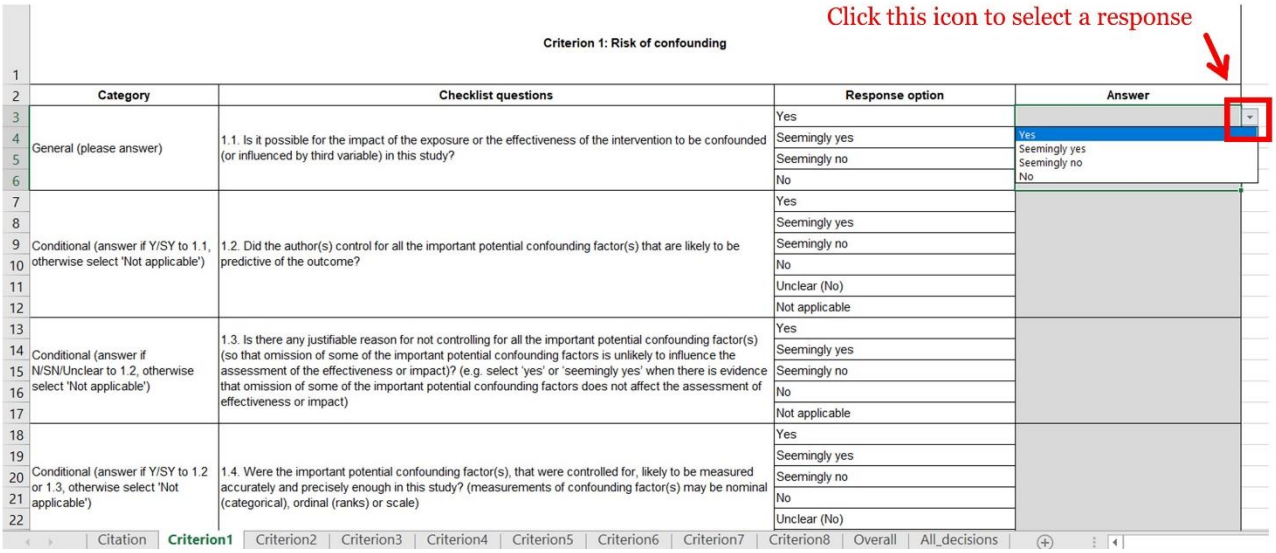

Figure A2. *How to answer a checklist question in the Excel version.*

Once assessors have answered the checklist questions, they will need to use the algorithm provided in the sheet to make a judgement about risk of bias for the risk-of-bias criterion. There is also a drop-down list for this, so assessors can just select the appropriate risk of bias (**Figure A3**).

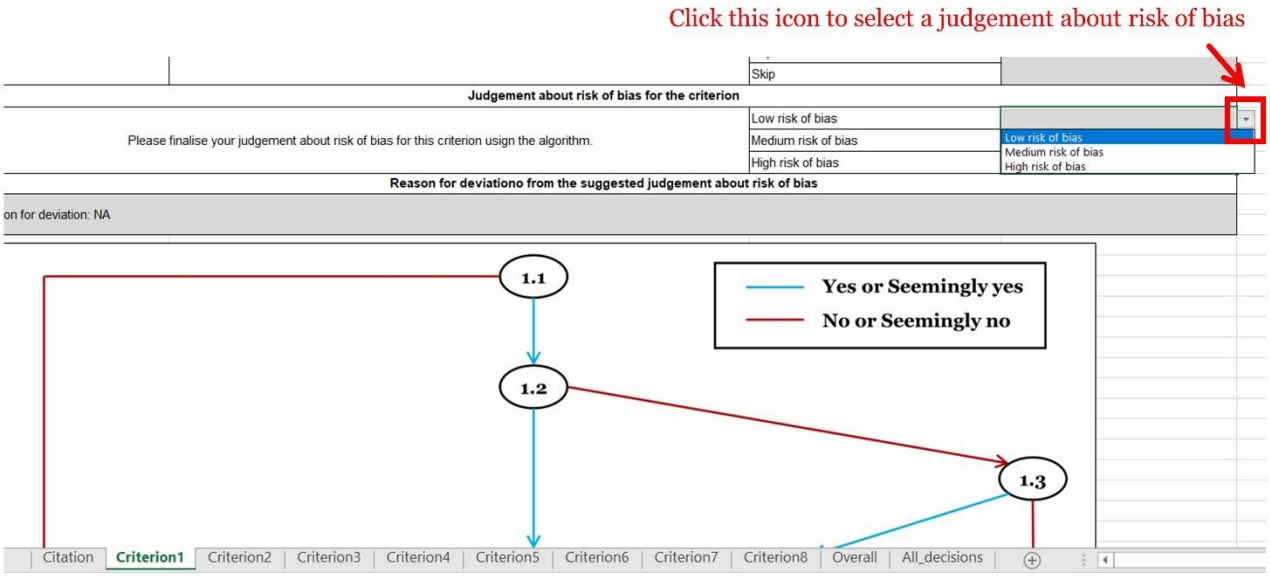

Figure A3. *How to select a judgement about risk of bias within a risk-of-bias criterion in the Excel version.*

Once assessors have judged risk of bias for all risk-of-bias criteria (from 'Criterion1' to 'Criterion8'), they can make an overall judgement about risk of bias for the study findings in the 'Overall' sheet. Judgement about risk of bias for individual risk-of-bias criteria should automatically appear in the corresponding cells shaded in grey (**Figure A4**). The criteria are also provided in the sheet so making an overall judgement about risk of bias should be straightforward. In the same way as selecting a judgement about risk of bias within a risk-of-bias criterion, assessors can select overall risk of bias using a drop-down list shaded in grey.

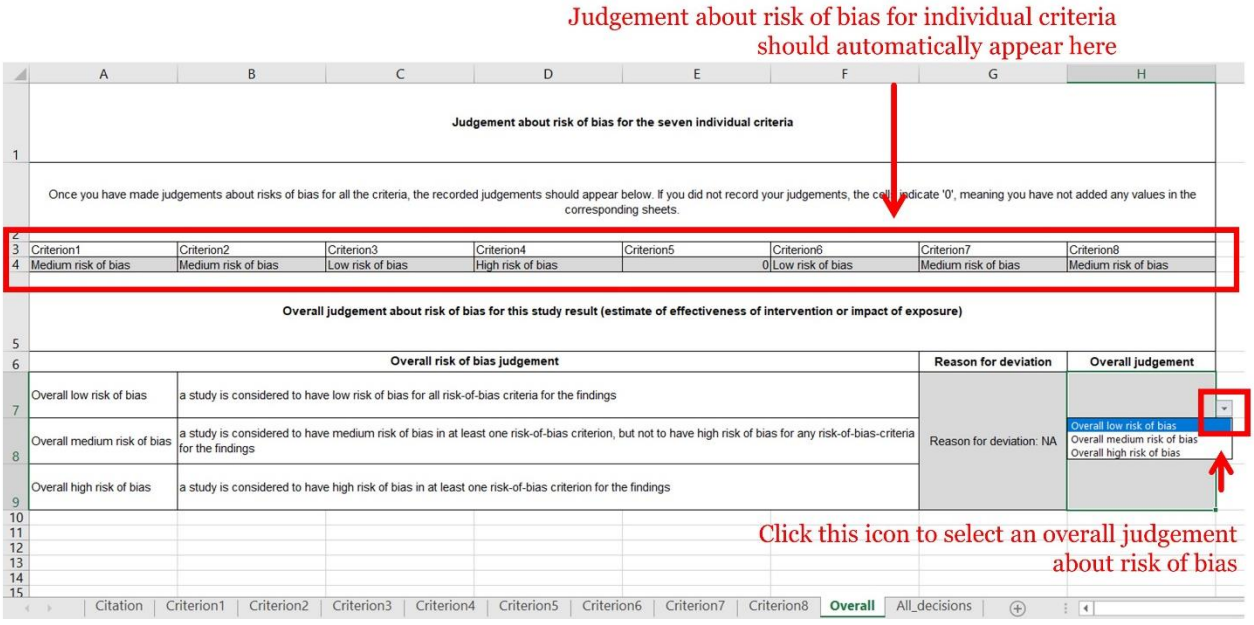

Figure A4. *How to make an overall judgement about risk of bias in the Excel version.*

Once assessors have answered all checklist questions, and have made judgements about risk of bias within individual risk-of-bias criteria and overall risk of bias, all decisions should automatically appear in one row (shaded in grey) in the 'All_decisions' sheet. Assessors should thus be able to merge their decisions on all included studies with ease (e.g. pasting the row to your data extraction sheet).

Part B: Applying the Tool

Criterion 1: Risk of Confounding

This criterion is concerned with bias due to uncontrolled variable (confounder or third variable) that influences both the intervention/exposure and the outcome.

Answering the Checklist Questions

Please answer the checklist questions in **Table B1** and record your responses.

Table B1. *Checklist questions for risk of confounding.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**1.1.** Is it possible for the impact of the exposure or the effectiveness of the intervention to be confounded (or influenced by third variable) in this study?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No
Conditional (answer if Y/SY to 1.1, otherwise select 'Not applicable')	**1.2.** Did the author(s) control for all the important potential confounding factor(s) that are likely to be predictive of the outcome?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No) ☐ Not applicable
Conditional (answer if N/SN/Unclear to 1.2, otherwise select 'Not applicable')	**1.3.** Is there any justifiable reason for not controlling for all the important potential confounding factor(s) (so that omission of some of the important potential confounding factors is unlikely to influence the assessment of the effectiveness or impact)? (e.g. select 'yes' or 'seemingly yes' when there is evidence that omission of some of the important potential confounding factors does not affect the assessment of effectiveness or impact)	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Not applicable
Conditional (answer if Y/SY to 1.2 or 1.3, otherwise select 'Not applicable')	**1.4.** Were the important potential confounding factor(s), that were controlled for, likely to be measured accurately and precisely enough in this study? (measurements of confounding factor(s) may be nominal (categorical), ordinal (ranks) or scale)	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No) ☐ Not applicable
Conditional (answer if you have answered 1.4, otherwise select 'Not applicable')	**1.5.** Did the author(s) analyse the effect appropriately by taking into account the important potential confounding factors, as well as the issue of accuracy and precision of the measurements of the potential confounding factor(s)?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No) ☐ Not applicable
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist	**1.6.** What are the predicted magnitude and the direction of bias due to confounding? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	☐ Intervention or exposure intolerably favoured * ☐ Intervention or exposure tolerably favoured ** ☐ Comparator intolerably favoured * ☐ Comparator tolerably favoured ** ☐ Intolerably towards no effect * ☐ Tolerably towards no effect ** ☐ Intolerably away from no effect *

question if they feel unfeasible)		☐ Tolerably away from no effect ** ☐ Unpredictable ☐ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B1**) to finalise your judgement about risk of bias for this criterion.

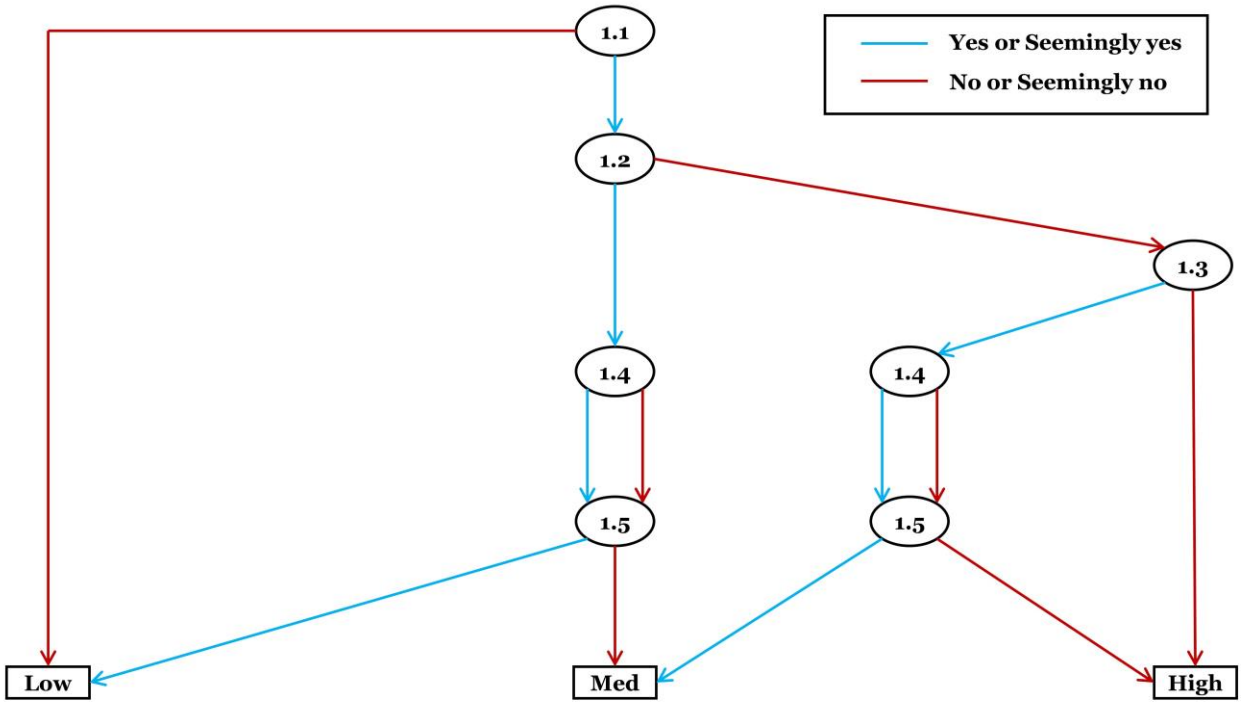

Figure B1. *Roadmap diagram for making judgement about risk of confounding. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B1** below.

Box B1. *Judgement about risk of confounding.*

☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
☐ High risk of bias (reason for deviation from the suggested judgement: _____)
Quantitative prediction of magnitude of bias (if available): _____

Criterion 2: Risk of Selection Bias

This criterion is concerned with bias arising from systematic differences in the selection of subjects or areas into the study or analysis after intervention or exposure.

Answering the Checklist Questions

Please answer the checklist questions in **Table B2** and record your responses.

Table B2. *Checklist questions for risk of selection bias.*

Category	Checklist Questions	Answer (Tick One Applies)

General (please answer whichever suitable)	**2.1.a.** Was the selection of subjects or areas into the study (or analysis) random or systematic (i.e. random or systematic sampling) and the selection was not altered # afterwards? (This applies when an attempt was not made to collect data of the entire, or nearly entire, inference population) OR **2.1.b.** Was the entire, or nearly entire, population of inference included in the study (or analysis), and the selection was not altered # afterwards? (This applies when an attempt was made to collect data of the entire, or nearly entire, inference population)	□ Answer 2.1.a □ Answer 2.1.b □ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
General (please answer whichever suitable)	**2.2.a.** Was/were the researcher(s) unaware (or blinded) of the selection until subjects or areas were assigned to or grouped as intervention/exposure and comparator, and the selection was not altered # afterwards? (when the study conducted the selection or grouping) OR **2.2.b.** Had the selection or grouping taken place before the conduct of the study (so that the researcher(s) were not involved in the selection or grouping at all) and the selection was not altered # afterwards?	□ Answer 2.2.a □ Answer 2.2.b □ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
General (please answer)	**2.3.** Did the start of follow-up (when specific individuals or particular areas, such as plots, who had received intervention/exposure or control treatments were followed up) or the start of post-intervention or post-exposure selection of subjects or areas (when subjects or areas were selected after intervention or exposure; this might be the case if individuals or areas were not specifiable so that follow-up to the same individuals or exact areas was not applicable due to the nature of study) coincide for most subjects or areas? (If only one subject or area was studied, select 'yes' or 'seemingly yes')	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
Conditional (answer if N/SN/Unclear to 2.1, 2.2 or 2.3, otherwise select 'Not applicable')	**2.4.** Was the follow-up or the selection of subjects or areas after intervention or exposure likely to be associated with any characteristics of the subjects or areas (that have different qualitative or quantitative values for different entities; e.g. pH, temperature, subspecies, age, sex, etc.)?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if Y/SY/Unclear to 2.4, otherwise select 'Not applicable')	**2.5.** Were the characteristics of the subjects or areas, that are likely to be associated with the follow-up or the selection of subjects or areas after intervention or exposure, also likely to be related to status of the intervention or exposure (or correlated with intervention or exposure variable(s); e.g. intervention intensity)?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if Y/SY/Unclear to 2.5, otherwise select 'Not applicable')	**2.6.** Were the characteristics of the subjects or areas, that are likely to be associated with the follow-up or the selection of subjects or areas after intervention or exposure, also likely to be predictive of the outcome?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if N/SN/Unclear to 2.1, 2.2 or 2.3, otherwise select 'Not applicable')	**2.7.** Did the author(s) estimate the potential bias(es) (systematic error(s)) due to haphazard or altered selection, or unequal timing of selection of subjects or areas into the study, and found that the bias(es) was/were minimal? Alternatively, is there evidence that the effect is measured accurately enough in relation to bias arising from selection of subjects or areas into the study or analysis after intervention or exposure?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No) □ Not applicable

Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**2.8.** What are the predicted magnitude and the direction of bias arising from systematic differences in the selection of subjects or areas into the study or analysis after intervention or exposure? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	☐ Intervention or exposure intolerably favoured * ☐ Intervention or exposure tolerably favoured ** ☐ Comparator intolerably favoured * ☐ Comparator tolerably favoured ** ☐ Intolerably towards no effect * ☐ Tolerably towards no effect ** ☐ Intolerably away from no effect * ☐ Tolerably away from no effect ** ☐ Unpredictable ☐ Skip

*# Assessors may accept slight alteration (e.g. exclusion of subjects may be considered as 'no alteration' if it is acceptable). * Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B2**) to finalise your judgement about risk of bias for this criterion.

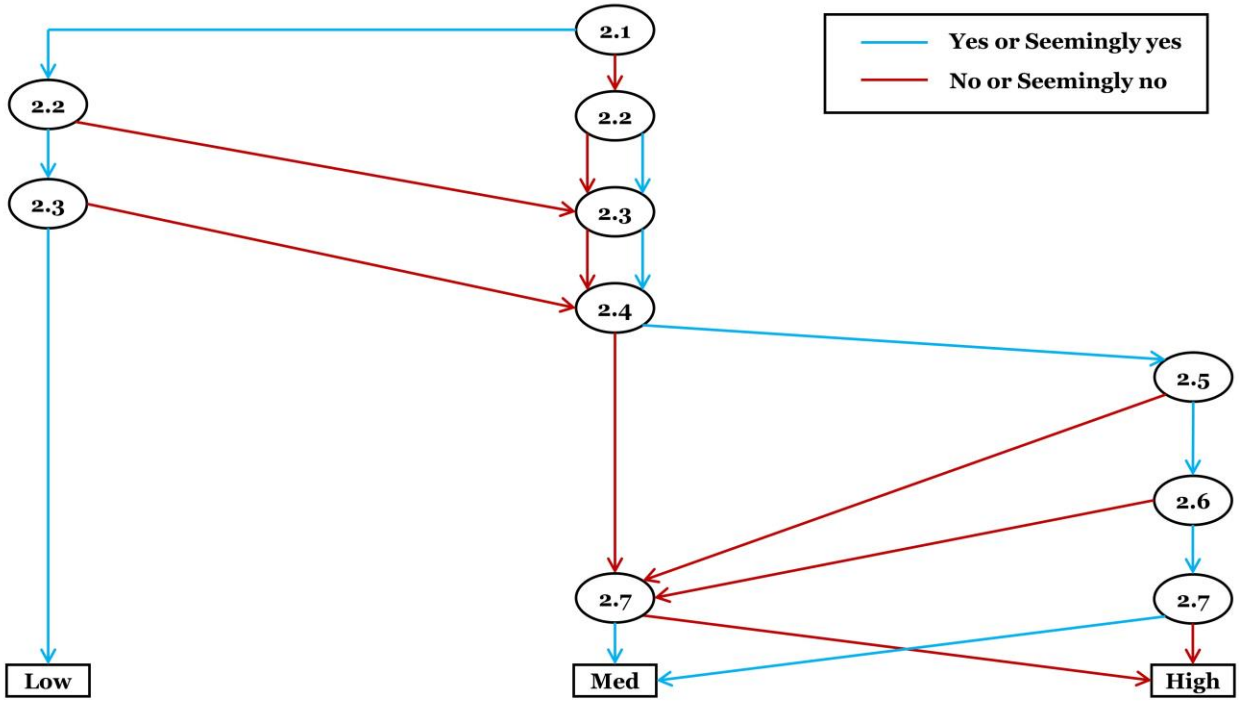

Figure B2. *Roadmap diagram for making judgement about risk of selection bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B2** below.

Box B2. *Judgement about risk of selection bias.*

☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)

□ High risk of bias (reason for deviation from the suggested judgement: _____) Quantitative prediction of magnitude of bias (if available): _____

Criterion 3: Risk of Misclassification Bias

This criterion is concerned with bias due to misclassification of intervention, exposure and/or comparator of interest.

Answering the Checklist Questions

Please answer the checklist questions in **Table B3** and record your responses.

Table B3. *Checklist questions for risk of misclassification bias.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**3.1.** Were the intervention or exposure (group) and the comparator (group) of interest clearly defined and the definitions are appropriate for measuring the effect of intervention or exposure (so that readers can understand, and the definitions allow the effect to be measured accurately and precisely enough)? (select 'no' or 'seemingly no' when classification is poorly defined and/or the choice of measure of exposure or intervention is unlikely to be accurate or precise enough, for example, when the use of an imprecise or inaccurate biomarker is defined as exposure)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
General (please answer)	**3.2.** Were the intervention or exposure (group) and comparator (group) pre-defined (i.e. before initiation of treatment or analysis) and there were no alterations # of the definitions afterwards? (e.g. select 'yes' or 'seemingly yes' when there is a study or analysis protocol defining the intervention or exposure as well as comparator, and there are no alterations # of the definitions afterwards)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
General (please answer)	**3.3.** Might classification of exposure, intervention or comparator (group) have been incorrect due to influence of some knowledge, experience or desire? (e.g. intentional misclassification of exposure to yield a desired outcome; unintentional misclassification due to prior knowledge or cognitive bias)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes)
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**3.4.** What are the predicted magnitude and the direction of bias due to misclassification of intervention, exposure and/or comparator of interest? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	□ Intervention or exposure intolerably favoured * □ Intervention or exposure tolerably favoured ** □ Comparator intolerably favoured * □ Comparator tolerably favoured ** □ Intolerably towards no effect * □ Tolerably towards no effect ** □ Intolerably away from no effect * □ Tolerably away from no effect ** □ Unpredictable □ Skip

*# Assessors may accept slight alteration (e.g. minor alteration of definition may be considered as 'no alteration' if it is acceptable). * Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B3**) to finalise your judgement about risk of bias for this criterion.

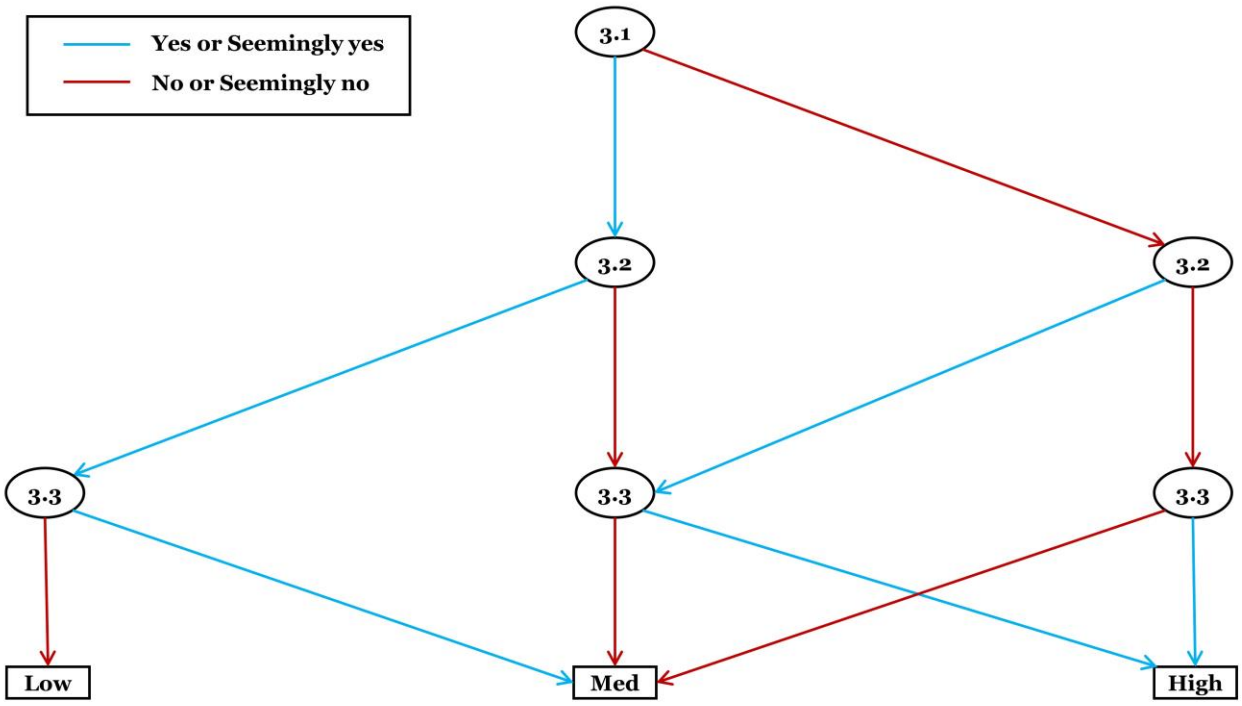

Figure B3. *Roadmap diagram for making judgement about risk of misclassification bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B3** below.

Box B3. *Judgement about risk of misclassification bias.*

- ☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ High risk of bias (reason for deviation from the suggested judgement: _____)

Quantitative prediction of magnitude of bias (if available): _____

Criterion 4: Risk of Performance Bias

This criterion is concerned with bias due to altered treatment procedure of interest.

Answering the Checklist Questions

Please answer the checklist questions in **Table B4** and record your responses.

Table B4. *Checklist questions for risk of performance bias.*

Category	Checklist Questions	Answer (Tick One Applies)
Conditional (answer if experimental treatments are applied in the study, otherwise select 'not applicable')	**4.1.** Were any of the persons, who applied treatments (intervention, exposure, alternative intervention, alternative exposure, or control), aware of the hypothesis that was being tested or the comparison that was being made to measure impact or effectiveness?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (Yes) ☐ Not applicable

Conditional (answer if experimental treatments are applied in the study, otherwise select 'not applicable')	**4.2.** Were there any alterations of intervention/exposure or control treatments of interest that might have an impact on the effectiveness of the intervention or the impact of the exposure? (e.g. deviated initiation, implementation and/or discontinuation)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if Y/SY/Unclear to 4.2, otherwise select 'Not applicable')	**4.3.** Were these deviated treatments unbalanced between intervention or exposure groups (when comparing two interventions or exposures; i.e. intervention of interest or exposure vs. alternative one), or inaccurately taken into account (when comparing intervention or exposure vs. control (no intervention or exposure); e.g. nitrogen fertilizer was mistakenly applied more than initially planned for one group but this deviation is not reflected on the data collection sheet, i.e., not occurred as recorded), and thus it might have influenced the measure of impact or effectiveness?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if experimental treatments are applied in the study, otherwise select 'not applicable')	**4.4.** Were both exposure/intervention and comparator treatments initiated and implemented as intended (or occurred as recorded) for all or nearly all subjects or areas?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No) □ Not applicable
Conditional (answer if Y/SY/Unclear to 4.2, or N/SN/Unclear to 4.4, otherwise select 'Not applicable')	**4.5.** Are the used analysis methods of the impact of the exposure or the effectiveness of the intervention appropriate in relation to bias due to altered treatment procedure of interest?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No) □ Not applicable
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible. Select 'not applicable' if no experimental treatments are applied)	**4.6.** What are the predicted magnitude and the direction of bias due to altered treatment procedure of interest? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	□ Intervention or exposure intolerably favoured * □ Intervention or exposure tolerably favoured ** □ Comparator intolerably favoured * □ Comparator tolerably favoured ** □ Intolerably towards no effect * □ Tolerably towards no effect ** □ Intolerably away from no effect * □ Tolerably away from no effect ** □ Unpredictable □ Not applicable □ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B4**) to finalise your judgement about risk of bias for this criterion.

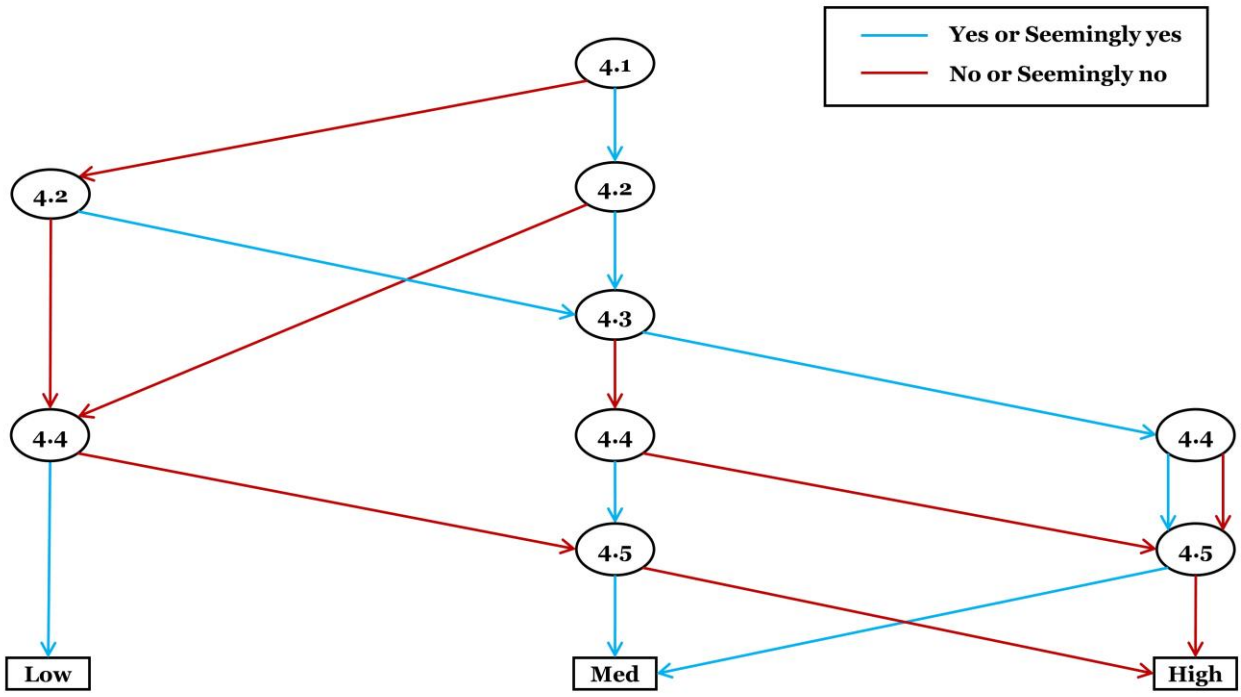

Figure B4. *Roadmap diagram for making judgement about risk of performance bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B4** below.

Box B4. *Judgement about risk of performance bias.*

- ☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ High risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ Not applicable (there are no experimental treatments applied in the study)

Quantitative prediction of magnitude of bias (if available): _____

Criterion 5: Risk of Detection Bias

This criterion is concerned with bias arising from systematic differences in measurement of outcomes of interest.

Answering the Checklist Questions

Please answer the checklist questions in **Table B5** and record your responses.

Table B5. *Checklist questions for risk of detection bias.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**5.1.** Was there any way for the outcome measure to be affected by knowledge of the exposure, intervention, subjects or areas, or desire for certain outcome (e.g. data collectors, who measured the outcome, were aware of the details of the study)?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (Yes)

General (please answer)	**5.2.** Was/were the person(s), who assessed the effectiveness of the intervention or the impact of the exposure, aware of the exposure or intervention received by subjects or areas?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes)
General (please answer)	**5.3.** Were the methods for measuring and analysing the outcome data the same across the groups?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (No)
General (please answer)	**5.4.** By looking at the available materials for this study, was/were any bias(es) in outcome measurement that was/were related to (or correlated with) the assessed exposure or intervention?	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes)
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**5.5.** What are the predicted magnitude and the direction of bias arising from systematic differences in measurement of outcomes of interest? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	□ Intervention or exposure intolerably favoured * □ Intervention or exposure tolerably favoured ** □ Comparator intolerably favoured * □ Comparator tolerably favoured ** □ Intolerably towards no effect * □ Tolerably towards no effect ** □ Intolerably away from no effect * □ Tolerably away from no effect ** □ Unpredictable □ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B5**) to finalise your judgement about risk of bias for this criterion.

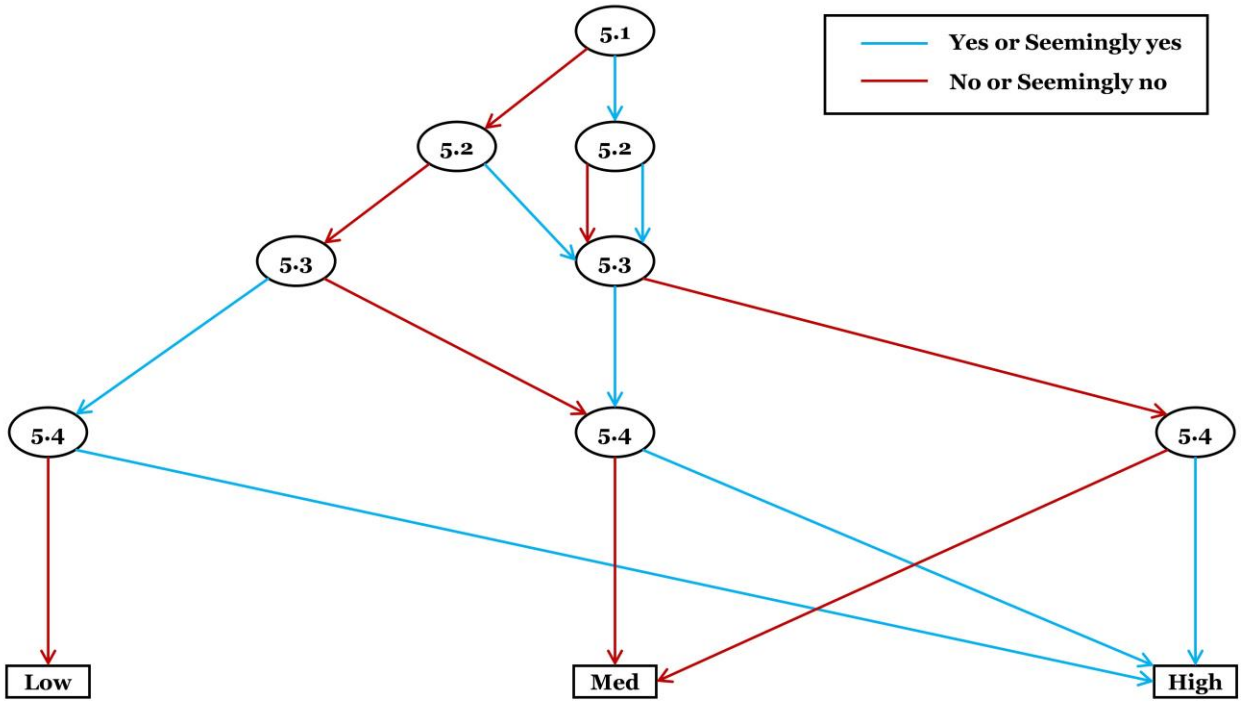

Figure B5. *Roadmap diagram for making judgement about risk of detection bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B5** below.

Box B5. *Judgement about risk of detection bias.*

☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
☐ High risk of bias (reason for deviation from the suggested judgement: _____)
Quantitative prediction of magnitude of bias (if available): _____

Criterion 6: Risk of Attrition Bias

This criterion is concerned with bias due to systematic differences in missing data between intervention/exposure and comparator groups.

Answering the Checklist Questions

Please answer the checklist questions in **Table B6** and record your responses.

Table B6. *Checklist questions for risk of attrition bias.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**6.1.** For the analysis/analyses of impact or effectiveness, were relevant outcome data available for all the included subjects or areas (or for nearly entire samples that could be equivalent of all included subjects or areas)?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No)
General (please answer)	**6.2.** Were any subjects or areas excluded after the start of the study because of missing some relevant data that was/were required for the analysis/analyses of the impact or effectiveness?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (Yes)

Conditional (answer whichever suitable if N/SN/Unclear to 6.1, or Y/SY/Unclear to 6.2, otherwise select 'Not applicable')	**6.3.a.** Are the percentages of missing data the same, or almost the same, across the groups? (This applies when an attempt was not made to collect data of the entire, or nearly entire, inference population) OR **6.3.b.** Are subjects or areas still representative of the population of inference? (This applies when an attempt was made to collect data of the entire, or nearly entire, inference population)	☐ Answer 6.3.a ☐ Answer 6.3.b ☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No) ☐ Not applicable
Conditional (answer if N/SN/Unclear to 6.1, or Y/SY/Unclear to 6.2, otherwise select 'Not applicable')	**6.4.** Is there any evidence that missing data has no or minimal (e.g. statistically non-significant) impact on the study findings (effectiveness of intervention or impact of exposure)?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (No) ☐ Not applicable
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**6.5.** What are the predicted magnitude and the direction of bias due to systematic differences in missing data between intervention/exposure and comparator groups? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	☐ Intervention or exposure intolerably favoured * ☐ Intervention or exposure tolerably favoured ** ☐ Comparator intolerably favoured * ☐ Comparator tolerably favoured ** ☐ Intolerably towards no effect * ☐ Tolerably towards no effect ** ☐ Intolerably away from no effect * ☐ Tolerably away from no effect ** ☐ Unpredictable ☐ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B6**) to finalise your judgement about risk of bias for this criterion.

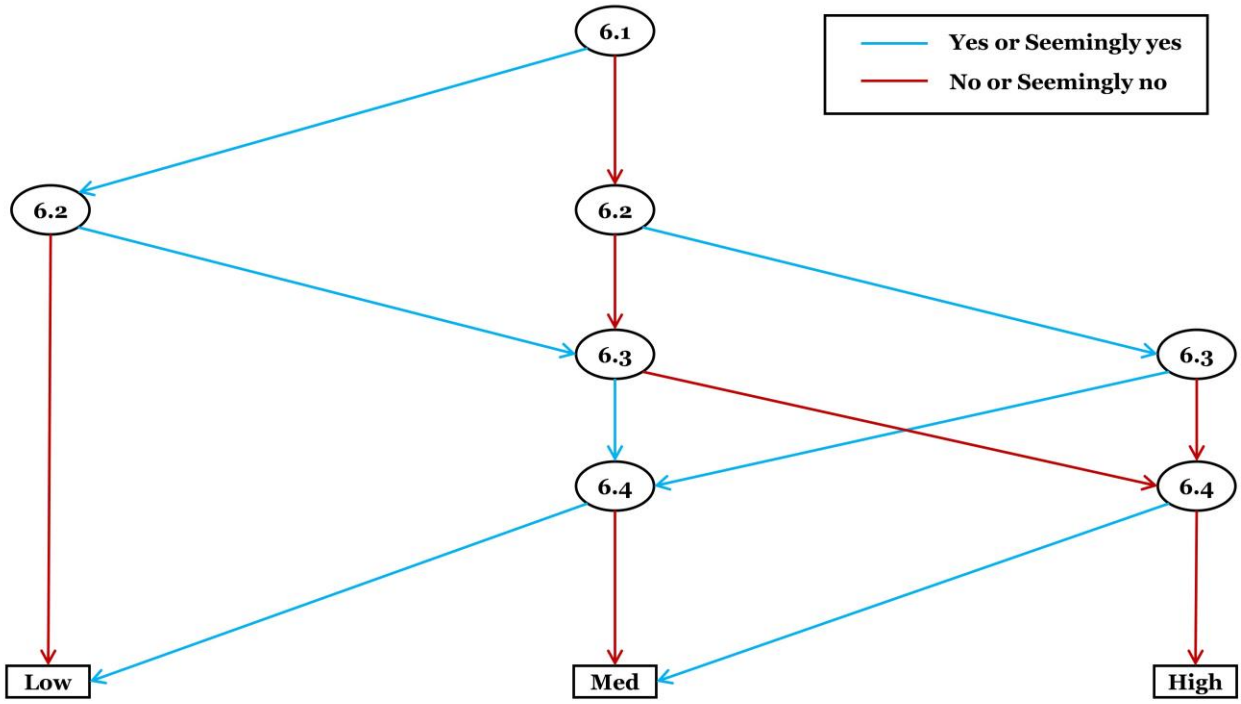

Figure B6. *Roadmap diagram for making judgement about risk of attrition bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B6** below.

Box B6. *Judgement about risk of attrition bias.*

- ☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
- ☐ High risk of bias (reason for deviation from the suggested judgement: _____)

Quantitative prediction of magnitude of bias (if available): _____

Criterion 7: Risk of Outcome Reporting Bias

This criterion is concerned with bias in reporting of study findings.

Answering the Checklist Questions

Please answer the checklist questions in **Table B7** and record your responses.

Table B7. *Checklist questions for risk of outcome reporting bias.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**7.1.** Are the reported relevant outcome data (or effect estimate) likely to be of (or based on) selected measurements of the outcome?	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (Yes)
General (please answer)	**7.2.** Are relevant outcome data likely to be unreported for some subgroup(s)? (i.e. only outcome data on certain subjects or areas with certain characteristic(s) (e.g. taxonomic group) or in certain conditions (e.g. intervention intensity) are available)	☐ Yes ☐ Seemingly yes ☐ Seemingly no ☐ No ☐ Unclear (Yes)
General (please answer)	**7.3.** Is/are the analysis/analyses of the causal relationship of interest (intervention-outcome or exposure-outcome) likely to	☐ Yes ☐ Seemingly yes

	be partially reported? (i.e. there is/are other relevant analysis/analyses of the causal relationship that is/are not reported)	☐ Seemingly no ☐ No ☐ Unclear (Yes)
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**7.4.** What are the predicted magnitude and the direction of bias in reporting of study findings? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	☐ Intervention or exposure intolerably favoured * ☐ Intervention or exposure tolerably favoured ** ☐ Comparator intolerably favoured * ☐ Comparator tolerably favoured ** ☐ Intolerably towards no effect * ☐ Tolerably towards no effect ** ☐ Intolerably away from no effect * ☐ Tolerably away from no effect ** ☐ Unpredictable ☐ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B7**) to finalise your judgement about risk of bias for this criterion.

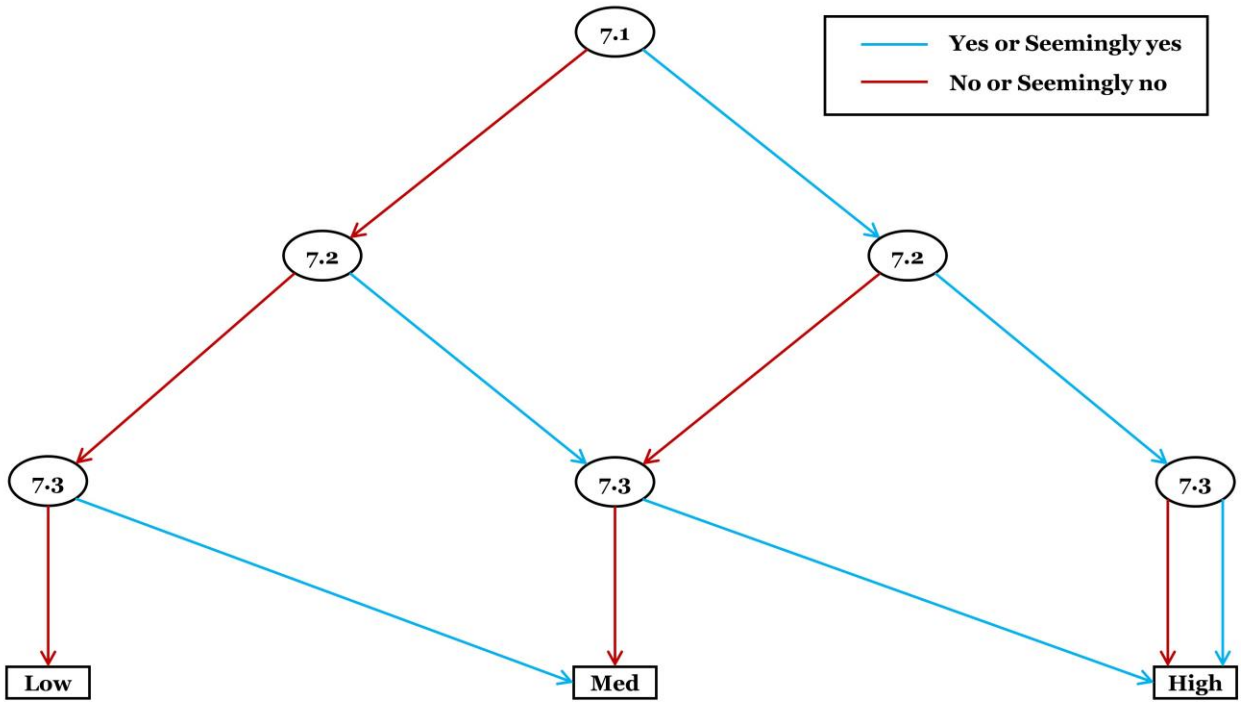

Figure B7. *Roadmap diagram for making judgement about risk of outcome reporting bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B7** below.

Box B7. *Judgement about risk of outcome reporting bias.*

□ Low risk of bias (reason for deviation from the suggested judgement: _____) □ Medium risk of bias (reason for deviation from the suggested judgement: _____) □ High risk of bias (reason for deviation from the suggested judgement: _____) Quantitative prediction of magnitude of bias (if available): _____

Criterion 8: Risk of Analysis Bias

This criterion is concerned with bias due to error in applied statistical methods.

Answering the Checklist Questions

Please answer the checklist questions in **Table B8** and record your responses.

Table B8. *Checklist questions for risk of analysis bias.*

Category	Checklist Questions	Answer (Tick One Applies)
General (please answer)	**8.1.** Is it likely that there is/are error(s) in the applied descriptive statistics? (e.g. miscalculations of sample sizes, means, medians, variances, ranges for intervention/exposure and comparator groups)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes)
Conditional (answer if inferential statistics are applied, otherwise select 'not applicable')	**8.2.** Is it likely that there is/are error(s) in the applied inferential statistics (including null hypothesis testing, estimation, coding)? (e.g. null hypothesis and alternative hypothesis mislabelled, miscalculations of differences between intervention/exposure and comparator, errors in coding, etc.)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Conditional (answer if inferential statistics are applied, otherwise select 'not applicable')	**8.3.** Were assumptions for the applied inferential statistics violated or the applied inferential statistics methods inappropriate? (e.g. normality not assumed when conducting a parametric test, equal or unequal variances not tested when testing for a difference, no justification for the choice of dependent and independent variables, a Pearson's correlation test was used when analysing a causal relationship, inappropriate comparison of multiple models to support the provided statement when some of the models do not relate to impact or effectiveness, inappropriate modelling which may affect an estimate of effectiveness or impact)	□ Yes □ Seemingly yes □ Seemingly no □ No □ Unclear (Yes) □ Not applicable
Optional (It is suggested that detailed rationale or empirical evidence be provided when predicting magnitude and direction of bias. Assessors may skip this optional checklist question if they feel unfeasible)	**8.4.** What are the predicted magnitude and the direction of bias due to error in applied statistical methods? (Note: quantitative assessment (e.g. through simulation) may be conducted to predict the magnitude and direction of bias for this study result)	□ Intervention or exposure intolerably favoured * □ Intervention or exposure tolerably favoured ** □ Comparator intolerably favoured * □ Comparator tolerably favoured ** □ Intolerably towards no effect * □ Tolerably towards no effect ** □ Intolerably away from no effect * □ Tolerably away from no effect ** □ Unpredictable □ Skip

** Intolerable means that the study result should not be considered as valid enough in relation to the predicted magnitude of bias. ** Tolerable means that the study result could be considered as valid enough in relation to the predicted magnitude of bias.*

Once you have answered the checklist questions, please use the diagram below (**Figure B8**) to finalise your judgement about risk of bias for this criterion.

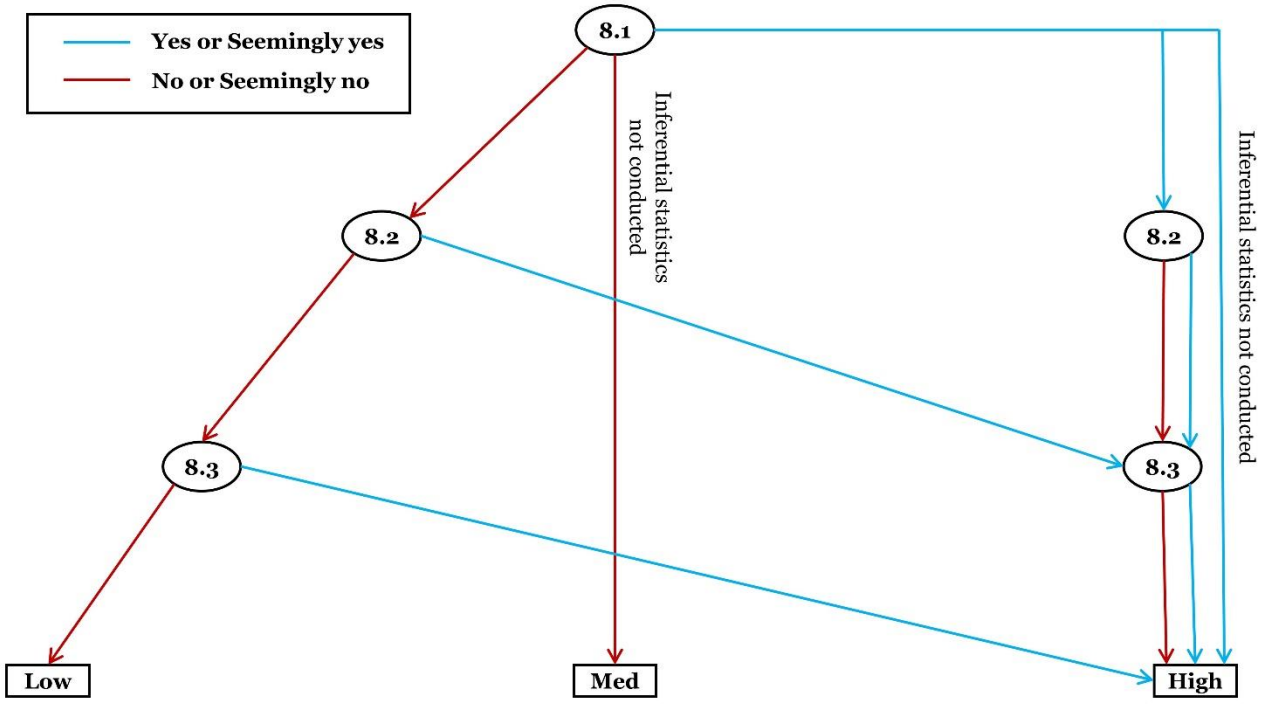

Figure B8. *Roadmap diagram for making judgement about risk of analysis bias. Note: if the optional question has been answered through quantitative assessment (e.g. through simulation), assessor's judgement about risk of bias for this criterion may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment.*

Please record your judgement about risk of bias for this criterion using **Box B8** below.

Box B8. *Judgement about risk of analysis bias.*

☐ Low risk of bias (reason for deviation from the suggested judgement: _____)
☐ Medium risk of bias (reason for deviation from the suggested judgement: _____)
☐ High risk of bias (reason for deviation from the suggested judgement: _____)
Quantitative prediction of magnitude of bias (if available): _____

Making an Overall Judgement about Risk of Bias for the Study Result

Once you have judged risk of bias for all criteria, please make an overall judgement about risk of bias for this study result using **Box B8** below. If some or all of the optional questions have been answered through quantitative assessment (e.g. through simulation), assessor's overall judgement about risk of bias may be upgraded or downgraded from the suggested judgement, depending on result of quantitative assessment. Also, if there is a justifiable reason for upgrading or downgrading overall judgement even when no quantitative assessments are conducted, assessors may deviate from the suggested judgement and record the reason.

Box B8. *Overall judgement about risk of bias for the study result.*

☐ Overall low risk of bias: a study is considered to have low risk of bias for all risk-of-bias criteria for the findings (reason for deviation from the suggested judgement: _____)

- ☐ Overall medium risk of bias: a study is considered to have medium risk of bias in at least one risk-of-bias criterion, but not to have high risk of bias for any risk-of-bias-criteria for the findings (reason for deviation from the suggested judgement: _____)
- ☐ Overall high risk of bias: a study is considered to have high risk of bias in at least one risk-of-bias criterion for the findings (reason for deviation from the suggested judgement: _____)

Quantitative prediction of magnitude of bias (if available): _____

References

1. Eisenhart, C. Expression of the uncertainties of final results. *Science* **1968**, *160*, 1201–1204, doi:10.1126/science.160.3833.1201.

2. Higgins, J.; Thomas, J.; Chandler, J.; Cumpston, M.; Li, T.; Page, M.; Welch, V. *Cochrane Handbook for Systematic Reviews of Interventions version 6.0*; Cochrane, 2019;

3. CEE Guidelines and Standards for Evidence Synthesis in Environmental Management VERSION 5.0. 2018. Available online: http://www.environmentalevidence.org/information-for-authors.

4. Sterne, J.A.C.; Savović, J.; Page, M.J.; Elbers, R.G.; Blencowe, N.S.; Boutron, I.; Cates, C.J.; Cheng, H.Y.; Corbett, M.S.; Eldridge, S.M.; et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ* **2019**, *366*, l4898, doi:10.1136/bmj.l4898.

5. Sterne, J.A.; Hernán, M.A.; Reeves, B.C.; Savović, J.; Berkman, N.D.; Viswanathan, M.; Henry, D.; Altman, D.G.; Ansari, M.T.; Boutron, I.; et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **2016**, *355*, i4919, doi:10.1136/bmj.i4919.

6. Matte, J.M.; Fraser, D.J.; Grant, J.W.A. Density-dependent growth and survival in salmonids: Quantifying biological mechanisms and methodological biases. *Fish Fish.* **2020**, *21*, 588–600, doi:10.1111/faf.12448.

7. Zvereva, E.L.; Kozlov, M. V. Biases in studies of spatial patterns in insect herbivory. *Ecol. Monogr.* **2019**, *89*, e01361, doi:10.1002/ecm.1361.

8. Cochran, W.G. *Sampling techniques*; Third Edit.; John Wiley & Sons: New York, USA, 1977.

9. Cochran, W.; Cox, G. *Experimental designs*; 2nd Editio.; John Wiley & Sons: New York, USA, 1957.

10. Toselli, G.; Sgolastra, F. Seek and you shall find: An assessment of the influence of the analytical methodologies on pesticide occurrences in honey bee-collected pollen with a systematic review. *Chemosphere* **2020**, *258*, 127358, doi:10.1016/j.chemosphere.2020.127358.

11. Li, S.; Zheng, X.; Liu, C.; Yao, Z.; Zhang, W.; Han, S. Influences of observation method, season, soil depth, land use and management practice on soil dissolvable organic carbon concentrations: A meta-analysis. *Sci. Total Environ.* **2018**, *631–632*, 105–114, doi:10.1016/J.SCITOTENV.2018.02.238.

12. Konno, K.; Akasaka, M.; Koshida, C.; Katayama, N.; Osada, N.; Spake, R.; Amano, T. Ignoring non-English-language studies may bias ecological meta-analyses. *Ecol. Evol.* **2020**, *10*, 6373–6384, doi:10.1002/ece3.6368.

13. Nunan, D.; Aronson, J.; Bankhead, C. Catalogue of bias: attrition bias. *BMJ evidence-based Med.* **2018**, *23*, 21–22, doi:10.1136/ebmed-2017-110883.

14. Steenland, K.; Schubauer-Berigan, M.K.; Vermeulen, R.; Lunn, R.M.; Straif, K.; Zahm, S.; Stewart, P.; Arroyave, W.D.; Mehta, S.S.; Pearce, N. Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: Strengths and limitations. *Environ. Health Perspect.* **2020**, *128*, 095002, doi:10.1289/EHP6980.