The Risk Of Bias In Non-randomized Studies – of Interventions, Version 2 (ROBINS-I V2) assessment tool (for follow-up studies)

20 November 2025

ROBINS-I V2 development group: Jonathan Sterne, Maya Brandt Mathur, Roy Elbers, Asbjørn Hróbjartsson, Alexandra McAleenan, Barney Reeves, Ian Shrier, Kate Tilling, Rebecca Armstrong, Nancy Berkman, Isabelle Boutron, James Carpenter, An-Wen Chan, Jon Deeks, Su Golder, David Henry, Peter Jüni, Jamie Kirkham, Menelaos Konstantinidis, Toby Lasserson, Yoon Loke, Luke McGuinness, Matthew Page, Jelena Savović, Bev Shea, Yoon Loke, David Mawdsley, Sasha Shepperd, Peter Tugwell, Jeff Valentine, Meera Viswanathan, Hugh Sharma Waddington, George Wells, Miguel Hernán, Julian Higgins

Contents

Outline of ROBINS-I V2

ROBINS-I aims to assess the risk of bias in a specific result from a non-randomized study that examines the effect of an intervention strategy, versus a comparator strategy, on an outcome. This document describes the ROBINS-I V2 tool for **follow-up (cohort) studies**. Throughout ROBINS-I V2, we refer to the "intervention strategy" and "comparatorstrategy" that are being compared. The comparator may be an alternative active intervention, standard care, or no intervention. Because there is no randomized assignment to the intervention or comparator, assignment is inferred from the intervention prescribed to or received by each participant. We refer to the participants classified into each strategy as the "intervention group" and "comparator group".

Risk-of-bias assessments should relate to risk of **material bias** rather than risk of any bias. Material bias should be interpreted as bias sufficient to cause an important change to the magnitude of the estimated effect, compared with the true value.

Before undertaking a ROBINS-I assessment (or series of assessments, e.g. in the context of a systematic review), users of the tool should specify the important confounding factors that are likely to influence the association between the intervention and the outcome (see section "At planning stage").

A key feature of the ROBINS-I approach is the specification, for each study, of the causal effect estimated by the result under consideration through specification of a hypothetical 'target trial'. This is essential for assessment of risk of bias, because the causal effect defines the result that would be seen (other than the impact of sampling variation) in the absence of bias.

If multiple assessors will implement ROBINS-I independently, the *Preliminary considerations* should be agreed between all assessors before each assessor works individually through evaluation of the confounding factors and bias domains.

ROBINS I V2 includes six domains of bias:

- Domain 1: Risk of bias due to confounding
- Domain 2: Risk of bias in classification of interventions
- Domain 3: Risk of bias in selection of participants into the study (or into the analysis)
- Domain 4: Risk of bias due to missing data
- Domain 5: Risk of bias arising from measurement of the outcome
- Domain 6: Risk of bias in selection of the reported result

Unlike version 1 of ROBINS-I. and RoB 2, ROBINS-I V2 does not include the domain "Risk of bias due to deviations from the intended intervention" (this is renamed as "Risk of bias due to protocol deviations" in the next version of RoB 2). In non-randomized studies, issues related to deviations from the intended intervention (protocol deviations) are relevant for analyses that estimate the 'per protocol' effect (see Guidance note 8). Bias arising in such analyses is addressed in Domain 1 (Variant B), through consideration of whether both baseline and time-varying confounding were adequately controlled for, using an appropriate analysis.

Each bias domain in ROBINS-I is addressed using a series of **signalling questions** that aim to gather important information about the study and the analysis being assessed. Most signalling questions have response options 'Yes', 'Probably yes', 'Probably no', 'No' and 'No information', with 'Yes' and 'Probably yes' having the same implications for risk of bias and similarly for 'No' and 'Probably no'. Some questions have additional response options (a 'weak' and a 'strong' version of 'Yes' or 'No') to help discriminate between higher and lower risk of bias. After the relevant signalling questions have been completed, an algorithm maps the answers to the signalling questions onto a proposed
judgement about **risk of bias** in the result that arises from this domain. The judgements and their broad interpretations are as follows.

Judgement	**Interpretation**
*Low risk of bias**	There is little or no concern about bias with regard to this domain.
Moderate risk of bias	There is some concern about bias with regard to this domain, although it is not clear that there is an important risk of bias.
Serious risk of bias	The study has some important problems in this domain: characteristics of the study give rise to a serious risk of bias.
Critical risk of bias	The study is very problematic in this domain: characteristics of the study give rise to a critical risk of bias, such that the result should

	generally be excluded from evidence syntheses.

*For Domain 1 (Risk of bias due to confounding), this is referred to as “Low risk of bias (except for concerns about uncontrolled confounding)”, in which confounding is very well addressed but cannot be eliminated as a possibility. This is because a risk of bias due to uncontrolled confounding cannot be excluded in an observational study.

ROBINS-I is intended to provide a framework for making informed and reasonable judgements about risk of material bias in studies of the effects of an intervention strategy (versus a comparator strategy) on an outcome. On occasion, answers to the signalling questions may not yield an appropriate risk of bias judgement based on the algorithm. Therefore, suggested risk of bias judgements produced by the algorithms can be overridden, in which case justification should be provided. We aim for transparency and reasonableness rather than mechanistic adherence to every word of the tool’s contents.

Optionally, a **predicted direction of bias** may be selected, balancing the various issues addressed within the domain. Response options for this depend on the type of bias being addressed.

After completing all six bias domains, an **overall judgement** is made for the risk of bias (and optionally for the predicted direction of any bias). The risk-of-bias Judgement is derived from the domain-level judgements using an algorithm. As for bias domain-level judgements, justification should be provided when the overall judgement suggested by the algorithm is overridden.

An online implementation of ROBINS-I V2 including automatic selection of relevant signalling questions and algorithm-derived risk-of-bias judgements is being developed and will be made available via www.riskofbias.info.

Index to guidance notes

The ROBINS-I V2 tool

At planning stage: list confounding factors

P1. List the important confounding factors relevant to all or most studies on this topic. Specify whether these are particular to specific intervention-outcome combinations.

Guidance note 1: listing confounding factors

A confounding factor is a prognostic factor that predicts whether participants receive the intervention or the comparator strategy. Important confounding factors are those that have the potential to introduce material bias into an estimated effect. Factors that are expected to have only very weak associations with the intervention or with the outcome, such that failure to account for them in the analysis will not have a material impact on the estimated effect of intervention on outcome, need not be considered here. Important confounding factors should be pre-specified at the planning stage, for example in the protocol of a systematic review that will include studies of the effects of interventions. The identification of potential confounding factors requires content knowledge and may usefully be informed by examination of relevant literature. Important confounding factors should be specified at the level of the broad research question (e.g. using a single list of confounding factors for a systematic review). This broad question may cover several specific interventions and/or outcomes. If confounding factors are specific to particular intervention-outcome combinations, then this should be stated.

Note that ROBINS-I does not address effect modification (the situation in which the effect of an intervention is different in different subsets of the population under study).

Mouse strain, sex, age, weight/general health, baseline MPTP-induced PD severity and baseline motor performance; MPTP dose/timing and injection consistency, nanoparticle batch/dose/timing; cage/litter/order/handling effects; exclusions/dropouts after model induction. Specific to the in vivo intervention-outcome comparison.

For each study result: preliminary considerations (parts A to D)

Guidance note 2: overview of preliminary considerations

The start point for a ROBINS-I assessment of an individual study is to specify the result from the study that is being assessed for risk of bias. The preliminary considerations questions should be answered in relation to the specific result that is being evaluated for the current ROBINS-I assessment.

In case of multiple alternative analyses being presented, it is important to specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Some characteristics of a study or a result may lead directly to the result being at critical risk of bias, and so make detailed risk-of-bias assessments unnecessary. A series of 'screening' questions in this section aim to identify such situations.

The **target trial specific to the study** is a hypothetical randomized trial, which need not be ethical or feasible, that compares the health effects of the same intervention and comparator strategies, conducted with the same eligibility criteria as the non-randomized study. In general, such target trials will not use blinding of participants or of health professionals administering interventions.

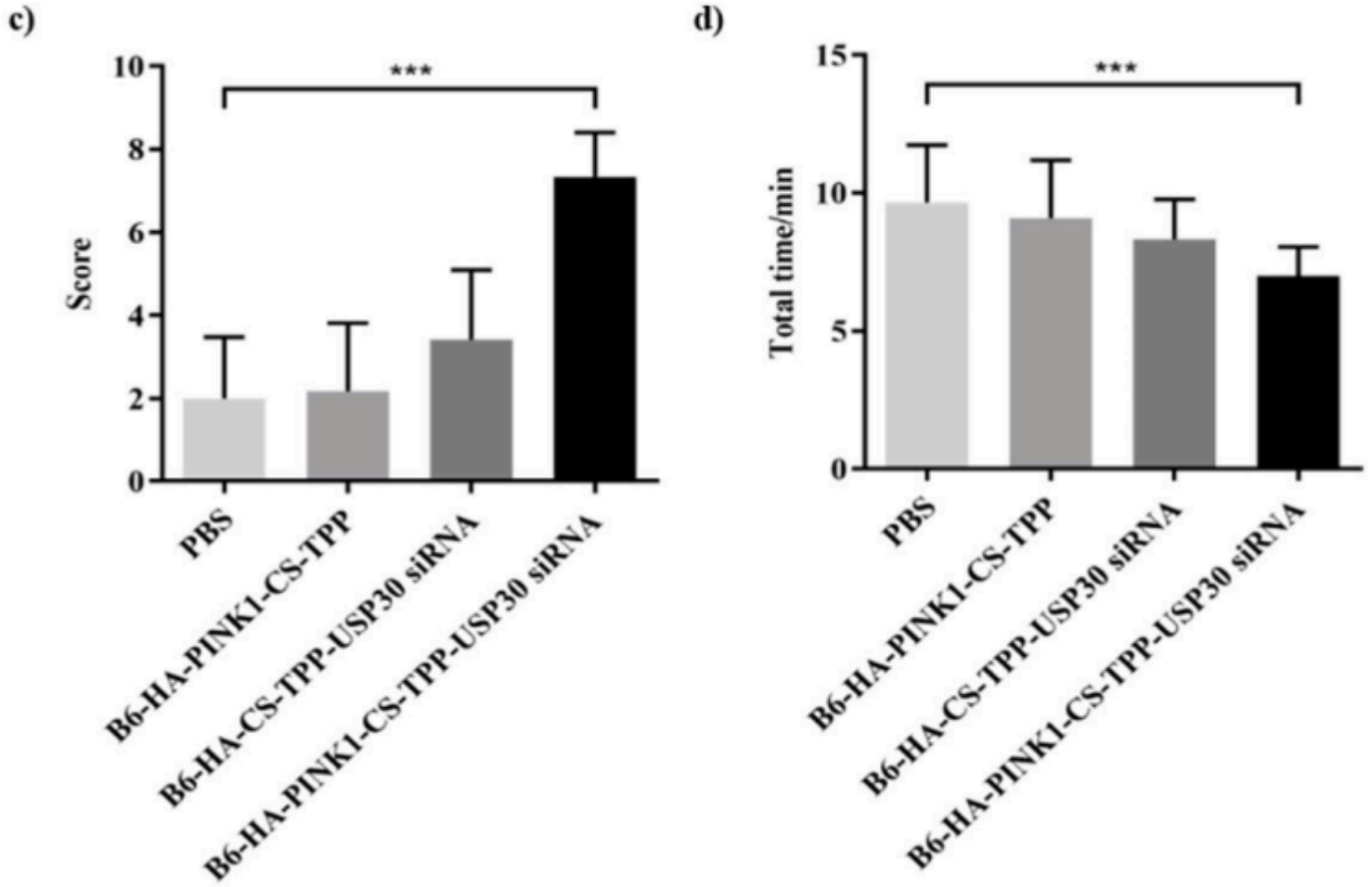
If multiple assessors will implement ROBINS-I independently, the questions in this section should be agreed between all assessors before each assessor works individually through the risk-of-bias assessment itself.

A. Specify the result being assessed for risk of bias

Guidance note 3: specifying the numerical result being assessed for risk of bias

A ROBINS-I assessment of risk of bias is specific to a particular study result. This is because different results from the same study may be at importantly different risks of bias (consider, for example, an unadjusted estimate of intervention effect compared with an estimate that is adjusted for numerous important confounding factors). Consequently, it may be necessary to undertake several ROBINS-I assessments of different results from the same study. If the study presents multiple alternative analyses, specify the numerical result (e.g. RR=1.52 (95% CI 0.83 to 2.77)) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

A1. Specify the numerical result being assessed



B6 peptide-HA-PINK1 antibody-CS-TPP-USP30 siRNA NPs vs PBS in MPTP-induced PD mice. Hanging score increased and rod-climbing time decreased; reported n = 12; statistical significance indicated in the caption ($P<0.05$), but exact means/CI/effect sizes are not given in text.

A2. Provide further details about this result (for example, location in the study report, reason it was chosen) [optional]

Chosen because it is the main in vivo therapeutic efficacy outcome most relevant to PD treatment, rather than purely nanoparticle characterization or in vivo

mechanism.

A3. Specify the outcome to which this result relates

Motor function/behavior after treatment, represented by hanging-test score and rod-climbing time, measured on day 7 after first treatment injection.

B. Decide whether to proceed with a risk-of-bias assessment

Guidance note 4: deciding whether to proceed with a risk-of-bias assessment

Some characteristics of a study or a result may lead directly to the result being at critical risk of bias, and so make detailed risk-of-bias assessments unnecessary. The questions in this section aim to identify such situations.

Question	Comments	Response options
B1 Did the authors make any attempt to control for confounding in the result being assessed?	Standardized key design factors, but no randomization, baseline equivalence or adjusted analysis is reported	PN
B2 If N/PN to B1: Is there sufficient potential for confounding that this result should not be considered further?	Potential confounding is important but not so severe that the animal experiment is uninterpretable; standardized experimental conditions partially limit bias.	PN

B3 Was the method of measuring the outcome inappropriate?	Hanging and rod-climbing tests are standard motor-behavior assays for mouse PD models, though blinding/reporting concerns remain.	N

If the answer to either B2 or B3 is 'Yes' or 'Probably yes', the result should be considered to be at 'Critical risk of bias' and no further assessment is required.

C. Specify the (hypothetical) target trial specific to the study

The target trial is either explicitly described by the primary study investigators or implied by the study's design and analysis. Alongside specifying the target trial, it is necessary to consider whether the analysis estimates the intention-to-treat effect or the per-protocol effect.

Guidance note 5: the target trial specific to the study

Evaluations of risk of bias are facilitated by considering the non-randomized study as an attempt to emulate a pragmatic randomized trial, which we refer to as the **target trial**. This is the hypothetical randomized trial whose results would be the same (other than due to sampling variation) as those from the non-randomized study under consideration, in the absence of bias. Its key characteristics are the eligible participants (specified by the inclusion/exclusion criteria) the intervention strategy and comparator strategy, the start and end of follow-up and the outcome assessed (see Table 1 of Cashin et al. (2025), Hernán and Robins (2016) and Hernán et al. (2025)).

Specifying the target trial based on a published study involves deducing or 'reverse engineering' the research question that the study addressed based on the analysis that was done. This may be difficult when the study itself does not clearly specify the question, or because the implicit intervention and comparator strategies implied by the analysis do not address a clinically relevant question.

Because it is hypothetical, ethics and feasibility need not be considered when specifying the target trial. For example, there would be no objection to a target trial that compared individuals who did and did not start smoking, even though such a trial would be neither ethical nor feasible in practice.

Differences between the target trial for the non-randomized study being considered and the generic research question of the review relate to issues of heterogeneity and/or generalizability rather than risk of bias.

Cashin AG, Hansford HJ, Hernán MA, Swanson SA, Lee H, Jones MD, Dahabreh IJ, Dickerman BA, Egger M, Garcia-Albeniz X, Golub RM, Islam N, Lodi S, Moreno-Betancur M, Pearson SA, Schneeweiss S, Sharp MK, Sterne JAC, Stuart EA, McAuley JH. Transparent reporting of observational studies emulating a target trial: the TARGET Statement. BMJ. 2025 Sep 3;390:e087179. doi: 10.1136/bmj-2025-087179.

Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* 2016;183:758-64; doi:10.1093/aje/kwv254.

Hernán MA, Dahabreh IJ, Dickerman BA and Swanson SA. The target trial framework for causal inference from observational data: why and when is it helpful? *Annals of Internal Medicine* 2025;178: 402-407. https://doi.org/10.7326/ANNALS-24-01871.

Guidance note 6: specifying the participants of the target trial

The eligible participants for the target trial are determined by the eligibility criteria for inclusion in the study and analysis being assessed for risk of bias. Specification of the eligible participants may require careful consideration. For example, Magid et al (2010) studied the comparative effectiveness of ACE inhibitors compared to beta blockers as second-line treatments for hypertension. From an initial cohort of 1.6 million patients, they restricted the analysis population to persons (1) with incident hypertension; (2) who were initially treated with a thiazide agent; (3) who had one of the two drugs of interest added as a second agent for uncontrolled hypertension; and (4) who did not have a contraindication to either drug. Their "comparative effectiveness" cohort included 15,540 individuals: less than 1% of the original cohort.

C1. Specify the participants and eligibility criteria

Male C57BL/6 mice, 8 weeks old, successfully exposed to MPTP 30 mg/kg Intraperitoneally for 5 days to induce a 5 days to induce a PD-like model;	Exclude mice not receiving model induction or not available for outcome testing

Guidance note 7: specifying the intervention and comparator strategies in the target trial

The intervention and comparator strategies may be point interventions that are administered at baseline, or be administered over a period of time rather than on a single occasion. Examples of point interventions include surgical interventions administered at a single time, a single vaccine dose, or starting a drug treatment. Note that, in non-randomized studies, the distinction between intention-to-treat and per-protocol effects (see Guidance note 8) is typically of little relevance for point interventions when participants are assigned (classified) into groups depending on the intervention they receive at baseline.

Examples of intervention strategies that are administered over time include surgery within 3 months of diagnosis, initial vaccination followed by a second dose 12 weeks later, or starting a drug treatment and continuing it for two years.

Analyses that estimate the intention-to-treat effect (see Guidance note 8) do not account for changes to the initial intervention. For such analyses detailed specification of the whole intervention and comparator strategy is not necessary.

For analyses that estimate the per-protocol effect (see Guidance note 8), the intervention and comparator strategies should be those whose effect is compared in the analysis that produced the result being assessed. This implies that the intervention changes that should not occur under the strategy (i.e. the protocol deviations) should be specified. For example:

- Suppose follow-up for individual participants is partitioned into time on drug A (intervention) and time on drug B (comparator). The intervention strategy is "Start on drug A and remain on drug A". The comparator strategy is "Start on drug B and remain on drug B". A protocol deviation occurs when participants stop the initial drug regimen.
- Suppose follow-up of comparison group participants is censored when they start the drug treatment without a clinical indication to do so. The comparator strategy is "Do not start the drug treatment during follow-up unless this is clinically indicated". A protocol deviation occurs when the comparison group participants start the drug treatment when it is not clinically indicated.
- Suppose follow-up of initially vaccinated participants is censored if they do not receive a second vaccine dose within 15 weeks. The intervention strategy is "Receive initial vaccination and then a second vaccine dose within 15 weeks". A protocol deviation occurs if a second vaccine dose is not received within 15 weeks.

C2. Specify the intervention strategy C3. Specify the comparator strategy	C2. B6 peptide-HA-PINK1 antibody-CS-TPP-ISP30 siRNA nanoparticles, 200 micrograms, intravenous tail injection every other day; behavior assessed on day 7 after first injection C3. PBS, 200 micrograms/vehicle equivalent, intravenous tail injection every other day under the same schedule

Guidance note 8: whether the analysis estimates the intention-to-treat effect or the per-protocol effect

ROBINS-I assessments of risk of bias due to confounding are tailored to whether the analysis estimates the 'intention-to-treat effect' or the 'per-protocol' effect. The intention-to-treat effect is the effect of assignment to the intervention versus the comparator, regardless of subsequent changes to the intervention received by participants. In non-randomized studies, assignment is often unknown, so it is usually inferred from the intervention started by or prescribed to each participant. If the analysis is estimating the intention-to-treat effect, Variant A of Domain 1 (confounding) is used.

Estimation of the intention-to-treat effect requires that participants' follow-up is not censored because of post-baseline changes to the initial intervention, for example if participants assigned to first-line cancer chemotherapy are censored when they start second-line therapy because of cancer recurrence.

It is sometimes necessary, or of interest, to account for changes to the intervention received by study participants over time because such changes are inconsistent with the intervention or comparator strategy. We refer to such changes as **protocol deviations**. For example, participants classified in a comparator "no intervention" group because they did not receive the intervention at baseline may deviate from the protocol by starting intervention during follow-up.

Changes to participants' initial intervention can be consistent with the intervention or comparator strategy. For example, stopping initial intervention because of intervention-related toxicities or switching to second-line treatment because of cancer progression are generally not considered to be protocol deviations so should not be accounted for in analyses.

The 'per-protocol' effect is the effect of receiving the intervention strategy versus the comparator strategy over time. Therefore, protocol deviations should be accounted for by analyses that estimate the per-protocol effect and, correspondingly, if the analysis accounts for protocol deviations then we can infer that it estimates the per protocol effect. Examples include: (i) interventions varied over time and follow-up for individual participants is partitioned according to the intervention received; or (ii) follow-up is censored when participants deviate from their initial intervention (e.g. participants assigned to a no-intervention comparison group start the intervention or those assigned to active intervention stop the intervention without a clinical indication).

When the analysis accounts for protocol deviations, it is necessary to account for 'time varying confounding', which is addressed in Variant B of Domain 1. Time-varying confounding occurs when time-varying prognostic factors influence changes to intervention after the start of follow-up. For example, in a cohort study of the effect of antiretroviral therapy (ART) on rates of AIDS and death in people with HIV, follow-up time for each participant was partitioned according to whether participants were receiving or not receiving ART. Because CD4 counts during follow-up influenced the decision to start ART, CD4 count was a time-varying confounder. Specialized analysis methods ('g-methods') should be used to estimate per-protocol effects, because standard regression models including time-varying confounders are likely to produce biased estimates of the effect of the intervention (versus the comparator) on the outcome.

C4. Did the analysis account for switches during follow-up between the intervention strategies being compared, or for other protocol deviations during follow-up?

☐ No (the analysis is estimating the intention-to-treat effect)

Analyzed as an intention-to-treat-like comparison of assigned treatment groups

☐ Yes (the analysis is estimating the per-protocol effect)

D. Information sources

Guidance note 9: information sources

Evaluation of a study should be based on the maximum possible amount of available information. In addition to published articles describing a study's methods and results, such information may be derived from the study protocol, unpublished reports or through correspondence with the study investigators.

Which of the following sources have you obtained to help you inform your risk of bias judgements (tick as many as apply)?

Journal article(s)
Study protocol
Statistical analysis plan (SAP)
Non-commercial registry record (e.g. ClinicalTrials.gov record)
Company-owned registry record (e.g. GSK Clinical Study Register record)
"Grey literature" (e.g. unpublished thesis)
Conference abstract(s)
Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
Individual participant data
Research ethics application
Grant database summary (e.g. NIH RePORTER, Research Councils UK Gateway to Research)
Personal communication with investigator
Personal communication with sponsor

Please specify any additional sources not listed above

Risk-of-bias assessment

Evaluation of confounding factors

Complete a row for each important confounding factor listed in advance (subsection (i) below); and either relevant to the setting of this particular study or identified by the study authors (subsection (ii)). **"Important" confounding factors are those for which, in the context of this study, adjustment is expected to lead to a meaningful change in the estimated effect of the intervention.**

Guidance note 10: evaluating confounding factors

Confounding is of fundamental importance to the analysis and interpretation of non-randomized studies of the effect of interventions on outcomes. It occurs when prognostic factors also predict receipt of the intervention or comparator strategy: we call these '**confounding factors**'. ROBINS-I addresses two types of confounding: baseline confounding and time-varying confounding. **Baseline confounding** occurs when one or more prognostic factors, present before the start of follow-up, predict whether participants start the intervention or comparator strategy. **Time-varying confounding** needs to be considered in analyses that estimate the per-protocol effect (see Guidance note 8) by allowing for time-varying interventions. Examples include analyses that partition follow-up time for individual participants according to intervention received, or that censor participants' follow-up when they deviate from their assigned intervention strategy. The time-varying confounding factors are prognostic factors that vary over time and predict post-baseline protocol deviations.

All the important confounding factors should be listed in this section. This includes both baseline confounding factors and, if relevant (if yes to C4), time-varying confounding factors.

It may not be possible to measure a confounding factor well, and we distinguish between the confounding factor and the **variables** used to measure it. These variables may be used, for example, as covariates in a regression analysis.

In the context of a particular study, variables need not be included in the analysis: (a) if they are not associated with the outcome, conditional on intervention received (noting that lack of a statistically significant association is not evidence of a lack of association); (b) if they are not associated with intervention; (c) if adjustment makes no or minimal difference to the estimated effect of intervention on outcome; (d) because the confounder was addressed in the study design, for example by restricting to individuals with the same value of the confounder; (e) because a negative control demonstrates that there was unlikely to have been confounding due to this variable or that uncontrolled confounding was likely to be minimal; or (f) because external evidence suggests that controlling for the variable is not necessary in the context of the study being assessed.

In some studies, researchers may include a very large set of potential confounding variables in an analysis without considering their associations with outcome and intervention. Users of ROBINS-I should focus on (i) the confounding factors they determined a priori to be important and (ii) other factors for which adjustment is expected to lead to an important change in the estimated effect of the intervention on the outcome in the context of the current study.

Users of ROBINS-I should evaluate the confounding factors that they prespecified as important for the intervention-outcome relationship under study. The tool also allows the user to evaluate a second list of any further confounding factors that are either relevant to the setting of this particular study or which the study authors identified as potentially important. It is likely that new ideas relating to confounding and other potential sources of bias will be identified after the drafting of the review protocol, and even after piloting data collection from studies selected for inclusion in the systematic review. For example, such issues may be identified because they are mentioned in

14

the introduction and/or discussion of one or more articles. This could be addressed in practice by explicitly recording whether potential confounders or other sources of bias are mentioned in the article.

In very rare situations it is possible that no confounding factors are present, either because interventions received are known to be unrelated to any prognostic factors for the outcome of interest, or because no such prognostic factors exist. In such situations, the risk of bias due to confounding may be assessed as low.

The purpose of this preliminary assessment of confounding factors is to review the extent to which the result being assessed was controlled for confounding, considering both the prespecified confounding factors and any further confounding factors identified as important in the context of the study being assessed. This enables users of ROBINS-I to answer the signalling questions for the Domain 1 assessment (Risk of bias due to confounding). "Important" confounding factors are those for which, in the context of this study, adjustment is expected to lead to an important change in the estimated effect of the intervention.

The preliminary assessment consists of the following steps for each confounding factor.

- determine which variables (if any) were measured for the factor;
- determine which of these variables were controlled for in the analysis;
- for variables that were not controlled for, look for evidence that controlling for the variable was not necessary in this particular study; • determine whether the confounding factor was measured validly and reliably by the variables used to measure it (this is assessed at the level of the confounding factor rather than the level of the individual variables used to measure the factor);
- determine the likely direction of bias if the analysis fails to adjust for this variable (alone).

The direction of bias, if the analysis fails to adjust for a particular variable (alone), will be that the effect estimate is biased *upwards* or biased *downwards*. For example, if older age predicts that a particular intervention is more likely to be received and the outcome is mortality, then this confounding would bias the estimated effect downwards: unless we adjust for age the intervention will appear more positively associated with higher mortality than it should. In the presence of *positive confounding* (the confounder is positively associated with both intervention and outcome, or negatively associated with both intervention and outcome), the bias will be upwards. In the presence of *negative confounding* (the confounder is positively associated with intervention and negatively associated with outcome, or vice versa), the bias will be downwards.

15

(i) Important confounding factors listed in advance

Confounding factor	Measured variable(s) for this factor, if any	Was this variable (or were these variables) controlled for in the analysis? (Y / N)	If this confounding factor was controlled for, was it measured validly and reliably by this variable (or these variables)?* (NA / Y / PY / PN / N / NI)	If this confounding factor was not controlled for, is there evidence that controlling for it was unnecessary?** (NA / Y / PY / PN / N)	OPTIONAL: Is failure to adjust for this confounding factor expected to bias the effect estimate upwards or downwards? (Upward bias (overestimate the intervention effect) / Downward bias (underestimate the intervention effect) / No information or unpredictable)	Comments
Mouse strain/sex/age	Y	Y by restriction	PY	NA		All treatment-Model mice appear to use male 8-week C57BL/6 mice; lowers confounding from age/sex/strain
Baseline MPTP severity/baseline motor performance	NI	N	NA	N	Unpredictable	Could bias either direction. Without baseline behavior after MPTP and before treatment, groups may have differed in severity

(ii) Additional important confounding factors relevant to the setting of this particular study, or identified by the study authors						
Confounding factor	Measured variable(s) for this factor, if any	Was this variable (or were these variables) controlled for in the analysis? (Y / N)	If this confounding factor was controlled for, was it measured validly and reliably by this variable (or these variables)?* (NA / Y / PY / PN / N / NI)	If this confounding factor was not controlled for, is there evidence that controlling for it was unnecessary?** (NA / Y / PY / PN / N)	OPTIONAL: Is failure to adjust for this confounding factor expected to bias the effect estimate upwards or downwards? (Upward bias (overestimate the intervention effect) /	Comments

					Downward bias (underestimate the intervention effect) / No information or unpredictable)	
Body weight/general health	NI	N	NA	N	unpredictable	Could affect MPTP sensitivity and motor performance
MPTP dose/timing/model induction	Y	PY by protocol	PY	NA	No info	Same protocol described, but no details on randomization/order/batch
Cage/litter/order/handling stress	NI	N	NA	N	Unpredictable	Could influence behavioral tests and treatment response
Nanoparticle batch/dose/treatment timing	Y	PY by protocol	PY	NA	No info	Dose/schedule reported; batch/randomization not reported
Assessor blinding/testing conditions	NI	N	NA	N	May overestimate treatment benefit if expectation favor NPs	Also affects measurement bias
Exclusions/dropouts after induction	NI	NI	NA	NI	No info	Missing-data concern because exclusions/attrition are not explicitly reported

* “Validity” refers to whether the confounding variable or variables accurately measure the confounding factor, while “reliability” refers to the precision of the measurement (more measurement error means less reliability).

** In the context of a particular study, variables need not be included in the analysis: (a)) if they are measured validly and reliably and are not associated with the outcome, conditional on intervention (noting that lack of a statistically significant association is not evidence of a lack of association; (b) if they are measured validly and reliably and are not associated with intervention; (c) if they are measured validly and reliably and adjustment makes no or minimal difference to the estimated effect of the primary parameter; (d) because the confounder was addressed in the study design, for example by restricting to individuals with the same value of the confounder; (e) because a negative control demonstrates that there was unlikely to have been confounding due to this variable or that uncontrolled confounding was likely to be minimal; or (f) because external evidence suggests that controlling for the variable is not necessary in the context of the study being assessed”.

Signalling questions and risk-of-bias judgements

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

1. Bias due to confounding

Guidance note 11: bias due to confounding

The questions in this domain focus on the confounding factors (see Guidance note 10) that were identified as important in the preliminary evaluation in section E.

We use the term **uncontrolled confounding** to refer to confounding that was not controlled by the design or analysis of the study – and is therefore likely to bias the estimated effect of intervention. This may arise because (i) confounding factors were not (or could not) be measured; (ii) variables used to measure confounding factors were insufficient to characterize the confounding factor; or (iii) variables that characterize the confounding factor were measured but not included in the analysis.

The answers to the signalling questions are based on contents of the completed table 'Evaluation of confounding factors'.

There are two variants of this domain.

Variant A is used when the analysis is estimating the intention-to-treat effect, when only baseline confounding needs to be addressed. Baseline confounding occurs when one or more prognostic factors, present before the start of follow-up, predict whether participants start the intervention or comparator strategy. Appropriate methods to control for baseline confounding include stratification, regression, matching, standardization, and inverse probability weighting. The analysis may control for individual variables or for estimated propensity scores. Inverse probability weighting is based on a function of the propensity score.

Variant B is used when the analysis is estimating the per-protocol effect, that is, the effect of receiving a specified intervention strategy versus a comparator strategy over time. This is the case when the analysis accounted for protocol deviations (switches during follow-up between the intervention strategies being compared, or other protocol deviations during follow-up). Such an analysis might be performed, for example, by partitioning follow-up for individual participants according to the intervention received, or by censoring follow-up when participants deviated from their initial intervention. When estimating the per protocol effect, both baseline confounding and time-varying confounding need to be addressed. It is usually **not appropriate** to attempt to control for time-varying confounding by including the time-varying confounders in regression models. Special methods ('g methods') that avoid 'conditioning on' time-varying confounders should be used. For example, in analyses that censor follow-up when participants deviate from their initial intervention strategy, one appropriate approach to analysis is to model the probability of such deviations over time, based on the baseline- and time—varying confounding factors, and then to adjusted for the censoring using inverse probability of censoring weights.

Domain 1, Variant A (the analysis is estimating the intention-to-treat effect so only baseline confounding needs to be addressed – if No to C4)

Signalling questions	Elaboration	Response
1.1 Did the authors control for all the important confounding factors for which this was necessary?	No randomization, baseline motor severity, baseline health.body weight, cage/order effects, or baseline adjustment are reported. This could materially affect a small behavioral study	SN

17

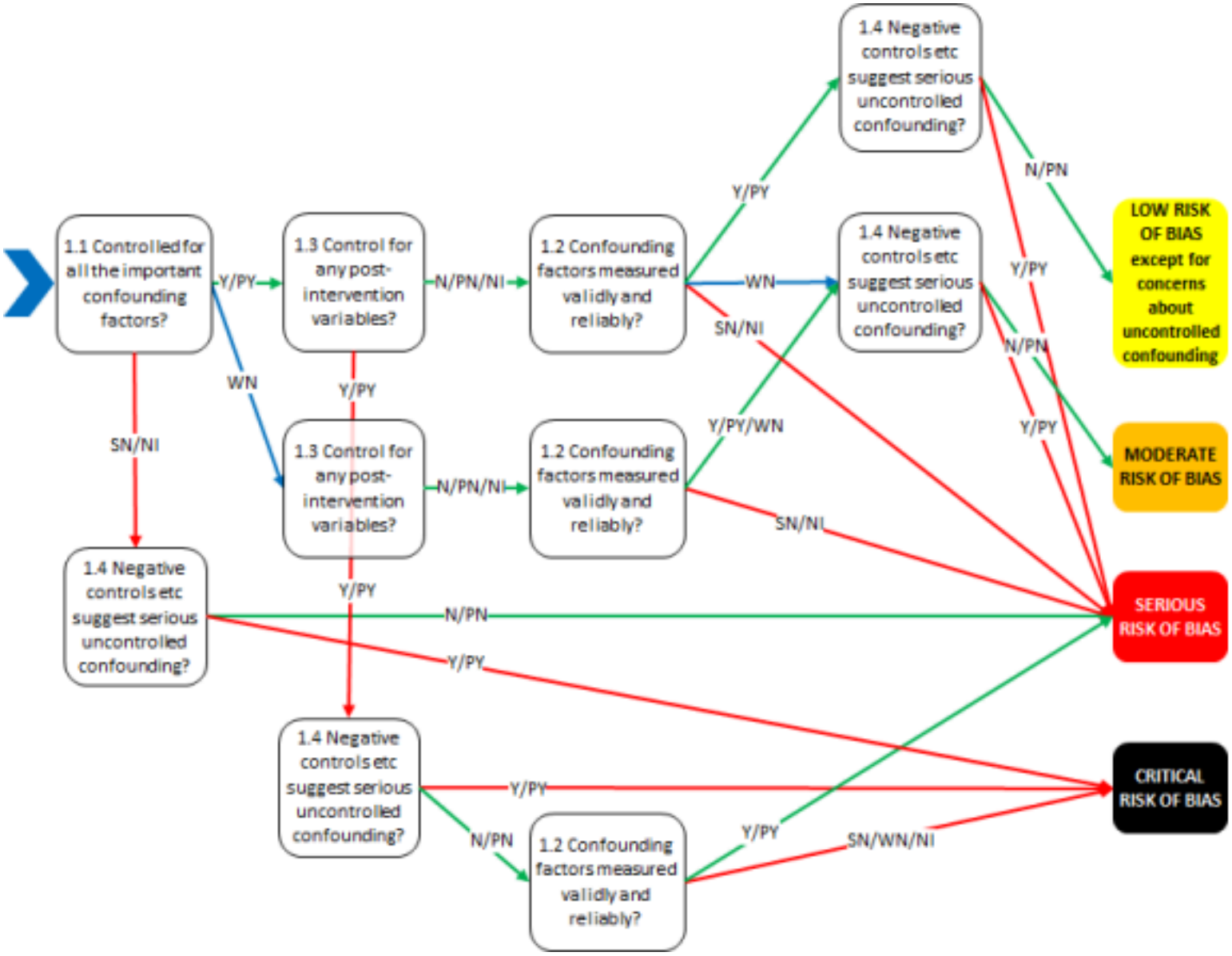
1.2 If Y/PY/WN to 1.1: Were confounding factors that were controlled for (and for which control was necessary) measured validly and reliably by the variables available in this study?	Not applicable because the key concern is lack of control for several important factors, not measurement of adjusted variables	NA

18

1.3 If Y/PY/WN to 1.1: Did the authors control for any post intervention variables that could have been affected by the intervention?	No adjusted analysis is described, so overadjustment is not apparent	NA/NI

1.4. Did the use of negative controls, quantitative bias analysis, or other considerations, suggest serious uncontrolled confounding?	Other considerations; absence of reported randomization/baseline equivalence suggest serious uncontrolled confounding; no negative-control analysis was provided	PY
Risk of bias judgement	Design standardization helps, but uncontrolled baseline severity and allocation/order effects could materially change the estimated treatment effect	Serious
Optional: What is the predicted direction of bias due to confounding?	Baseline severity imbalance could favor either treatment or comparator; measurement/reporting biases may favor intervention but are assessed in other domains	Unpredictable

Algorithm for reaching default risk of bias judgement:



Domain 1, Variant B (the analysis is estimating the per-protocol effect so both baseline and time-varying confounding need to be addressed – if Yes to C4)

Variant B of Domain 1 is used when the analysis accounted for protocol deviationssuch as switches during follow-up between the intervention strategies being

compared, or stopping active intervention without a clinical indication to do so. Accounting for protocol deviations might be done for example by partitioning follow-up for individual participants according to the intervention received, or by censoring follow-up when participants deviated from their initial intervention. It is then necessary to control for both baseline and time—varying confounding.

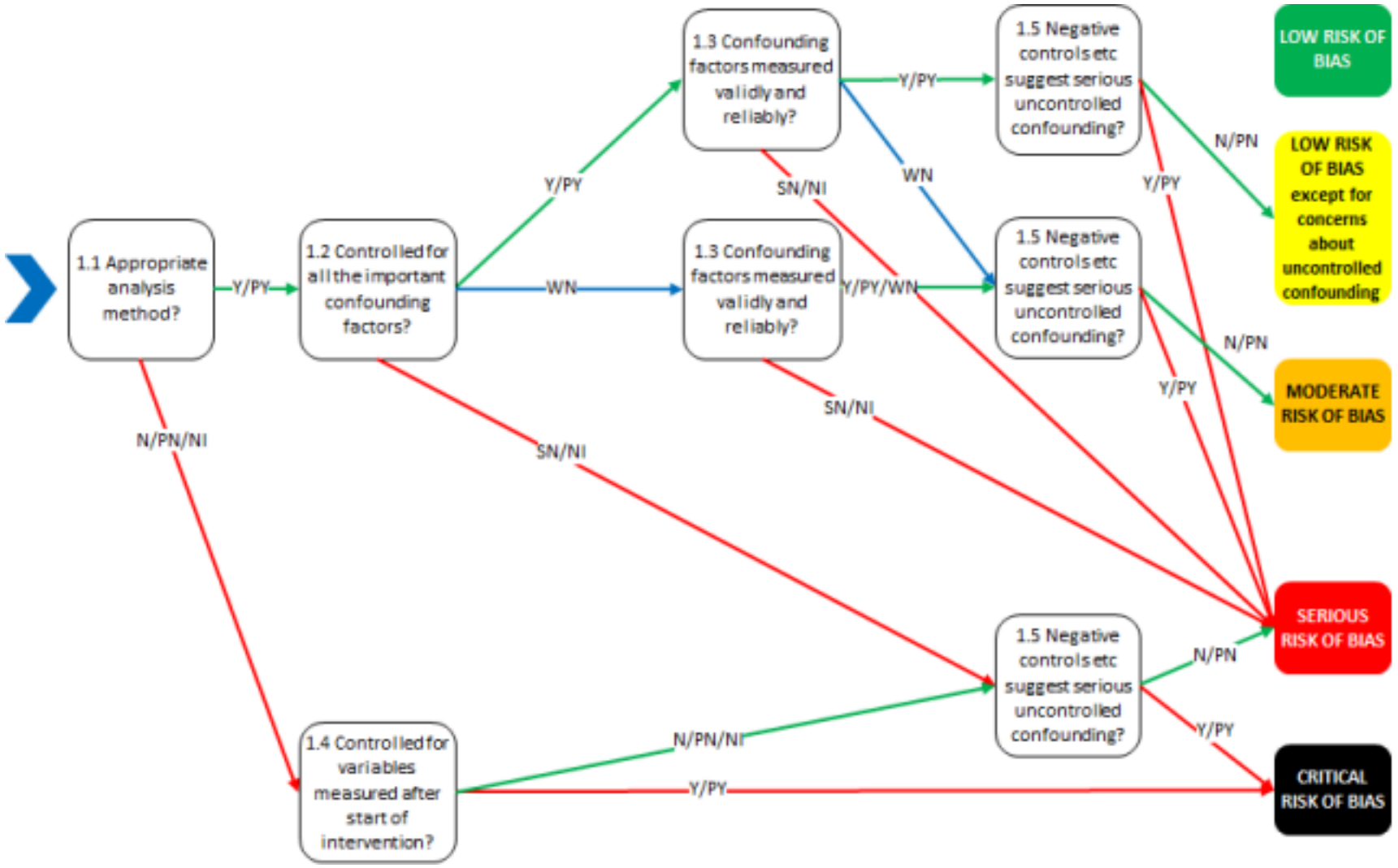
Signalling questions	Elaboration	Response
1.1 Did the authors use an analysis method that was appropriate to control for time-varying as well as baseline confounding?	Appropriate methods to control for time-varying confounding ('g methods') include those based on inverse probability weighting. Standard regression models that include time-varying confounders may be problematic when the time-varying confounders are affected by prior intervention (this is also known as treatment confounder feedback).	Y / PY / PN / N / NI
1.2 If Y/PY to 1.1: Did the authors control for all the important baseline and time-varying confounding factors for which this was necessary?	The important confounding factors are those specified in the *Preliminary consideration of confounding factors*. For studies that allow for time-varying interventions, for example by partitioning follow up time according to intervention received or by censoring participants' follow-up when they deviate from their assigned intervention strategy, the analysis should control for both the baseline confounding factors and for time-varying confounders (prognostic factors that predict changes to intervention received). The preliminary assessment will have determined whether there were important confounding factors that were not controlled for and should have been (because there was no evidence that controlling for the variable was unnecessary). Failure to control for all important confounding factors may lead to bias. Answer 'Y' or 'PY' if all the important confounding factors for which it is was deemed necessary to control (under *Preliminary consideration of confounding factors*) were indeed controlled for. Also answer 'Y' or 'PY' in the (very rare) situation that there are no confounding factors and an unadjusted analysis is presented. Answer 'WN' if most of the important confounding factors for which it was deemed necessary to control (under *Preliminary consideration of confounding factors*) were controlled for, and the any uncontrolled confounding (because not all important confounding factors were controlled for) was not likely to be substantial. This would be the case, for example, if the factors that were not controlled for were likely to be highly correlated with factors that were controlled for.	NA / Y / PY / WN (no, but uncontrolled confounding was probably not substantial) / SN (no, and uncontrolled confounding was probably substantial) / NI

	Answer 'SN' if there is at least one important confounding factor that should have been controlled for but was not, and the failure to control for this factor is likely to have a material impact on the estimated effect of intervention.	
1.3 If Y/PY/WN to 1.2: Were confounding factors that were controlled for (and for which control was necessary) measured validly and reliably by the variables available in this study?	Appropriate control of confounding requires that the variables adjusted for are valid and reliable measures of the confounding factors. For some topics, a list of valid and reliable measures of confounding factors will be specified in the review protocol but for others such a list may not be available. Study authors may cite references to support the use of a particular measure. If authors control for confounding variables with no indication of their validity or reliability, then subjectivity of the measure should be evaluated. In the (very rare) situation that there are no confounding factors and an unadjusted analysis is presented, answer 'Y' or 'PY'.	NA / Y / PY / WN (no, but the extent of measurement error in confounding factors was probably not substantial) / SN (no, and the extent of measurement error in confounding factors was probably substantial) / NI
1.4 If N/PN/NI to 1.1: Did the authors control for time varying factors or other variables measured after the start of intervention?	This question is asked if an inappropriate analysis method has been used to control for time-varying confounding factors. In such a situation, controlling for (conditioning on) time-varying factors measured after the start of intervention is likely to lead to bias if these are also on the causal pathway from the intervention to the outcome.	NA / Y / PY / PN / N / NI

1.5 Did the use of negative controls, or other considerations, suggest serious uncontrolled confounding?	Use of a "negative control" – exploration of an alternative analysis in which no association should be observed – can sometimes suggest that the result is subject to uncontrolled confounding, if similar associations are identified for the result being assessed and the negative control. If the study did not use negative controls and no other considerations suggest uncontrolled confounding, answer 'N'. Answer 'Y' or 'PY' if negative controls indicate that the result being assessed suffers from material bias due to confounding.	Y / PY / PN / N
Risk of bias judgement	See algorithm.	Low (except for concerns about uncontrolled confounding) / Moderate / Serious / Critical

Optional: What is the predicted direction of bias due to confounding?	If the likely direction of bias can be predicted, it is helpful to state this. A judgement about the predicted direction of bias should take into account all uncontrolled confounding from omitted important confounders and may therefore require judgements about the relative impact of confounding factors operating in different directions. The effect of confounding is to bias the estimated effect upwards or downwards – to overestimate or to underestimate the intervention effect. If the true intervention effect is above 0 (above 1 for a ratio measure) then an upward bias will also represent a bias away from the null; if the true intervention effect is below 0 (below 1 for a ratio measure) then a downward bias will also represent a bias away from the null. The situation is not always so clear. For example, if the true effect is above 0 then a downward bias may bring the estimate towards the null or beyond it (to a value less than 0).	Upward bias (overestimate the effect) / Downward bias (underestimate the effect) / Unpredictable

Algorithm for reaching default risk of bias judgement:



2. Bias in classification of interventions

Guidance note 12: bias in classification of interventions

This domain addresses bias arising from misclassification of the intervention strategies being compared. The first three questions in the domain address a type of misclassification bias arising from **immortal time**, time during which outcome events are prevented from happening by the way the analysis was done. Such analytical problems generally occur when the intervention and comparator strategies cannot be distinguished at the time when follow-up would have started in the target trial, making it difficult or impossible to know which intervention strategy participant were following. This is the first of two types of bias arising from immortal time that are addressed in ROBINS-I V2 (the second type, which is usually less serious, is addressed in *Domain 3, Bias in selection of participants into the study*).

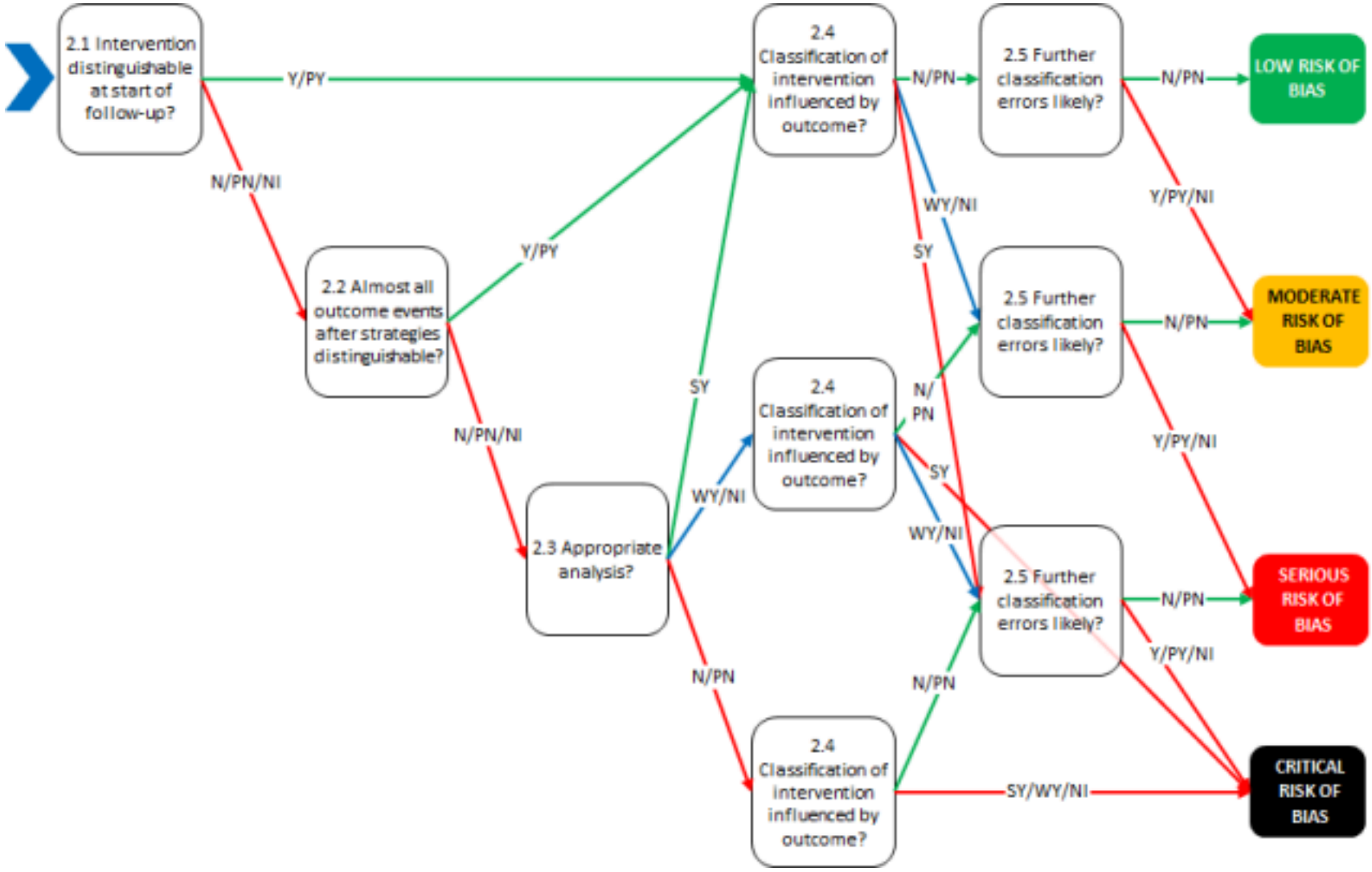
Misclassification is generally considered to be **differential** if it is related to subsequent outcomes; or **non differential** if it is unrelated to subsequent outcomes. The fourth and fifth signalling questions in the domain address differential misclassification and non-differential misclassification, respectively. Differential misclassification, which is generally considered to be more serious than non-differential misclassification, may occur when classification of intervention status is influenced by knowledge of the outcome or risk of the outcome.

Signalling questions	Elaboration	Response options
2.1 Were the intervention strategies distinguishable at the time when follow-up would have started in the target trial?	Groups were defined by injected PBS or specific nanoparticle formulations	Y
2.2 If N/PN/NI to 2.1: Did all or nearly all outcome events occur after the intervention and comparator strategies could be distinguished?	Strategies were distinguishable	NA

2.3 If N/PN/NI to 2.2: Did the analysis avoid problems arising from intervention strategies that are not distinguishable at the start of follow-up?	No immortal-time classification problem apparent	NA
2.4 Was classification of intervention status influenced by knowledge of the outcome or risk of the outcome?	Classification was based on treatment administered; not later outcome.	N

2.5 Were further classification errors (not influenced by knowledge of the outcome or risk of the outcome) likely?	No evidence that treatment formulations were confused, though allocation documentation is sparse	PN
Risk of bias judgement	The treatment groups are clearly defined by administered formulations	Low
Optional: What is the predicted direction of bias in classification of interventions?	No important intervention misclassification was identified	No clear bias/direction

Algorithm for reaching default risk of bias judgement:



3. Bias in selection of participants into the study (or into the analysis)

Guidance note 13: bias in selection of participants into the study or into the analysis

In their book *Causal Inference: What If*, Robins and Hernán define the term 'selection bias' as encompassing "various biases that arise from the procedure by which individuals are selected into the analysis". ROBINS-I focusses on selection biases that affect estimates of the effect of an intervention (versus a comparator) on an outcome: these arise from 'conditioning on a common effect of the intervention and the outcome'. 'Conditioning' on a variable happens when we restrict the analysis to people with a chosen value or range of values of the variables, or if we include the variable in a regression model for the effect of the intervention on the outcome.

A well-known example of selection bias is from studies of the effect of folic acid supplementation on the risk of neural tube defects in infancy. Many such studies were restricted, for pragmatic reasons, to babies born alive. However, both folic acid supplementation and foetal neural tube defects affect whether a pregnancy ends in live birth. The analysis was affected by selection bias because it was restricted to live births and excluded pregnancies that did not end in a live birth.

Selection bias is addressed in two ROBINS-I bias domains. This domain (Domain 3, Bias in selection of participants into the study (or into the analysis) addresses selection biases arising because participants, periods of follow-up or outcome events were omitted from the analysis because of participants characteristics that were measured, or events that occurred, after the start of the intervention strategies.

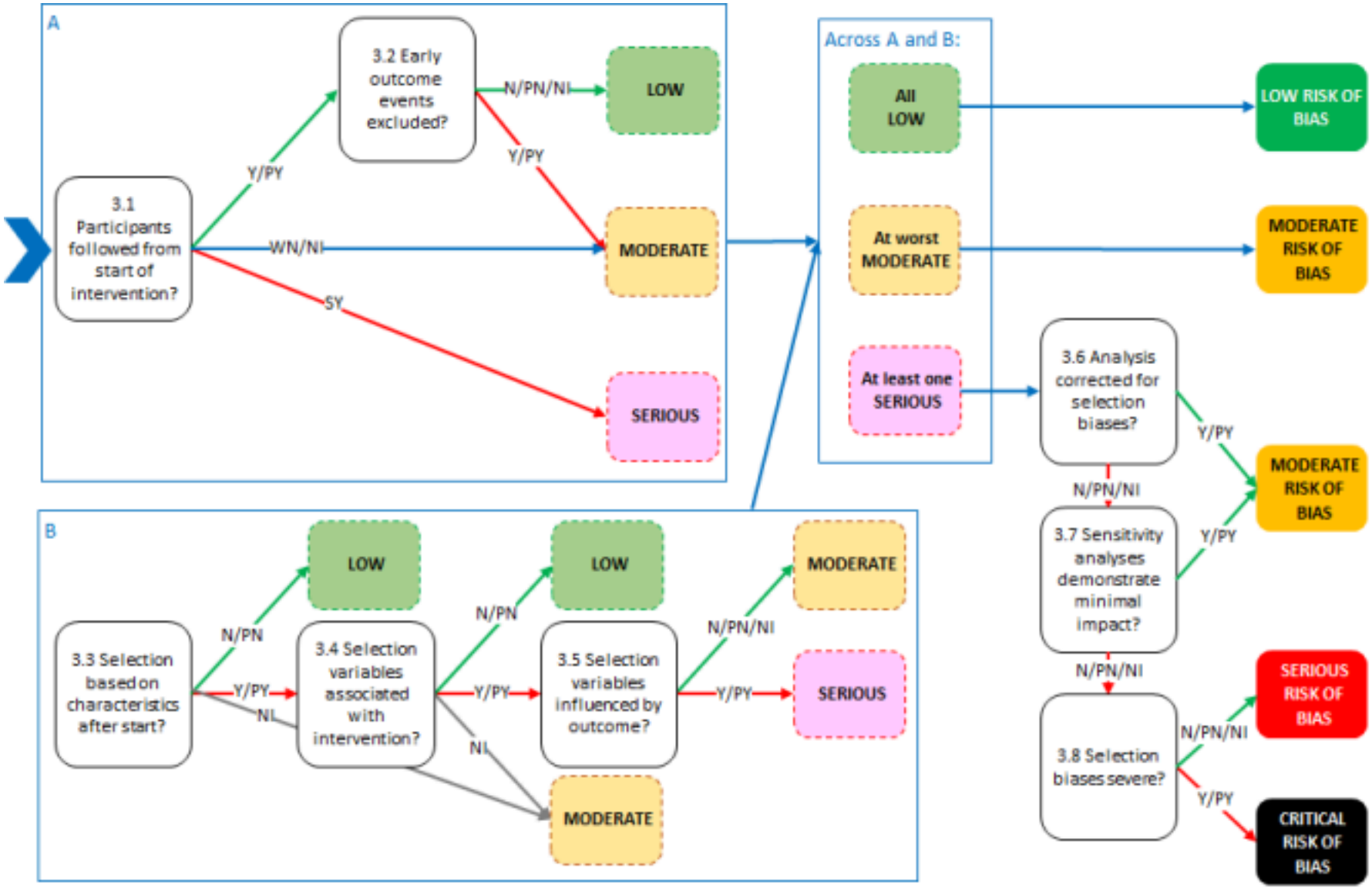
Domain 5 (Bias due to missing data) addresses selection biases that arise when because participants are excluded from the analysis because of missing data, or when analyses accounting for missing data cannot correct for the selection bias.

Signalling questions	**Elaboration**	**Response options**
A. Questions about prevalent user bias and immortal time		
3.1 Did follow up in the analysis begin at the start of the intervention strategies being compared?	Behavioral follow-up appears tied to treatment initiation and assessed day 7 after first injection	PY

3.2 If Y/PY to 3.1: Were outcome events during a period of follow-up after the start of the interventions excluded from the analysis?	No period after treatment start appears to have been omitted	N
B. Questions about other types of selection bias		
3.3 Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention (additional to the situations addressed in 3.1 and 3.2)?	No post-intervention selection into analysis is described	PN
3.4 If Y/PY to 3.3: Were the post-intervention variables that influenced selection likely to be associated with intervention?	No such selection process described	NA
3.5 If Y/PY to 3.4: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA
C. Questions about analysis, sensitivity analyses and severity of the problem		
3.6 If SN to 3.1 or Y/PY to 3.5: Is it likely that the analysis corrected for all of the potential selection biases identified above?	No substantial post intervention selection bias identified	NA

3.7 If N/PN/NI to 3.6: Did sensitivity analyses demonstrate that the likely impact of the potential selection biases identified above was minimal?	**Not needed**	NA
3.8 If N/PN/NI to 3.7: Were potential selection biases identified above sufficiently severe that the result should not be included in a quantitative synthesis?	Not applicable	NA
Risk of bias judgement	Low for ROBINS-I mechanisms, but downgraded slightly because animal inclusion/exclusion details and n values are not fully transparent	Low/moderate
~~Optional: What is the predicted direction of bias in selection of participants into the study?~~	~~If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized as being in favour of the intervention, as being in favour of the comparator, or as towards (or away from) the null.~~	~~Favours intervention / Favours comparator / Towards null /Away from null / Unpredictable~~

Algorithm for reaching default risk of bias judgement



4. Bias due to missing data

Guidance note 14: bias due to missing data

Missing outcome data may arise, among other reasons, through attrition (loss to follow up), missed appointments and incomplete data collection. Additionally, in non randomized studies data may be missing for characteristics including interventions received and confounders.

A general rule for consideration of bias due to missing data is that we should consider biases introduced by the missing data, compared with the effect estimate from an analysis in which all the data we intended to collect were available. Unfortunately, a single threshold for an acceptable proportion of missing data cannot meaningfully be defined. For example, a result based on 95% complete outcome data might be biased if the outcome was rare and if reasons for missing outcome data were strongly related to intervention group. Therefore, the potential for bias due to missing data should be assessed unless complete data on intervention status, the outcome and confounding variables were available all, or nearly all, participants.

Considerations of bias due to missing data depend on how the analysis accounted for the missing data. Different signalling questions should be answered depending on three types of analysis. The first is that a **complete case analysis**, restricted to participants with complete data on all the intervention, outcome and confounding variables, was performed. In this situation, an important consideration is whether missingness of individual participants from the analysis is related to the true value of the outcome for those participants. The second is that missing data were **imputed**, which means that estimated or assumed values were assigned to participants with missing data. Imputed data should not lead to bias if the data are 'missing at random' (see the elaboration for signalling question 4.8) and an appropriate imputation method is applied. Other types of analysis are addressed by a separate, general, signalling question. The final signalling question asks whether sensitivity analyses were performed that demonstrated that the impact of missing data is minimal.

Further discussion of many of these issues can be found in Lee et al (J Clin Epidemiol 2021;134:79-88) and Hughes et al (IJE 2019;48:1294-1304).

Signalling questions	Elaboration	Response options
4.1 Were complete data on intervention status available for all, or nearly all, participants?	Intervention assignment is clear for plotted groups, but individual animal records are unavailable	PY
4.2 Were complete data on the outcome available for all, or nearly all, participants?	No attrition/exclusion statement. Methods say four four groups of three mice for animal treatment, while Fig. 11 behavior caption reports n = 12, creating uncertainty	PN / NI

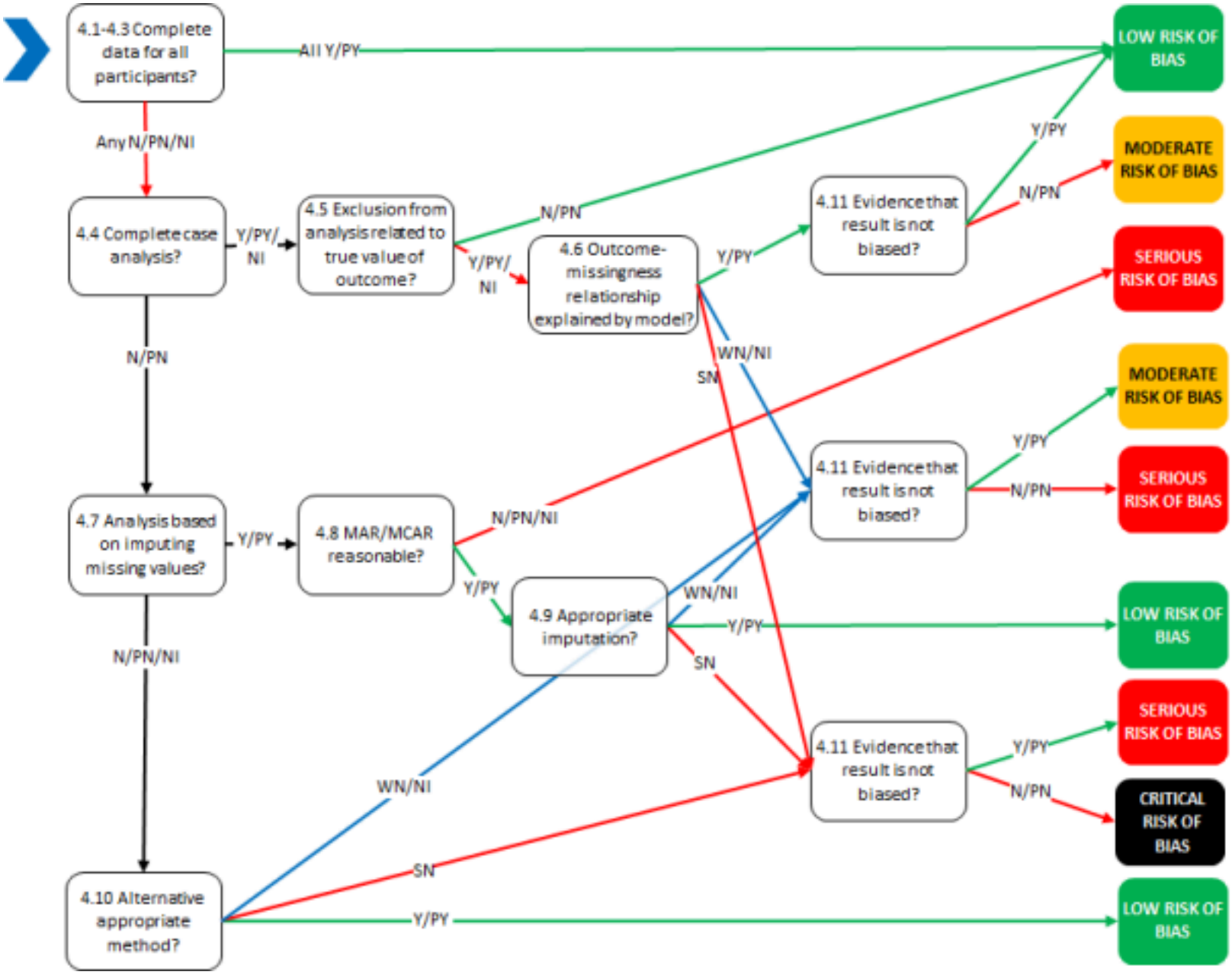
4.3 Were complete data on important confounding variables available for all, or nearly all, participants?	Important baseline covariates were not reported or used in analysis	NI
4.4 If N/PN/NI to 4.1, 4.4 or 4.4: Is the result based on a complete case analysis?	Likely complete-case based on plotted animals, but not explicitly stated	PY /NI
4.5 If Y/PY/NI to 4.4: Was exclusion from the analysis because of missing data (in intervention, confounders or the outcome) likely to be related to the true value of the outcome?	No exclusion/dropout information provided	NI

4.6 If Y/PY/NI to 4.5: Is the relationship between the outcome and missingness likely to be explained by the variables in the analysis model?	No model for missingness or covariates reported	NA / NI
4.7 If N/PN to 4.4: Was the analysis based on imputing missing values?	No imputation reported	N

4.8 If Y/PY to 4.7: Is it reasonable to assume that data were 'missing at random' (MAR) or 'missing completely at random' (MCAR)?	No imputation reported	N
4.9 If Y/PY to 4.8: Was imputation performed appropriately?	No imputation reported	N
4.10 If N/PN/NI to 4.7: Was an appropriate alternative method used to correct for bias due to missing data?	No weighting or sensitivity analysis reported	N

4.11 If PN/N/NI to 4.1, 4.2 or 4.3 AND (Y/PY/NI to 4.5 OR WN/SN/NI to 4.9 OR WN/SN/NI to 4.10): Is there evidence that the result was not biased by missing data?	No attrition table or sensitivity analysis	N/NI
Risk of bias judgement	There is insufficient information about missing/excluded animals and inconsistent n reporting, but no directed evidence of major differential missingness	Moderate
~~Optional: What is the predicted direction of bias due to missing data?~~	~~If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized as being in favour of the intervention, as being in favour of the comparator, or as towards (or away from) the null.~~	~~Favours intervention / Favours comparator / Towards null /Away from null / Unpredictable~~

Algorithm for reaching default risk of bias judgement:


4.1-4.3 Complete data for all participants?
All Y/PY
LOW RISK OF BIAS
Any N/PN/NI
4.4 Complete case analysis?
Y/PY/ NI
4.5 Exclusion from analysis related to true value of outcome?
N/PN
Y/PY/ NI
4.6 Outcome-missingness relationship explained by model?
Y/PY
4.11 Evidence that result is not biased?
Y/PY
N/PN
MODERATE RISK OF BIAS
SERIOUS RISK OF BIAS
WN/NI
SN
N/PN
4.7 Analysis based on imputing missing values?
Y/PY
4.8 MAR/MCAR reasonable?
N/PN/NI
Y/PY
4.11 Evidence that result is not biased?
Y/PY
N/PN
MODERATE RISK OF BIAS
SERIOUS RISK OF BIAS
4.9 Appropriate imputation?
WN/NI
Y/PY
SN
LOW RISK OF BIAS
N/PN/NI
4.11 Evidence that result is not biased?
Y/PY
N/PN
SERIOUS RISK OF BIAS
CRITICAL RISK OF BIAS
WN/NI
SN
4.10 Alternative appropriate method?
Y/PY
LOW RISK OF BIAS

5. Bias in measurement of the outcome

Guidance note 15: bias in measurement of the outcome

Bias may be introduced if outcomes are misclassified or measured with error. Misclassification or measurement error of outcomes may be non-differential or differential.

Non-differential measurement error is unrelated to the intervention received. It can be systematic (for example when measurement of blood pressure is consistently 5 units too high in every participant) – in which case it will not affect precision or cause bias; or it can be random (for example when measurement of blood pressure is sometimes too high and sometimes too low in a manner that does not depend on the intervention or the outcome) – in which case it will affect precision without causing bias.

Differential measurement error is measurement error related to intervention received. It will bias the intervention-outcome relationship. This is often referred to as detection bias. Examples of situations in which detection bias can arise are (i) if outcome assessors are aware of intervention received (particularly when the outcome is subjective); (ii) different methods (or intensities of observation) are used to assess outcomes of participants receiving different interventions; and (iii) measurement errors are related to intervention received (or to a confounder of the intervention-outcome relationship).

Blinding of outcome assessors aims to prevent systematic differences in measurements according to intervention received. However, blinding is frequently not possible or not performed for practical reasons.

Signalling questions	Elaboration	Response options
5.1 Could measurement or ascertainment of the outcome have differed between intervention groups?	The same behavioral tests appear to be used across groups	PN
5.2 Were outcome assessors aware of the intervention received by study participants?	The paper does not report blinding of behavioral testers or histology/cell-count assessors	NI

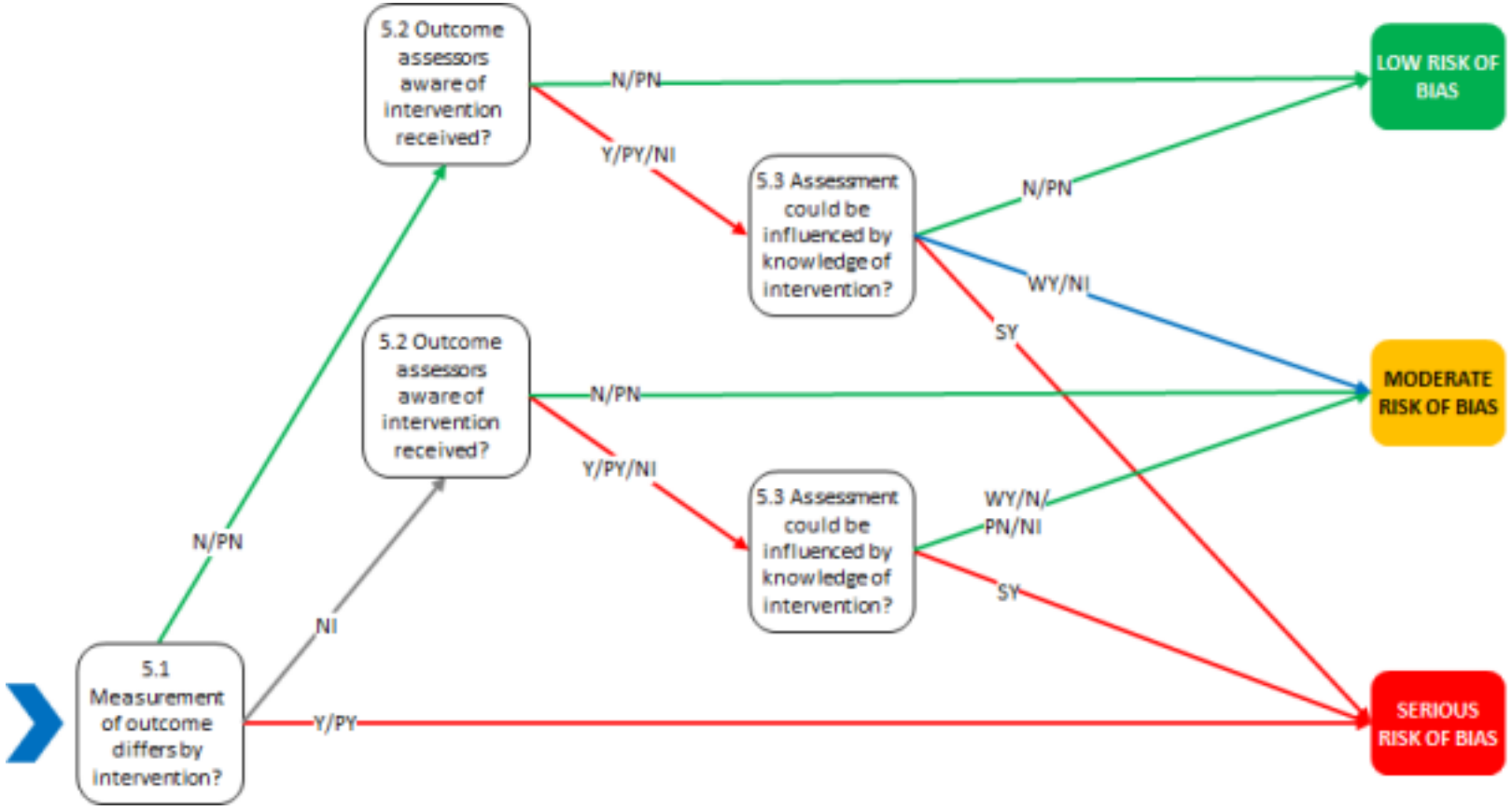
5.3 If Y/PY/NI to 5.2: Could assessment of the outcome have been influenced by knowledge of the intervention received?	Hanging score and rod-climbing timing involve observation/judgement; histology neuron counts can also be influenced if unblinded. Direction likely favors the expected beneficial NP group	SY/PY

39

Risk of bias judgement	Lack of reported blinding for behavioral and histological outcomes creates a material risk of differential outcome assessment.	Serious
Optional: What is the predicted direction of bias in measurement of outcomes?	Knowledge of treatment could bias scoring, timing, image selection, or cell counting toward better outcomes in the nanoparticle group.	Favours intervention / Away from null

40

Algorithm for reaching default risk of bias judgement:



6. Bias in selection of the reported result

Guidance note 16: bias in selection of the reported result

Selective reporting can arise for both harms and benefits of an intervention, although the motivations (and direction of bias) underlying selective reporting of effect estimates for harms and benefits may differ. Selective reporting may arise, for example, from a desire for findings to be newsworthy (or sufficiently noteworthy to merit publication), or from commercial considerations, or from a desire to demonstrate that there is not evidence of a harmful effect of an intervention.

Selective outcome reporting occurs when the effect estimate for an outcome measurement was selected from among analyses of multiple outcome measurements for the outcome domain. Examples include: use of multiple measurement instruments (e.g. pain scales) and reporting only the most favourable result; reporting only the most favourable subscale (or a subset of subscales) for an instrument when measurements for other subscales were available; reporting only one or a subset of time points for which the outcome was measured.

Selective analysis reporting occurs when results are selected from effects estimated in multiple ways: e.g. carrying out analyses of both change scores and post intervention scores adjusted for baseline; multiple analyses of a particular measurement with and without transformation; multiple analyses of a particular outcome with and without adjustment for potential confounders (or with adjustment for different sets of potential confounders); multiple analyses of a particular outcome with and without, or with different, methods to take account of missing data; a continuously scaled outcome converted to categorical data with different cut-points; multiple composite outcomes analysed for one outcome domain, but results were reported only for one (or a subset) of the composite outcomes. (Reporting an effect estimate for an unusual composite outcome might be evidence of such selective reporting.)

Selection of a subgroup from a larger cohort: The cohort for analysis may have been selected from a larger cohort for which data were available on the basis of a more interesting finding. Subgroups defined in unusual ways (e.g. an unusual classification of subgroups by dose or dose frequency) may provide evidence of such selective reporting.

The best evidence that results were not selectively reported is available if a pre-specified, publicly available analysis plan is available (e.g. from a link in a publication or from an online platform) and is in line with the reported results. Protocols for non-randomized studies are increasingly being registered, although there is inconsistency across platforms (Malmsiø et al, 2022). An analysis plan that is sufficiently detailed to permit full assessment of selective reporting may seldom be available for observational studies. In the absence of a protocol or analysis plan, clues can sometimes be gained by comparing Methods sections with Results sections.

Malmsiø D, Frost A, Hróbjartsson A. A scoping review finds that guides to authors of protocols for observational epidemiological studies varied highly in format and content. J Clin Epidemiol. 2022 Dec 20;154:156-166. doi: 10.1016/j.jclinepi.2022.12.012.

Signalling questions	Elaboration	Response options
6.1 Was the result reported in accordance with an available, pre-determined analysis plan?	No protocol, registry, or statistical analysis plan was available	NI

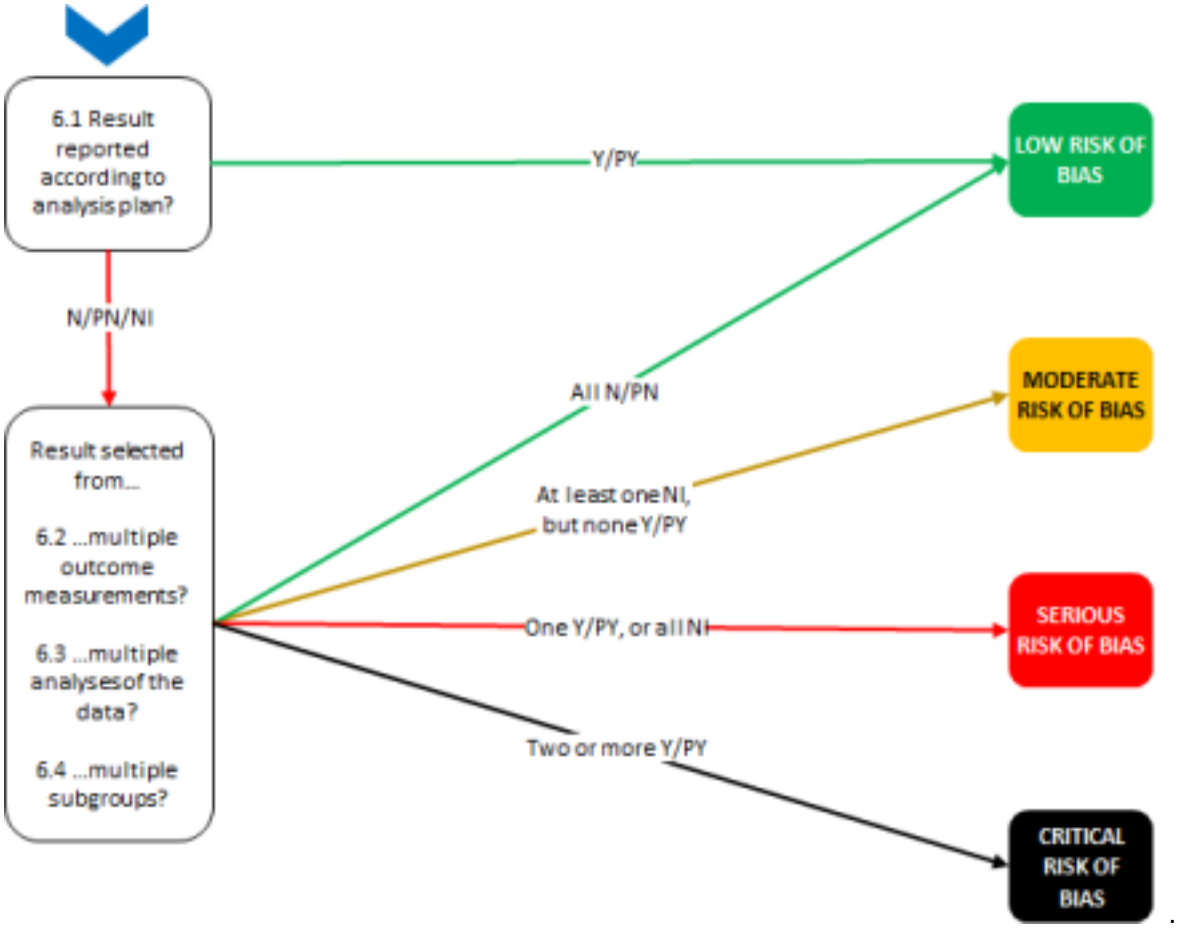
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...		
6.2 ... multiple outcome *measurements* (e.g. scales, definitions, time points) within the outcome domain?	The paper includes multiple outcomes domains and measures: behavior, TEM morphology, neuron count, USP30 expression, fluorescence localization, OCR/ECAR, ATP/SOD, cell viability, and multiple formulations/molecular weights	PY

6.3 ... multiple *analyses* of the data?	Multiple comparisons and treatment variants are presented; exact planned analyses, corrections, and full result set are not available	PY/ NI

6.4 ... multiple *subgroups*?	There are multiple models/cell types and formulations, but no clear subgroup selection within mouse behavioral analysis. Still, intentions are not available	PN / NI

Risk of bias judgement	No prospective plan plus many possible outcomes/analyses creates substantial selective-reporting concern.	Serious
Optional: What is the predicted direction of bias in selection of the reported result?	Selective presentation is more likely to emphasize favorable nanoparticle effects	Favours intervention Away from null

Algorithm for reaching default risk of bias judgement:



Overall risk of bias

Guidance note 17: overall risk of bias

ROBINS-I defaults to setting the overall risk of bias for a result to be equal to the risk-of-bias judgement for the domain with the greatest risk of bias. For example, if the 'worst' judgement across domains is of serious risk of bias, then the result would be judged as at serious risk of bias overall. However, the user may override this to judge the result to be at greater risk of bias if there are problems in several domains. For example, if several domains are assessed to be at serious risk of bias, and it is considered that these problems are likely to be compounded, then it may be reasonable to judge the result to be at critical risk of bias overall.

Predicting the direction of bias overall may be difficult. Risk-of-bias judgements for the individual domains might be used to inform the influence of that domain to the likely direction of bias overall.

Overall risk of bias	See algorithm.	Serious risk
What is the predicted direction of bias?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized as being in favour of the intervention, as being in favour of the comparator, or as towards (or away from) the null. Alternatively, if there direction is driven by bias due to confounding, the direction may be an upwards bias (overestimate the effect) or a downward bias (underestimate the effect).	Upward bias (overestimate the effect) / Towards null /Away from null

Algorithm for reaching overall risk of bias judgement:

Judgement	**Interpretation**	**How reached**
Low risk of bias except for concerns about uncontrolled confounding	There is the possibility of uncontrolled confounding that has not been controlled for (given the observational nature of the study), but otherwise little or no concern about bias in the result	*Low risk of bias except for concerns about uncontrolled confounding* in Domain 1 and *Low risk of bias* in all other domains

Moderate risk of bias	There is some concern about bias in the result, although it is not clear that there is an important risk of bias	At least one domain is at *Moderate risk of bias*, but no domains are at *Serious risk of bias* or *Critical risk of bias*
Serious risk of bias	The study has some important problems: characteristics of the study give rise to a serious risk of bias in the result	At least one domain is at *Serious risk of bias*, but no domains are at *Critical risk of bias* OR Several domains are at *Moderate*, leading to an additive judgement of *Serious risk of bias*
Critical risk of bias	The study is very problematic: characteristics of the study give rise to a critical of bias in the result, such that the result should generally be excluded from evidence syntheses.	At least one domain is at *Critical risk of bias* OR Several domains are at *Serious risk of bias*, leading to an additive judgement of *Critical risk of bias*