Appendices

Appendix A. Characteristics of systematic reviews assessed using the ROBINS-I instrument

Pilot Round	Title (Reference)	Exposure	Outcome	Number of Studies	Number of raters per study	Study design (n)
Round 1	Draft protocol for systematic review to evaluate the evidence for an association between bisphenol A (BPA) exposure and obesity[1]	BPA	Overweight and obesity	14	2	Cohort (2) Cross-sectional (12)
Round 2	The Navigation Guide-Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth[2]	PFOA	Fetal growth (i.e., birth weight)	17	2	Cohort (7) Cross-sectional (10)
Round 3	The Correlation between Polybrominated Diphenyl Ethers (PBDEs) and Thyroid Hormones in the General Population: A Meta-Analysis[3]	PBDEs	Thyroid function as measured by thyroid simulation hormones (TSHs) or thyroid hormone thyroxine (T4)	17	3	Cohort (3) Case-control (1) Cross-sectional (13)

BPA: bisphenol A; PBDE: polybrominated diphenyl ethers; PFOA: perfluorooctanoic acid.

[1] Thayer K, Rooney A, Boyles A, Holmgren S, Walker V, Kissling G, U.S. Department of Health and Human Services: Draft protocol for systematic review to evaluate the evidence for an association between bisphenol A (BPA) exposure and obesity. National Toxicology Program 2013.
[2] Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, Woodruff TJ: The Navigation Guide-Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth. Environmental health perspectives 2014.
[3] Zhao XM, Wang HL, Li J, Shan ZY, Teng WP, Teng XC: The Correlation between Polybrominated Diphenyl Ethers (PBDEs) and Thyroid Hormones in the General Population: A Meta-Analysis. Plos One 2015, 10(5).

Appendix B. Detailed methods of the evaluation of ROBINS-I and development of the RoB instrument for NRS of exposures

Methods

Instruments

Initially released as ACROBAT-NRSI in 2014 and renamed as ROBINS-I in 2016, this study used both iterations of the instrument when assessing understanding and applicability to environmental exposure studies; however, we refer to the instrument as ROBINS-I throughout even if the earlier version was used[4].

Signaling question response options include 'Yes', 'Probably yes', 'Probably no', 'No', and 'No information', and are complemented by free text fields to capture response judgments. Raters use the signaling question and free-text responses to make domain-level judgments about RoB. Domain- and study-level response options include 'Low', 'Moderate', 'Serious', and 'Critical' RoB. The individual study-level RoB is typically taken from the most severe of the domain-level judgments, unless the rater feels that the individual study should be rated as having greater RoB than that based on several affected domains. Domain-level responses across a body of evidence (across studies) allow an assessment of how much the domain-level RoB judgments may contribute to the trustworthiness of the entirety of evidence.

[4] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I et al: ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016, 355:i4919. ACROBAT-NRSI: A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions. Accessed 24 September 2014. [https://sites.google.com/site/riskofbiastool/].

Preparation for an evaluation using this instrument includes populating both a project- and an individual study-level protocol[6]. For each research question, raters complete one project-level protocol, identifying their target randomized trial research question. The target randomized trial research question identifies the population, intervention, comparison, and outcomes of interest. Based on this target trial, raters identify the nature of the target comparison (i.e., effect of interest), potential confounders and the relationship between them and the confounding domains for the research project. It also includes addressing possible co-interventions that could have an impact on the study outcomes, and the result(s) being assessed. For each individual study eligible to answer the review question, reviewers complete a study-level protocol. Text fields in the study-level protocol reflect those in the project-level protocol, to facilitate the abstraction of information from each individual study to determine generalizability and applicability to answering the project-level research question. Raters extract information to assess whether or not confounders and co-interventions identified as critical were addressed in the individual study and whether the individual study identified additional confounders or co-interventions.

Systematic reviews selected for pilot testing

We assessed the utility of ROBINS-I by piloting the instrument on all primary studies included in two previously published systematic reviews and studies identified from a draft case study protocol developed by OHAT as part of its early efforts to implement systematic review[5]. We selected previously published systematic reviews and a draft case study protocol that presented

[5] Thayer K, Rooney A, Boyles A, Holmgren S, Walker V, Kissling G, health UDo, services h: Draft protocol for systematic review to evaluate the evidence for an association between bisphenol A (BPA) exposure and obesity. National Toxicology Program 2013. Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, Woodruff TJ: The Navigation Guide-Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth. Environmental health perspectives 2014. Zhao XM, Wang HL, Li J, Shan ZY, Teng WP, Teng XC: The Correlation between Polybrominated Diphenyl Ethers (PBDEs) and Thyroid Hormones in the General Population: A Meta-Analysis. Plos One 2015, 10(5).

both persistent and non-persistent chemicals, as well as included primary studies featuring a variety of NRS designs (cohort, case-control, and cross-sectional). Using ROBINS-I, raters evaluated studies identified in two systematic reviews and one draft case study protocol of environmental epidemiological studies: 1) exposure to bisphenol A (BPA) and its association with obesity; 2) developmental exposure to perfluorooctanoic acid (PFOA) and its effect on fetal growth; and 3) exposure to polybrominated diphenyl ethers (PBDEs) and its effect on thyroid function (Appendix A). Each of the reviews represented a collection of 14-19 studies, most of which were cross-sectional in design but also included several cohort studies.

Evaluation of selected systematic reviews

For the first and second rounds of user testing, which informed initial revisions to the instrument, two raters independently responded to the signaling questions and provided domain- and study-level RoB judgments according to the ROBINS-I instrument from each study within the selected systematic reviews on BPA and PFOA into Microsoft Excel. A third rater reviewed the results, established consensus, and determined overall RoB for each study. In the third round of user testing, the three raters independently applied the modified instrument to a systematic review looking at the impact of PBDEs on thyroid function. The three raters then agreed on overall RoB for each study.

Our rating protocol, developed for each review, identified the hypothetical (i.e., target) randomized trial, potential confounders, and possible co-exposures of interest. Initially, raters identified pre-specified chemical confounders and possible co-exposures related to the health outcomes. We used sources such as the PhenX Toolkit (https://www.phenxtoolkit.org/) to

identify key confounders for the health outcomes[6]. Topic-specific experts provided guidance to address raters' unfamiliarity with the topic of each systematic review. When raters recognized additional confounders or co-exposures mentioned in the studies, these were added to the protocol; all studies were then re-evaluated so that raters considered the most comprehensive lists of confounders and co-exposures.

In the three rounds of pilot testing, raters received a form to identify and document barriers and facilitators to the use of the ROBINS-I in studies of environmental health. Also, we asked raters to provide descriptions of their understanding of each signaling question in the ROBINS-I instrument to identify areas requiring additional clarity and/or rewording. When deciding to modify ROBINS-I for the subsequent rounds of pilot testing, we considered modifications suggested by raters: for example, repeated misunderstanding of specific signaling questions over the multiple rounds of pilot testing.

Data analysis

When discrepancies were identified during the first and second round of testing, the third reviewer discussed with the two raters to determine the basis for the discrepancy, i.e., confusion on the item or differences of opinion on the raters' observations. We discussed differences related to the clarity of the item and either reworded the item or provided additional guidance for the question, as necessary. Similarly, in the third round of user testing, all three raters that

[6] Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H: The PhenX Toolkit: get the most from your measures. American journal of epidemiology 2011, 174(3):253-260.

provided the review of the studies discussed and arrived at a consensus on the response to each instrument item and overall RoB.

Appendix C. Modifications made as a result of three rounds of pilot testing and external consultation

Methods used during pilot-testing of ROBINS-I and subsequent modifications of the instrument	1) Development of Step I: a. *A priori*, topic-specific experts of the environmental exposures of interest provided input to Step 1 of the instrument, identifying critical confounders, potential co-exposures, and identifying characteristics of the exposure and health outcome measurement accuracy, such as its persistence. b. Raters consulted a database on chemical and environmental exposures, the PhenX Toolkit (https://www.phenxtoolkit.org/), to identify potential confounders [19]. c. Topic-specific experts provided background information for raters when applying ROBINS-I or the modified instrument. 2) Completion of Step II & III a. To improve reliability of responses, at minimum, two raters independently applied the instrument to each study in the systematic reviews, and compared and discussed their evaluations to reach consensus. b. Topic-specific experts performed additional piloting of the modified instrument.
Round 1: BPA and obesity	3) Replacement of the word 'intervention' with 'exposure' throughout the document; 4) Additional written instructions to address how to respond to signalling questions about temporality in a study of cross-sectional design a. For example, when responding to question 1.6 "Did authors avoid adjusting for post-intervention variable", we added "In a cross-sectional study, post-exposure variables are not studied and thus the action of adjusting or not adjusting for them does not present a risk to bias in the study. Therefore, the response option selected should represent that the risk to bias is not present or minimally present, not that the question is 'Not applicable'." 5) Additional instructions in conversations to address the subjectivity of the answer choices (for example the difference between 'Yes' and 'Probably Yes') and importance of explanations for why an answer choice was selected 6) Additional instructions in conversation to raters to minimize the use of the response option 'N/A'
Round 2: PFOA and fetal growth	1) Additional questions added to Domain 3. Bias in measurement of exposure to assess the exposure: a. "Is there a concern that the variation in exposure levels across groups was insufficient to potentially identify associations with health outcomes?" b. "Is there a concern that the exposure assessment did not capture the relevant time window of exposure with respect to the health outcome?" c. "Are there concerns that missing exposure data (including methods used to input data) may have resulted in exposure misclassification?" 2) Additional question added to Domain 3. Bias in measurement of exposure to assess temporality of exposure and outcome measurements: a. "Was information on exposure status recorded prior to outcome assessment?"
Round 3: PDBE and thyroid function	1) Additional fields added to Step I of the instrument: a. "List the criterial used to determine the accuracy of exposure measurement" b. "List the possible co-exposures that could differ between exposure groups and could have an impact on study outcomes" 2) Additional fields added to Step II of the instrument: a. "List the criteria used to determine the accuracy of exposure measurement"

<table>
<tr><td></td><td>b. “Factors to consider when evaluating health outcome assessment”</td></tr>
<tr><td>Consultation with topic-specific experts and ROBINS-I instrument developers</td><td>1) Discussions with topic-specific experts and comparison across instruments led to modifications made to the wording of questions in Domain 3: Bias in measurement of exposure and the inclusion of an additional question (3.7):
<table>
<tr><th>ROBINS-I (Bias in classification of intervention)</th><th>Modified instrument for assessing RoB in environmental exposure studies (Bias in measurement of exposure)</th></tr>
<tr><td>3.1 Is the intervention well defined?</td><td>3.1 Is exposure status well defined?</td></tr>
<tr><td>3.2 Was information on intervention status recorded at the time of intervention?</td><td>3.2 Did entry into the study begin with start of the exposure?</td></tr>
<tr><td>3.3 Was information on intervention status unaffected by knowledge of the outcome or risk of the outcome?</td><td>3.3 Was information on exposure status recorded prior to outcome assessment?</td></tr>
<tr><td></td><td>3.4 Could classification of exposure status have been affected by knowledge of the outcome or risk of the outcome?</td></tr>
<tr><td></td><td>3.5 Are the levels, duration, or range of exposure of the population at risk sufficient or adequate to detect an effect of exposure?</td></tr>
<tr><td></td><td>3.6 Is the follow-up period adequate to allow for the development of the outcome of interest?</td></tr>
<tr><td></td><td>3.7 Were exposure assessment methods robust (including methods used to input data)?</td></tr>
</table>
2) Discussions with ROBINS-I instrument developers lead to the following modifications:
a. Reorganization of questions 3.5 and 3.6 into the project- and study-level protocols as measures to assess indirectness and study eligibility, not RoB
b. Agreement of replacing ‘intervention’ with ‘exposure’ throughout the instrument; replacement of ‘target trial’ with ‘target experiment’; expansion of future guidance to distinguish between ROBINS for intentional interventions and modified ROBINS for unintentional exposures; and expansion of future guidance to highlight scenarios specific to environmental and occupational exposures.</td></tr>
</table>

Appendix D. Risk of Bias Instrument for Non-randomized Studies of Exposure

The risk of bias instrument for non-randomized studies of exposure

Step I: At the review level

Specify the research question

Participants	
Experimental exposure	
Control exposure	

List the confounding domains relevant to all or most studies

List the possible co-exposures that could differ between exposure groups and could have an impact on study outcomes

List the criteria used to determine the accuracy of exposure measurement

Factors to consider when evaluating health outcome assessment

Step II: For each study

Specify a target experiment specific to the study:

☐	The protocol-specified target experiment fully applies	***OR***	Participant	
			Experimental exposure	
			Control exposure	

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of exposure.

Is your aim for this study…?

- ☐ to assess the effect of initiating exposure (as in an intention-to-treat analysis)
- ☐ to assess the effect of initiating and adhering to exposure (as in a per-protocol analysis)
- ☐ other (specify)

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, 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.

Preliminary consideration of confounders

Complete a row for each important confounding area (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding areas are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the exposure. "Validity" refers to whether the confounding variable or variables fully measure the area, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding areas listed in the review protocol				
Confounding area	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding area measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is adjusting for this variable (alone) expected to move the effect estimate up or down?
			Yes / No / No information	Favor intervention / Favor control / No information

(ii) **Additional confounding areas relevant to the setting of this particular study, or which the study authors identified as important**				
Confounding area	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding area measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is adjusting for this variable (alone) expected to move the effect estimate up or down?
			Yes / No / No information	Favor intervention / Favor control / No information

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of exposure; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of criteria used to determine the accuracy of measurement of exposure and outcome

Complete a row for each measure listed in the study for the (i) exposure and (ii) outcome. Of the measures listed in the protocol, consider the sensitivity, specificity, and confidence in the methods used in the study.

(i) Exposure measurement method listed in the study		
Method of measurement	Measured exposure	Is the exposure measured validly and reliably by this method (or these methods)?
		Yes / No / No information

(ii) Outcome measurement method listed in the study		
Method of measurement	Measured outcome	Is the outcome measured validly and reliably by this method (or these methods)?
		Yes / No / No information

Preliminary consideration of co-exposures

Complete a row for each important co-exposure (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

"Important" co-exposures are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the exposure.

(i) Co-exposures listed in the review protocol		
Co-exposure	Is there evidence that controlling for this co-exposure was unnecessary (e.g., because it was not administered)?	Is presence of this co-exposure likely to favor outcomes in the experimental or the control group
		Favor experimental / Favor comparator / No information
		Favor experimental / Favor comparator / No information
		Favor experimental / Favor comparator / No information

(ii) Additional co-exposures relevant to the setting of this particular study, or which the study authors identified as important		
Co-exposure	Is there evidence that controlling for this co-exposure was unnecessary (e.g., because it was not administered)?	Is presence of this co-exposure likely to favor outcomes in the experimental or the control group
		Favor experimental / Favor comparator / No information
		Favor experimental / Favor comparator / No information
		Favor experimental / Favor comparator / No information

Step III: For each study: risk of bias assessment

Risk of bias assessment (cohort-type studies)

Bias due to confounding	1.1 **Is there potential for confounding of the effect of exposure in this study? If N or PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signaling questions need be considered	Y / PY / PN / N	[Description]
	If Y/PY to 1.1, answer 1.2 and 1.3 to determine whether there is a need to assess time-varying confounding:		
	1.2. **<u>If Y or PY to 1.1</u>:** Was the analysis based on splitting follow up time according to exposure received? **If N or PN to 1.2**, answer questions 1.4 to 1.6, which relate to baseline confounding	NA / Y / PY / PN / N / NI	[Description]
	1.3. **If Y or PY to 1.2**: Were exposure discontinuations or switches likely to be related to factors that are prognostic for the outcome?	NA / Y / PY / PN / N / NI	[Description]
	If N or PN to 1.3, answer questions 1.4 to 1.6, which relate to baseline confounding		
	1.4. Did the authors use an appropriate analysis method that adjusted for all the critically important confounding areas?	NA / Y / PY / PN / N / NI	[Description]
	1.5. **If Y or PY to 1.4**: Were confounding areas that were adjusted for measured validly and reliably by the variables available in this study?	NA / Y / PY / PN / N / NI	[Description]
	1.6. Did the authors avoid adjusting for post-exposure variables?	NA / Y / PY / PN / N / NI	[Description]

	If Y or PY to 1.3, answer questions 1.7 and 1.8, which relate to time-varying confounding and to baseline confounding		

	1.7. Did the authors use an appropriate analysis method that adjusted for all the critically important confounding areas and for time-varying confounding?	NA / Y / PY / PN / N / NI	[Description]
	1.8. **If Y or PY to 1.7**: Were confounding areas that were adjusted for measured validly and reliably by the variables available in this study?	NA / Y / PY / PN / N / NI	[Description]
	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to confounding?	Favors experimental / Favors comparator / Unpredictable	[Rationale]
Bias in selection of participants into the study	2.1. Was selection of participants into the study (or into the analysis) based on variables measured after the start of the exposure? **<u>If N or PN to 2.1 go to 2.4</u>**	Y / PY / PN / N / NI	[Description]
	2.2. **<u>If Y/PY to 2.1:</u>** Were the post-exposure variables that influenced selection associated with exposure?	Y / PY / PN / N / NI	[Description]
	2.3. **<u>If Y/PY to 2.2:</u>** Were the post-exposure variables that influenced eligibility selection influenced by the outcome or a cause of the outcome?	NA / Y / PY / PN / N / NI	[Description]
	2.4 Do start of follow-up and start of exposure coincide for most participants?	NA / Y / PY / PN / N / NI	[Description]
	2.5 If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?	NA / Y / PY / PN / N / NI	[Description]

	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to selection of participants into the study?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]
Bias in classification of exposures	3.1 Is exposure status well defined?	Y / PY / PN / N / NI	[Description]
	3.2 Did entry into the study begin with start of the exposure?	Y / PY / PN / N / NI	[Description]
	3.3 Was information used to define exposure status recorded prior to outcome assessment?	Y / PY / PN / N / NI	[Description]
	3.4 Could classification of exposure status have been affected by knowledge of the outcome or risk of the outcome?	Y / PY / PN / N / NI	[Description]
	3.5 Were exposure assessment methods robust (including methods used to input data)?	Y / PY / PN / N / NI	[Description]
	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to measurement of outcomes or exposures?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]
Bias due to departures from intended exposures	4.1. Is there concern that changes in exposure status occurred among participants? **If your aim for this study is to assess the effect of initiating and adhering to an exposure (as in a per-protocol analysis), answer questions 4.2 and 4.3, otherwise continue to 4.4 if Y or PY to 4.1.**	Y / PY / PN / N / NI	[Description]
	4.2. Did many participants switch to other exposures?	Y / PY / PN / N / NI	[Description]

	4.3. Were the critical co-exposures balanced across exposure groups?	Y / PY / PN / N / NI	[Description]
	4.4. **<u>If NY/PN PY to 4.1, or Y/PY to 4.2, or 4.3:</u>** Were adjustment techniques used that are likely to correct for these issues?	NA / Y / PY / PN / N / NI	[Description]
	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to departures from the intended exposures?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]
Bias due to missing data	5.1 Were there missing outcome data?	Y / PY / PN / N / NI	[Description]
	5.2 Were participants excluded due to missing data on exposure status?	Y / PY / PN / N / NI	[Description]
	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Y / PY / PN / N / NI	[Description]
	5.4 **<u>If Y/PY to 5.1, 5.2 or 5.3:</u>** Are the proportion of participants and reasons for missing data similar across exposures?	NA / Y / PY / PN / N / NI	[Description]
	5.5 **<u>If Y/PY to 5.1, 5.2 or 5.3:</u>** Were appropriate statistical methods used to account for missing data?	NA / Y / PY / PN / N / NI	[Description]
	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to missing data?	Favors experimental / Favors comparator / Towards null	[Rationale]

		/Away from null / Unpredictable	
Bias in measurement of outcomes	6.1 Could the outcome measure have been influenced by knowledge of the exposure received?	Y / PY / PN / N / NI	[Description]
	6.2 Was the outcome measure sensitive?	Y / PY / PN / N / NI	[Description]
	6.3 Were outcome assessors unaware of the exposure received by study participants?	Y / PY / PN / N / NI	[Description]
	6.4 Were the methods of outcome assessment comparable across exposure groups?	Y / PY / PN / N / NI	[Description]
	6.5 Were any systematic errors in measurement of the outcome unrelated to exposure received?	Y / PY / PN / N / NI	[Description]
	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to measurement of outcomes?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]
Bias in selection of	Is the reported effect estimate likely to be selected, on the basis of the results, from…?		
the reported result	7.1. … multiple outcome *measurements* within the outcome domain?	Y / PY / PN / N / NI	[Description]
	7.2 … multiple *analyses* of the exposure-outcome relationship?	Y / PY / PN / N / NI	[Description]
	7.3 … different *subgroups*?	Y / PY / PN / N / NI	[Description]

	Risk of bias judgement	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the predicted direction of bias due to selection of the reported result?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]
Overall bias	**Risk of bias judgement**	Low / Moderate / Serious / Critical / NI	[Support for judgement]
	Optional: What is the overall predicted direction of bias for this outcome?	Favors experimental / Favors comparator / Towards null /Away from null / Unpredictable	[Rationale]