Synthesis methods data dictionary

No.	Reporting item	Item question, listed response options	Notes and examples
CHARACTERISTICS			
1.1	Source database	• HE	Health Evidence
		• HSE	Health Systems Evidence
1.2	Database quality score	Enter the quality score assigned by the source database.	**Note:** The Health Evidence (HE) database uses an in-house tool, assigning all reviews a score out of 10 points (Health Evidence 2018). Reviews scoring 8 or higher are characterised by the tool as strong, scores between 5-7 are characterised as moderate, and lower scoring reviews are characterised as weak. Quality scores for HE records are in the record metadata in EPPI-Reviewer. The Health Systems Evidence (HSE) database uses the AMSTAR tool (Shea, Grimshaw et al. 2007), assigning a score out of 11 points, with no descriptors assigned. Quality scores for HSE records are found by following the URL in the EPPI-Reviewer record to the database record.
1.3	Review question/objectives	Statement of the review question or objectives in PICO format if possible (e.g. collect the title or objectives of a Cochrane review, or write similar free text statement)	**Note:** Objectives are preferable if available in a clearly stated form, as this may capture secondary objectives that can be helpful in understanding the intention to create comparisons later in this form.
1.4	Number included studies	Enter the number of included studies	
1.5	Guidance	Did the review cite any specific source of guidance for their overall review methods?	**Note:** We are looking for general systematic review methods. Do not code for methods for specific tools or techniques such as a RoB tool or statistical method.
		• Cochrane review	The review was published in the *Cochrane Database of Systematic Reviews of Interventions*
		• Cochrane methods (but not a Cochrane review)	**Note:** No assessment required as to whether methods have been adhered to. **EXAMPLE:** specific reference to Cochrane methods or Cochrane MECIR standards, citation of *Cochrane Handbook for Systematic Reviews of Interventions*, or methods published by a Cochrane Review Group
		• JBI review	The review was published by JBI (the Joanna Briggs Institute.
		• JBI methods (but not a JBI review)	**EXAMPLE:** The review cited the JBI Reviewer's Manual (or a similar official JBI methods publication).
		• PRISMA	**Note:** This option should be coded whenever the PRISMA checklist or guidance paper is cited, regardless of whether it is noted as reporting or conduct guidance. Referring to or presenting a PRISMA flow diagram is not sufficient.
		• Other (describe)	The review cited other sources of methodological or reporting guidance. **EXAMPLE:** Campbell Collaboration, Task Force on Community Preventive Services, MOOSE Reporting Guidelines for Meta-analyses of Observational Studies or other published methodological papers.
		• No	No guidance is cited.
		• Unclear	Some information is provided, but it is unclear whether it is a citation of general methods guidance.

No.	Reporting item	Item question, listed response options	Notes and examples
1.6	Protocols & registration		**Note:** Protocols and registry records are noted, but are not generally required for coding. The only exception is if the text of the review refers explicitly to the protocol for details of a specific method needed for coding **EXAMPLE:** "See protocol for full taxonomy of intervention groupings."
1.6.1	Registry record	Does the review cite a registry record?	
		• Yes (add details)	If yes, collect registry ID or URL.
		• Yes but not available	
		• No	There is no registry record.
		• No information	No information is provided to indicate whether a registry record exists.
1.6.2	Protocol	Does the review cite a protocol? (or is it a Cochrane or Campbell review for which we know a protocol is available?)	**Note:** For updates, a previously published review can perform the function of a protocol. Note that a record in PROSPERO should be recorded in the question above as a registry record, not a protocol (although the authors may describe this as a protocol).
		• Yes (add details)	If yes, collect citation and URL.
		• Yes but not available	**Note:** This may include any protocol that is referred to but not publicly available (including those that may be available on request from the authors).
		• No	No protocol is available.
		• No information	No information is provided to indicate whether a protocol exists.
POPULATION			
2.1	Population for review	Population characteristics for inclusion of studies in the review.	Capture verbatim detail to as evidence of accuracy of interpretation - summary version can be extrapolated later, e.g. from review question captured above.
2.1.1	Eligible population		**Note:** e.g. from title or eligibility criteria. Use more detailed description in eligibility criteria only if it further clarifies the scope of the review.
2.1.2	Population setting	Capture any aspects of setting that define the population eligible for this review (e.g. rurality, socioeconomic status, LMIC, etc.).	**Note:** If no setting features related to the population are specified, enter 'any'. Note that a separate question is asked later about aspects of the intervention setting, e.g. hospital, community, etc.
INTERVENTIONS			
3.1	Intervention for review	Overall intervention characteristics for inclusion of studies in the review.	Capture verbatim detail to as evidence of accuracy of interpretation - summary version can be extrapolated later, e.g. from review question captured above.
3.1.1	Eligible interventions	Intervention characteristics of the overall review	
3.1.2	Intervention category	• Health systems and delivery of care	**EXAMPLE:** Interventions to improve prescribing for older people receiving polypharmacy. Although the study aims to improve medication prescribing, and medications are treatments for disease, intervention whose effectiveness is being tested aimed at improving the delivery of care, rather than testing the effects of the medications themselves.
		• Consumer communication and	**Note:** This category can also be used to code interventions that focus on consumer experience of a

No.	Reporting item	Item question, listed response options	Notes and examples
		participation	condition, or the experience of carers and families.
		• Prevention, treatment or management of a health condition	**Note:** In some manner, any review in this study could be considered relevant to the 'prevention, treatment or management of a health condition'. Rather than selecting this code for every review, aim to select one category that most closely describes the direct action of the intervention.
		• Population health	
		• Other	
		• Unclear	
OUTCOMES			
5.2	Outcomes for synthesis	Grouping of outcome characteristics for synthesis within the review.	
5.2.1	What was the first primary outcome?	Specify the first primary outcome of in the review	Selected as follows: (1) The first specified primary outcome (2) The first outcome in the statement of objectives/hypothesis (3) The first outcome in the Methods (4) The first outcome in the Results. **Note:** The following questions will only be answered in relation to this outcome, as responding for multiple separate outcomes introduces complexity in the coding of unrelated outcomes that can obscure outcome groupings. For the purposes of this question, the word 'outcome' can be interpreted as an outcome, a broad domain including multiple outcomes, or a specific outcome measure - the outcome should be collected exactly as stated in the review, regardless of breadth or level.
STUDY DESIGNS			
6.1	Study designs for the review	Study design characteristics for inclusion of studies in the review	
6.1.1	Eligible study designs	What study designs were eligible for inclusion in the review?	Code this item based on any study design criteria listed as part of the review's eligibility criteria. If no study design criteria are stated, infer this from the included study designs.
		• RCTs only	This category includes quasi-randomised as well as truly randomised trials.
		• NRS only	
		• Both RCTs and NRS	
		• Unclear	
SYNTHESIS METHODS			
7.1	Main approach used in the review		This question to be coded on what is used in the review in practice. Purpose of this question is a high-level indicator to us of what we will find used in the review.
		• MA for all syntheses	
		• MA for most syntheses	
		• A mix of MA and other	
		• Other for most syntheses	

No.	Reporting item	Item question, listed response options	Notes and examples
		• Other for all syntheses	
		• No synthesis	
		• Other (describe)	
		• Unclear	
7.2	Methods specified		Code if methods were specified/described. **Note:** For this question, we don't need the description to include exhaustive details required to replicate the synthesis, e.g. all specifications used in statistical software, as long as the synthesis method used is clear. Methods such as trim and fill, used to explore or account for publication bias, are not of interest here (they are additions or adjustments to the main synthesis method).
		• Meta-analysis of effect estimates	Code if meta-analysis is specified. **Note:** This can include an explicit statement is made in the methods that MA could not be used due to characteristics of the included studies (e.g. heterogeneity, outcome reporting, indicating a post hoc decision once data were available). **EXAMPLE:** "Due to the heterogeneity of study continuity and mortality measurements, it was not possible to combine them to produce an estimate of effect size."
		• Combining P values (describe specific method)	
		• Vote counting - direction of effect	Vote counting using the direction of effect means that authors are counting the direction of the effect estimate from each study, regardless of whether it is statistically significant or not.
		• Vote counting - statistical significance	Vote counting using statistical significance means that authors are counting only statistically significant effects, and treating those that are not statistically significant as no direction/no effect.
		• Vote counting - other/unclear	Vote counting of some kind is specified, but it is unclear what basis will be used to 'count' studies as being in one direction or another, or how a decision will be made about the overall direction of evidence.
		• Binomial or sign test	Was a sign(binomial) test specified to assess the statistical significance of evidence for the existence of an effect?
		• Median - weighted	Were effect estimates to be summarised using the median to estimate the centre of the distribution, and was weighting used in this calculation?
		• Median - no weighting	Were effect estimates to be summarised using the median to estimate the centre of the distribution, but without the use of weighting?
		• Median - unclear weighting	Were effect estimates to be summarised using the median to estimate the centre of the distribution, but it was unclear whether weighting was used in this calculation?
		• Interquartile range	Were effect estimates to be summarised as an interquartile (IQ) range?
		• Range	Were effect estimates to be summarised as a range?
		• 'Narrative synthesis' with no further detail	The authors have stated that they will use "narrative synthesis", or similar terms such as "narrative summary" or "qualitative synthesis", without elaborating what this means in detail.

No.	Reporting item	Item question, listed response options	Notes and examples
		• Other (describe)	
		• No synthesis	The authors stated that no synthesis methods would be used - individual study results listed or summarised in tables or text
		• Unclear	
		• No information	No information was provided specifying what synthesis methods would be used in the review (i.e. in the Methods section or ahead of the reporting of any results).
7.3	Methods used in practice		Code these methods if actually used in practice. **Note:** Methods such as trim and fill, used to explore or account for publication bias, are not of interest here (they are additions or adjustments to the main synthesis method).
		• Meta-analysis of effect estimates	
		• Combining P values (describe specific method)	
		• Vote counting - direction of effect	Vote counting using the direction of effect means that authors are counting the direction of the effect estimate from each study, regardless of whether it is statistically significant or not. **Note:** Informal summary sentences appearing in the Discussion are not sufficient to code as vote counting as a method in the analysis. **EXAMPLE:** A statement that “48 out of 56 studies showed an improvement in screening attendance”, where it is clear that statistical significance has not been used to define improvement (e.g. based on the text, or by comparing to a results table).
		• Vote counting - statistical significance	Vote counting using statistical significance means that authors are counting only statistically significant effects, and treating those that are not statistically significant as no direction/no effect. **Note:** Informal summary sentences appearing in the Discussion are not sufficient to code as vote counting as a method in the analysis. **EXAMPLE:** A statement that "The majority of individual studies reported significant increases in physical activity following the intervention (42 out of a possible 60).”.
		• Vote counting - other/unclear	Vote counting of some kind is being used, but it is unclear what basis is being used to 'count' studies as being in one direction or another, or how a decision was made about the overall direction of evidence. **Note:** Informal summary sentences appearing in the Discussion are not sufficient to code as vote counting as a method in the analysis. **EXAMPLE:** the authors may conclude that the evidence "overall" supports a particular conclusion, without explaining how they reconcile differences in the available data.
		• Binomial or sign test	Was a sign(binomial) test used to assess the statistical significance of evidence for the existence of an effect?
		• Median - weighted	Effect estimates were summarised using the median to estimate the centre of the distribution, but no weighting was used.
		• Median - no weighting	Were effect estimates summarised using non-parametric statistics (e.g. median, IQ range, range)? Use this option if weighting was not used.

No.	Reporting item	Item question, listed response options	Notes and examples
		• Median - unclear weighting	Effect estimates were summarised using the median to estimate the centre of the distribution, but it was unclear whether weighting used in this calculation.
		• Interquartile range	Were effect estimates to be summarised as an interquartile (IQ) range?
		• Range	Were effect estimates to be summarised as a range?
		• Other (describe)	
		• No synthesis	No synthesis methods were used - individual study results listed or summarised in tables or text.
		• Unclear	
7.4	Hierarchy?	Was a hierarchy of preferred synthesis methods described?	**Note:** A statement that "narrative synthesis" or similar will be used if MA not possible does not constitute a hierarchy, as 'narrative synthesis' is not really a specific synthesis method.
		• Yes	
		• No	
		• Unclear	
		• n/a	No synthesis methods described or used
7.4.1	If yes, was hierarchy linked to type of data available?	Was this hierarchy explicitly linked to the type of data available from the included studies?	**Example:** Effect estimates, only P values, etc.
		• Yes	
		• No	
		• Unclear	
		• n/a	No hierarchy
7.5	Rationale?	Was a rationale given for the synthesis method intended or used?	
		• Yes - rationale for MA	The authors gave a rationale for the use of meta-analysis. **Note:** Do not code for statements about the model selected without a reason. **EXAMPLE:** For a particular model (e.g. random-effects used for reasons related to heterogeneity), or a description of the circumstances under which meta-analysis would proceed (e.g. homogeneity).
		• Yes - rationale against MA	The authors gave a reason why meta-analysis could not be conducted.
		• Yes - rationale for other method	The authors gave a positive reason why a method other than synthesis was used. **Note:** Do not code for negative reasons why meta-analysis could not be used.
		• Yes - rationale against other method	The authors gave a reason why a method other than meta-analysis could not be used.
		• No	
		• Unclear	
		• n/a	No summary or synthesis methods specified.
7.6	Selection between multiple outcomes/measures?	In the presence of multiplicity (multiple outcomes within a domain or measurement methods/tools within an	**Note:** The method can be prespecified or post hoc, but must be described somewhere in the text. Do not code based on inferring from use in practice. We are interested in here in selecting between multiple measures within a PICO grouping - e.g. multiple measures within a stated timeframe category, or

No.	Reporting item	Item question, listed response options	Notes and examples
		outcome), was a method explicitly stated for selecting between the eligible options?	multiple eligible measures for an outcome. Other cases, such as a decision to prefer change from baseline over end point measures, or selection between multiple intervention arms or comparisons within a study, are not of interest and should not be coded.
		• Yes - selection between outcomes within a domain	**EXAMPLE:** 'Anxiety' and 'depression' in a 'mental health' domain'.
		• Yes - selection between measures within an outcome	**EXAMPLE:** Selection between multiple eligible measurement tools for depression.
		• Yes - selection of time points within an outcome timeframe	**EXAMPLE:** "If studies provided data for multiple follow-up time points, we extracted data for the time furthest from baseline."
		• Yes - other (describe)	
		• Yes - analysis method includes more than one effect estimate per study and accounts for dependency	**EXAMPLE:** "a three-level meta-analysis technique was used. By applying a multilevel approach to meta-analysis, there is no need for aggregating or selecting data, implying that all relevant effect sizes can be extracted from primary studies (see also Assink et al. 2015; Assink and Wibbelink 2016)."
		• No - reported all outcomes/measures	The review explicitly stated that all outcomes from each included study were reported.
		• No - reported multiple outcomes/measures but unclear if all	The review appeared to report multiple outcomes measures from included studies, and did not appear to be selective about reporting significant outcomes, giving the impression of being complete, although there's nothing explicit to state that this was in fact complete.
		• Unclear	
		• No information	
7.6.1	If a selection mechanism was used, what was the basis for selection?	• Order of reporting	Authors selected results based on the order in which they were reported in the original paper. **EXAMPLE:** "If studies used multiple questionnaires measuring one outcome measure (e.g., both CES-D and HADS-D were used to measure depression), only the first described questionnaire was included in the analyses."
		• Hierarchy of preferred outcomes/measures	A list of possible options is reported with a preferred order for selection depending on which options are available from the included studies. **EXAMPLE:** A list of measurement tools for Depression with a preferred order.
		• Mean or median of available results	A result is selected or calculated to reflect the centre of the set of available results. **EXAMPLE:** the mean or median.
		• Clinical relevance	Results for inclusion in synthesis were chosen for reasons of relevance to the research question. **Note:** Studies that select results on the basis of time (e.g. the measure taken closest or furthest from the end of the intervention) should be given this code, as this is likely done for clinical reasons, such as to distinguish either an immediate or lasting clinical effect (e.g. "Since some studies contained more than one follow-up time point, we chose the follow-up time point which was closest to 6-12 wk postpartum, as this period has the highest risk for the development of postpartum depression.").
		• Core outcome set	The results selected are based on an established core outcome set for the condition or intervention of

No.	Reporting item	Item question, listed response options	Notes and examples
			interest to the review.
		• Methodological reasons	Methodological reasons include selection designed to improve the accuracy of the effect estimate. **EXAMPLE:** Where outcome measures are selected for adjustment of confounders (e.g. "We selected reported study estimates that adjusted for potential confounding variables for inclusion in meta-analysis over reported estimates that did not adjust for potential confounding variables."); or for completeness of data (e.g. "When two or more papers contained the same or some of the same study population, the paper that described the largest population was used.").
		• Consistency with other included studies	The results are selected based on which measures, interventions or comparators are most directly comparable to other included studies in the synthesis.
		• Other (describe)	
		• No information	A selection mechanism was used, but no information was provided to describe the basis for selection.
		• Unclear	Some information was provided, but the basis for selection remained unclear.
		• n/a	No methods relating to multiplicity were observed in the methods or results of the review.
7.7	Heterogeneity		
7.7.1	Methods specified	Were any methods specified to investigate or deal with heterogeneity in the review?	**Note:** Statistics such as I^2, chi^2 to measure heterogeneity, or the use of a random-effects meta-analysis model, are not considered methods to investigate or deal with heterogeneity and should not be coded.
		• Subgroup analysis	
		• Meta-regression - using inverse variance weighting	**Note:** Terms such as 'random-effects model' indicate the use of inverse-variance weighting in meta-regression, as per use in meta-analysis.
		• Meta-regression - other or unclear weighting	
		• Prediction intervals	**Note:** Equivalent term: predictive interval
		• Non-parametric methods	Were non-parametric methods specified to investigate heterogeneity in the review? **EXAMPLE:** Kruskal-Wallis one way ANOVA (or Mann-Whitney U test for two groups)
		• Other (describe)	
		• No	No methods were specified to investigate or deal with heterogeneity.
		• Unclear	Some information was provided but it remains unclear.
7.7.2	Methods used in practice	Were methods used in practice to investigate or deal with heterogeneity in the review?	**Note**: Statistics such as I2, chi2 to measure heterogeneity, or the use of a random-effects meta-analysis model, are not considered methods to investigate or deal with heterogeneity and should not be coded.
		• Subgroup analysis	
		• Meta-regression - using inverse variance weighting	**Note:** Terms such as 'random-effects model' indicate the use of inverse-variance weighting in meta-regression, as per use in meta-analysis.
		• Meta-regression - other or unclear weighting	

No.	Reporting item	Item question, listed response options	Notes and examples
		• Prediction intervals	**Note:** Equivalent term: predictive interval
		• Non-parametric methods	Were non-parametric methods used to investigate heterogeneity in the review? **EXAMPLE:** Kruskal-Wallis one way ANOVA (or Mann-Whitney U test for two groups)
		• Other (describe)	
		• No	No methods were used to investigate or deal with heterogeneity.
		• Unclear	Some information was provided, but it is unclear.
		• n/a	No heterogeneity present, so methods for exploring heterogeneity were not required.
7.8	Changed plans		
7.8.1	Authors' comments on changed plans		**Note:** This set of three questions relates to explicit statements by the authors about synthesis methods they wanted to use and couldn't. There is no need to code these questions for other changes in methods not explicitly noted by the authors.
7.8.1.1	Authors stated synthesis methods could NOT be applied as planned	Did the authors state that one or more of the intended synthesis methods could NOT be applied as planned?	**Note:** Not having enough studies to conduct a planned analysis should not be coded here - this question assumes that even if there were enough studies, the planned synthesis method could not proceed for another reason. If it relates to a comparison, not having enough studies can be coded as a PICO grouping that did not proceed (item for companion study, not reported in this data dictionary). Not having sufficient studies for a subgroup or other secondary analysis does not need to be coded.
		• Yes	
		• No	
		• Unclear	Some information was provided, but it is unclear.
		• n/a	No synthesis methods were specified.
7.8.1.2	If yes, what were the reasons?	If the authors stated a synthesis method could not proceed, what were their reasons?	
		• Not enough studies available	**EXAMPLE:** None or only 1 or 2 in a planned group.
		• Outcome not reported	Studies available that meet criteria and measured the outcome, but data not reported.
		• Incompletely reported outcome or effect estimate	Outcome was reported, but not in enough detail to use for planned synthesis.
		• Different outcome measures across studies	Authors reported that differences in the outcomes measured prevented synthesis. **EXAMPLE:** "Studies reported different maternal health knowledge outcome categories which could not be combined in a meta-analysis."
		• Different effect measures across studies	Authors reported that differences in effect measures used to report results prevented synthesis. **EXAMPLE:** "different statistical estimates used to report associations between continuity of care and mortality."
		• Bias in included studies	Authors reported that results included in the synthesis for one or more studies were at high risk of bias and synthesis would not be appropriate.
		• Bias due to missing results	Authors were concerned about bias due to missing results from a synthesis (i.e. they suspected that the

No.	Reporting item	Item question, listed response options	Notes and examples
			outcome of interest was unavailable from one or more studies due to selective reporting), so decided the synthesis would not be appropriate.
		• Clinical diversity (P or I)	Authors decided the studies were too diverse in population or intervention, and synthesis would not be appropriate.
		• Methodological diversity	Authors stated that study designs were too diverse (e.g. RCTs and NRSIs), and synthesis would not be appropriate.
		• Statistical heterogeneity	Observed statistical heterogeneity was high (e.g. measured using I-squared), and authors decided synthesis would not be appropriate.
		• Other (describe)	
		• Unclear	Some information was provided, but it is unclear.
		• No information	No information was provided on the reasons a planned synthesis method that could not proceed.
		• n/a	The authors did not state that a planned synthesis method could not proceed.
7.8.1.3	If yes, what action did they take?	If the authors stated a grouping could not take place, did they take any action or make changes to their analysis in response?	
		• Combined the planned PICO groupings at a higher/broader level	This may have been in accordance with a planned contingency for sparse data.
		• Split the planned PICO grouping to a narrower/more specific level	
		• Did not conduct any synthesis for a planned comparison	
		• Other (describe)	
		• Unclear	Some information was provided, but it is unclear.
		• No information	No information was provided on what approach was taken in response to a planned synthesis method that could not proceed.
		• n/a	The authors did not state that a planned synthesis method could not proceed.
7.8.2	Could they have done more?	Was there the potential to do more synthesis than the authors did? Could the stated reasons for not proceeding have been overcome with readily available methods?	**Note:** This is a judgement call. Acknowledge that we don't have specific clinical knowledge in every area, or perfect knowledge of the available data.
		• Yes - more could have been done with the data.	It is clear that more could have been done with the available data to enable synthesis plans to appear possible. **Note:** Don't assume anything complex, or that would require the guidance of a statistician to realise it was possible.

No.	Reporting item	Item question, listed response options	Notes and examples
			EXAMPLE: Relatively straightforward transformations or calculations, readily imputed statistics, including calculating effect estimates from separate intervention group means.
		• Yes - could have grouped more	Yes, it is clear that synthesis could have proceeded if the authors had taken less of a 'splitting' approach, and been comfortable with more diversity in their analysis. **Note:** Give the benefit of the doubt - don't assume specialist knowledge of the field, but broader groupings appear clearly possible and not inappropriate. **EXAMPLE:** Where the outcome of physical activity is measured in different studies using steps per day, minutes per week, and METS minutes per day, it is straightforward to conduct a standardised analysis of these measures to generate a synthesised effect estimate for physical activity.
		• Yes - other (describe)	
		• No	
		• Unclear	It is uncertain whether the authors could have done more synthesis than they did.
		• n/a	The authors did not state that a planned synthesis method could not proceed.
7.9	Summary without synthesis		This section applies to the reporting of results from individual studies that are NOT included in a meta-analysis for any reason, including single studies that have been deemed not suitable/possible to include in the MA, outcomes/comparisons for which synthesis methods other than MA are used, and outcomes/comparisons for which no synthesis at all is conducted. **Note:** It is not necessary to code the reporting of individual study results that are included in MA (whether in the text or on forest plots).
7.9.1	Main approach to reporting study-level results in text	For reviews that presented the results of individual studies NOT included in a meta-analysis (whether other synthesis methods were used or not), what was the main approach to summarising the results of individual studies in the text of the review?	**Note:** Only one option should be selected.
		• Effect estimates at the study level	**EXAMPLE:** RR, OR, MD (with or without measures of precision (e.g. CIs), etc.
		• Summary statistics at the intervention group level	**EXAMPLE:** Intervention group summary statistics (e.g. mean (SD), control group mean (SD), medians, etc., presented separately, not used to calculate an effect estimate for the study as a whole.
		• Statements about statistical significance	The authors mainly made statements about whether or not results were statistically significant. (This is different to the selective reporting of results based on significance, captured in the next question). **EXAMPLE:** "Five studies reported health-related quality of life scores (Glasgow 2005; Glasgow 2006; Lorig 2010; Quinn 2011; Smith 2000) but none showed statistically significant differences. In one study, both the control and the intervention group showed improvement on the PAID-2 scale (Glasgow 2005) but there was no significant difference between the two groups at 12 months." [Pal, from complex interventions project]
		• Qualitative statements about	The authors used mainly statements with no quantification of effect estimates, where no numerical

No.	Reporting item	Item question, listed response options	Notes and examples
		direction of effect	results are reported, and it's not clear that the statement refers to statistical significance. **EXAMPLE:** "The e-health program resulted in significantly improving physical health outcomes with small to moderate effect sizes on primary health outcomes of patients with diabetes (refs). However, not all outcomes improved, and in some measures, comparable effect sizes were seen in the intervention and control group." [Eland-de Kok, from complex interventions project]
		• Study-level results not reported	Study-level results were available that were not included in a meta-analysis, and were not reported in the text.
		• Other (describe)	
		• A mix	No dominant approach was used, and different approaches were used for different studies. **Note:** Do not code individual approach options in addition to coding as 'A mix'. If more than one method was consistently used, the specific methods should be coded.
		• Unclear	It is uncertain what the main approach to reporting results in the text was.
		• n/a	No study-level results were available that were not included in meta-analysis. Reporting not required.
7.9.2	Main approach to reporting study-level results in tables or graphs	For reviews that summarised results without MA for at least one outcome, what was the main approach to summarising the results of individual studies in tables?	**Note:** This does not apply to the reporting of individual study data within a MA (e.g. on the forest plot). Only one option should be selected.
		• Effect estimates at the study level	**EXAMPLE:** RR, OR, MD, with or without measures of precision (e.g. CIs).
		• Summary statistics at the intervention group level	**EXAMPLE:** Intervention group m(SD), control group m(SD), median, range, without calculating an effect estimate for the study as a whole.
		• Statements about significance	The authors mainly made statements about whether or not results were statistically significant. (This is different to the selective reporting of results based on significance, captured in the next question).
		• Qualitative statements about direction of effect	The authors used mainly statements with no quantification of effect estimates, where no numerical results are reported, and it's not clear that the statement refers to statistical significance.
		• Study-level results not reported	Study-level results were available that were not included in a meta-analysis, and were not reported in tables or graphs.
		• Other (describe)	
		• A mix	No dominant approach was used, and different approaches were used for different studies. **Note:** Do not code individual approach options in addition to coding as 'A mix'. If more than one method was consistently used, the specific methods should be coded.
		• Unclear	It is uncertain what the main approach to reporting results in the tables was.
		• n/a	No study-level results were available that were not included in meta-analysis. Reporting not required.
7.9.3	Selective reporting of statistically significant results?		This question refers to the selective reporting of results that are statistically significant in the text or tables.

No.	Reporting item	Item question, listed response options	Notes and examples
		• Yes	Code 'yes' when there is clear evidence in all comparisons/outcomes that there are results available from multiple studies, but the subset of studies that did not show statistically significant effects were either not reported, were less completely reported, or were only reported as supplementary data. **EXAMPLE:** Only statistically significant results are reported numerically in the review, and other results are reported only as 'NS'. There is a statement in the text such as "three out of the five studies found a statistically significant effect", although no numerical results are report. The results table in the study reports numerical results only where they are statistically significant, and forms the basis of discussion in the text of the review. A supplementary file includes a more complete data table including non-significant results that is not discussed in the text of the review.
		• Yes for some outcomes (not all)	As above, but only done for some of the comparisons/outcomes reported in the review. Other comparisons/outcomes appear to have all results presented, irrespective of whether statistically significant effects were found.
		• No	There was no evidence of selective reporting of statistically significant results.
		• Unclear	It is unclear whether selective reporting of statistically significant results took place, but it is suspected. **EXAMPLE:** Only one outcome per study is reported, and the proportion that are statistically significant appears implausible, but there is no direct evidence to indicate selectivity. Classification of a study as significant appears to be based on selective use of specific intervention elements or subgroups, rather than the study as a whole, but this is not clear. Reporting of complete results appears to be more common in statistically significant results, but this may be determined by the information reported in the primary studies. Significance is not reported clearly enough to assess.
		• n/a	No study-level results were available that were not included in meta-analysis. Reporting not required.
7.9.4	Study-level effect estimates reported consistently where possible?	Did the authors report consistent data on the effect estimates for each study (where possible)?	
		• Yes	Code as 'yes' when it appears results across studies are consistently reported as effect estimates wherever possible, and there are no instances of inconsistent reporting that could readily be resolved. **EXAMPLE:** Results are reported consistently as effect estimates, except where no effect estimate was reported by an included study and one could not be calculated (e.g. a qualitative statement that there was no effect on an outcome).
		• No	Code as 'no' when there are instances of results that are not reported as effect estimates, but effect estimates were almost certainly available or could be readily calculated. **EXAMPLE:** Results for one or more studies are reported as the mean(SD) for each intervention group, from which an effect estimate could have been readily calculated by the review authors. Only qualitative statements were used to describe results in the text and tables for every study.
		• Unclear	There is some information that raises doubt as to whether effect estimates have been reported

No.	Reporting item	Item question, listed response options	Notes and examples
			consistently wherever possible, but it is unclear.
		• n/a	No study-level results were available that were not included in meta-analysis. Reporting not required.
7.10	Figures	Were figures used in the review?	**Note:** Do not code other figures such as logic models, clinical figures, PRISMA flow diagrams or RoB summaries not designed to convey the results of studies or synthesis. They are not relevant to this study.
		• Forest plot	
		• Box/whisker plot	
		• Bubble plot	
		• Albatross plot	
		• Harvest plot	
		• Effect direction plot	
		• Stacked bar plot	
		• Funnel plot	**Note:** Do not code if funnel plots are discussed (e.g. in the context of considering publication bias) but not actually presented in the review.
		• Other (describe)	
		• No synthesis or summary figures	There were no figures in the review used to display the results of included studies or synthesised results.
ROBIS			
8.1	ROBIS selected signalling questions		Assessment of seven items selected from the ROBIS tool (Whiting, Savovic, et al., 2018), as indicators of good practice in the core methods of a systematic review. Items should be coded in accordance with the ROBIS guidance (which sets a relatively high bar for conduct) on the basis that we are using these items to differentiate reviews of high methodological quality from those that might achieve an acceptable score on other 'quality tools' (i.e. those reported in the source databases for this study).
8.1.1	Were eligibility criteria unambiguous?	ROBIS signalling question 1.3: Were eligibility criteria unambiguous?	From ROBIS guidance: "Specific information about the characteristics of eligible studies must be provided, as far as possible avoiding any ambiguities about the types of study, population, interventions, comparators and outcomes. Criteria should be sufficiently detailed that the review could be replicated using the criteria specified."
		• Yes	
		• No	
		• Unclear	
8.1.2	Appropriate range of electronic sources searched?	ROBIS signalling question 2.1: Did the search include an appropriate range of database/electronic sources for published and unpublished reports?	From ROBIS guidance: "The assessor needs to judge what constitutes an appropriate range of databases. This will vary according to review topic. It is anticipated that at a minimum a MEDLINE and EMBASE search would be conducted. Searches of material published as conference reports should also be considered along with a search of research registers. Guidance on the appropriate range of databases can be found in SR guidance such as the Cochrane Handbook."
		• Yes	To code "yes", reviews should search at least Medline, EMBASE AND at least one trials register. **Note:** The Cochrane Central Register of Controlled Trials includes both Medline and EMBASE and is an acceptable alternative. If Embase not included, searching several alternative databases may be

No.	Reporting item	Item question, listed response options	Notes and examples
			acceptable. Judgement can be used. **EXAMPLE:** A review of maternity care did not search Embase, but searched a range of other databases: "PubMed, CINAHL, Lilacs, AJOL, WHO RHL, and Popline were searched, along with ongoing trials registers (ISRCT register, ICTRP register), and the White Ribbon Respectful Maternity Care Repository (that collects items relevant to respectful care or disrespect and abuse from around the world, including audits, service evaluations, and formal research). We also included regular AMDD monthly RMC updates, as they were issued." A review of studies assessing the impact of roll-out of a vaccine in South American countries did not search Embase, but did search the geographically focussed Embase. In addition, no trials register was searched, but the review's eligibility criteria did not include clinical trials, but rather population-level studies of hospitalisation rates that are unlikely to be registered.
		• No	
		• Unclear	
		• No information	No information is available on the sources searched.
8.1.3	Appropriate additional sources searched?	ROBIS signalling question 2.2: Were methods additional to database searching used to identify relevant reports?	From ROBIS guidance: "Additional methods such as citation searches, contacting experts, reference checking, handsearching etc. should have been performed."
		• Yes	To code "yes", reviews should search at least references of included studies, plus at least one additional source (i.e. references of other reviews, citations search for included studies, expert contacts, website searches for reports not published in journals).
		• No	Code "no" if the above criteria are not met, or only electronic searches are listed and no additional sources are stated as being searched.
		• Unclear	
		• No information	No information is available on the sources searched.
8.1.4	Efforts to minimise error in data collection?	ROBIS signalling question 3.1: Were efforts made to minimise error in data collection?	From ROBIS guidance: "In order to minimize bias and errors in the data collection process this should involve at least two reviewers. Ideally this should be done independently but extraction by one reviewer and detailed checking by a second reviewer is also acceptable."
		• Yes	
		• No	
		• Unclear	
		• No information	No information is available on the data collection process.
8.1.5	RoB assessed appropriately?	ROBIS signalling question 3.4: Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	From ROBIS guidance: "If risk of bias was not formally assessed then this question should be answered as "No". If a formal assessment was carried out then, assessors will need to use their judgement regarding whether the criteria used were appropriate. If an accepted published tool was used for the appropriate design, such as the Cochrane Risk of Bias tool for RCTs or QUADAS-2 for DTA studies, then this will be

No.	Reporting item	Item question, listed response options	Notes and examples
			fairly straightforward and this question can be answered as Yes."
		• Yes	Code as "yes" if reviews used Cochrane or similar domain-based assessment tool appropriate for included study designs. It may be acceptable not to use a domain-base tool for non-randomised studies - although ROBINS-I exists, it is not yet considered industry standard. **Note:** Other tools that assess similar domains (e.g. JBI) may be considered acceptable if reported by domain and not only as a numerical score. Some tools that can produce a numerical score (e.g. PEDRO), are widely used within some fields, and assessments can be reported in a way that is consistent with contemporary best practice in risk of bias assessment (i.e. by domain rather than as a score).
		• No	Code as "No" if a review used a tool that incorporates issues unrelated to bias; modified or invented their own tool; reported tools using only summary scores and did not report assessments by domain or item; or used a reporting guideline to assess bias (e.g. STROBE).
		• Unclear	
		• No information	No information is available on whether risk of bias was assessed.
8.1.6	Heterogeneity addressed?	ROBIS signalling question 4.4: Was between-studies variation (heterogeneity) minimal or addressed in the synthesis?	From ROBIS guidance: "This question targets variation in results of the studies rather than the variation in their characteristics.... If a random-effects model has been used appropriately to allow for heterogeneity and/or further subgroup/ meta-regression analyses run to explore heterogeneity, the assessor might answer “Yes”. However, random effects meta-analysis of studies that are extremely diverse in either characteristics or results may yield a meaningless result, particularly if the results of the studies point in opposing directions of effect."
		• Yes	To code "yes", reviews should both assess the presence of heterogeneity in the results and (if present) explore its causes (for example using subgroup analysis or meta-regression where meta-analysis is used, or otherwise structured discussion of results grouped by possible causal characteristics). Reviews that assessed results and found no heterogeneity can also be coded "yes".
		• No	**Note:** ROBIS guidance indicates that reviews that do not use meta-analysis in response to the presence of heterogeneity should be coded "yes". For this study, this is not considered an appropriate approach, and reviews should still explore the possible causes of heterogeneity, as this may identify explanations that can restore certainty in the synthesis (as per guidance on assessment of inconsistency using the GRADE approach, which focuses on unexplained heterogeneity, see Chapter 14, Cochrane Handbook 2019). Reviews that identify heterogeneity but do not explore its causes should be coded "no".
		• Unclear	
		• No information	No information is available on the assessment of or response to heterogeneity.
8.1.7	Were biases addressed in the synthesis?	ROBIS signalling question 4.6: Were biases in primary studies minimal or addressed in the synthesis?	From ROBIS guidance: "Assessors are encouraged to answer this question as “No” if they judge there to be important bias in constituent studies that has been ignored by the reviewers. For example, if risk of bias has not been evaluated in the systematic review, ROBIS assessors should answer “No”. Equally, if risk of bias has been assessed but reviewers have not incorporated it into findings/ conclusions this should also receive a “No” answer to this question." **Note:** To code "yes", reviews should (if risk of bias is present) either use risk of bias to structure

No.	Reporting item	Item question, listed response options	Notes and examples
			summaries or synthesis, or use risk of bias in the interpretation of results (e.g. through incorporation of bias into an assessment of the certainty of the evidence using GRADE (see Chapter 14, Cochrane Handbook 2019) or a similar approach. Reviews that assess studies and find all results at low risk of bias can also be coded "yes".
		• Yes	
		• No	
		• Unclear	
		• No information	There is no information available on whether risk of bias was addressed in the synthesis.