Journal of Clinical Epidemiology 163 (2023) 79–91

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Critical elements of synthesis questions are incompletely reported: survey of systematic reviews of intervention effects

Miranda S. Cumpston[a], Joanne E. McKenzie[a], Rebecca Ryan[b], James Thomas[c], Sue E. Brennan[d,*]

[a]*Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria 3004, Australia*
[b]*Centre for Health Communication & Participation, La Trobe University, Melbourne, Victoria 3086, Australia*
[c]*EPPI-Centre, UCL Social Research Institute, University College London, Gower St, London, WC1E 6BT, UK*
[d]*School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria 3004, Australia*

Accepted 26 September 2023; Published online 29 September 2023

Abstract

Objectives: To examine the characteristics of population, intervention and outcome groups and the extent to which they were completely reported for each synthesis in a sample of systematic reviews (SRs) of interventions.

Study design and setting: We coded groups that were intended (or used) for comparisons in 100 randomly sampled SRs of public health and health systems interventions published in 2018 from the *Health Evidence* and *Health Systems Evidence* databases.

Results: Authors commonly used population, intervention and outcome groups to structure comparisons, but these groups were often incompletely reported. For example, of 41 SRs that identified and/or used intervention groups for comparisons, 29 (71%) identified the groups in their methods description before reporting of the results (e.g., in the Background or Methods), 12 (29%) defined the groups in enough detail to replicate decisions about which included studies were eligible for each synthesis, 6 (15%) provided a rationale, and 24 (59%) stated that the groups would be used for comparisons. Sixteen (39%) SRs used intervention groups in their synthesis without any mention in the methods. Reporting for population, outcome and methodological groups was similarly incomplete.

Conclusion: Complete reporting of the groups used for synthesis would improve transparency and replicability of reviews, and help ensure that the synthesis is not driven by what is reported in the included studies. Although concerted effort is needed to improve reporting, this should lead to more focused and useful reviews for decision-makers.

Keywords: Systematic reviews; Evidence synthesis; Research questions; PICO; Reporting; Comparisons

1. Introduction

Systematic reviews (SRs) collate and synthesize research, and are used to inform decision making by clinicians, consumers, policy makers and researchers [1]. In SRs examining the effects of interventions, authors commonly define an overarching research question using the 'PICO' framework, in which the populations, interventions, comparators and outcomes are specified (the 'PICO for the review') [2].

Within a single SR, many research questions may be addressed. Reviews may aim to investigate the effects of a range of interventions (e.g., counseling, education or communication interventions for behavior change), or the effects of the same intervention in different populations (e.g., higher and lower socioeconomic groups). To address these questions, comparisons between interventions must be specified and the criteria that define each comparison considered (the 'PICO for each synthesis') [3]. The PICO for each synthesis can critically influence the findings of a SR because it determines which studies are eligible for each synthesis. Take, for example, a review of psychosocial interventions for smoking cessations described in Box 1. Without modifying the eligibility criteria for the SR as a whole, the synthesis may be structured to address different questions (broader or narrower) by grouping study characteristics and outcomes differently in comparisons.

The *Cochrane Handbook for Systematic Reviews of Interventions* [3,5] includes guidance for specifying the PICO

* Corresponding author. School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Rd, Melbourne, Victoria 3004, Australia. Tel.: +61-3-9903-0366; fax: +61-3-9903-0556.
E-mail address: sue.brennan@monash.edu (S.E. Brennan).

https://doi.org/10.1016/j.jclinepi.2023.09.013

What is new?

Key findings

- Methodological requirements for systematic reviews (SRs) have focused on specifying criteria for including studies in a review (the 'PICO for the review'), yet decisions about which groups of studies are eligible for each synthesis can have a critical impact on findings.
- We examined the characteristics of population, intervention and outcome groups and the extent to which they were completely reported for each synthesis in 100 SRs.
- Fewer than half defined groups in enough detail to replicate decisions about which studies were eligible for each synthesis; 39% used intervention groups in the synthesis that were not mentioned in the methods.

What this adds to what was known?

- This study is the first to show that the PICO for each synthesis is incompletely reported in many SRs.

What is the implication and what should change now?

- Complete reporting of synthesis questions is needed to ensure clear, replicable and well-focused syntheses, and help ensure that synthesis is not driven by what is reported in included studies.

Box 1 Example: questions for each synthesis.

In an SR of psychosocial interventions for smoking cessation [4], the eligibility criteria for the review included *any* psychosocial intervention to help people stop smoking during pregnancy.

One research question was whether psychosocial interventions in general were effective for smoking abstinence in late pregnancy. To address this question, all types of psychosocial interventions were included in a single comparison.

Other, more specific, research questions examined whether each of the psychosocial interventions that are commonly used to support people to stop smoking during pregnancy (e.g., counseling, health education, feedback) were effective. To address these questions, separate comparisons were presented to estimate the effects of each type of psychosocial intervention compared with usual care. A question could be specified for each synthesis using the PICO framework, for example, ''What is the effect of smoking cessation counseling (intervention) during pregnancy (population) compared to usual care (comparator) on 'smoking abstinence in late pregnancy' (outcome)?''

In addition to specifying intervention groups for comparisons, groups were defined for secondary analyses that examined whether the size of the effect of psychosocial interventions was modified by certain factors. For these analyses, explicit criteria were defined for three subgroups: single component interventions, multicomponent interventions, and tailored interventions.

PICO = Population, Intervention, Comparator, Outcome. Adapted from: Chapter 3, Cochrane Handbook for Systematic Reviews of Interventions [3].

for each synthesis. Complete reporting involves specifying each of the synthesis questions to be addressed and the PICO criteria for including studies in each synthesis (i.e., defining each of the PICO groups) [3]. In specifying groups, authors should.

1. Identify the basis for groups (as part of the rationale, e.g., we grouped populations by age to understand the effects in both younger children and teenagers).
2. Name (label) *and* define each group in sufficient detail to allow replication of decisions about which studies are eligible for each group (e.g., children were defined as up to age 10, while teenagers were aged 11 or older).
3. State the role of each group in the synthesis (i.e., for comparisons, or for subgroup or sensitivity analyses).

Fig. 1 shows variations of complete (blue pathways) and incomplete (orange pathways) specification of groups for comparison. Incomplete reporting occurs when one or more essential elements are not reported in the description of methods (e.g., groups are named but not defined in sufficient detail for replication, or it is unclear whether groups are intended for comparisons or another purpose in synthesis). In the most extreme cases, authors may use groups in the synthesis but not identify any groups in the methods description, or may identify groups but then use different groups in synthesis (orange boxes).

Clearly specifying the synthesis questions and PICO for each synthesis benefits both authors and readers of SRs, as well as editors and peer referees. These should preferably be planned at the protocol stage so as to minimize post hoc decisions based on the results of the included studies. Such planning can be of particular benefit where the authors expect to encounter diversity in populations, settings, interventions and outcomes [5,6]. Reporting the PICO for each synthesis enables readers to understand the research

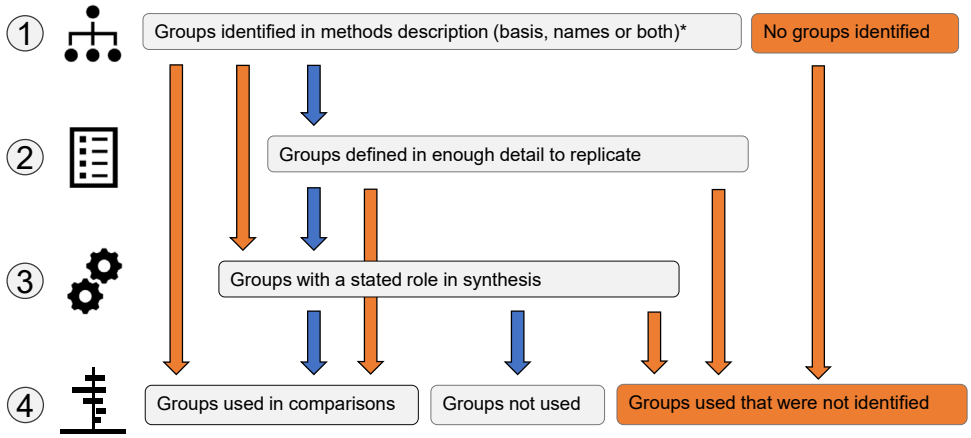

Fig. 1. Variations on complete and incomplete specification of the PICO for each synthesis. Blue: all essential elements of fully specified groups are reported (noting that specified groups may not be used in practice for legitimate reasons, such as when no studies within the particular group are identified). Orange: One or more essential elements of specification are missing. *Complete reporting requires both the basis for groups and naming each group. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

questions addressed. Enhanced transparency may increase confidence in the methods used, and therefore in the findings of the SR [3].

Given the critical role in determining the findings of the SR, better understanding is needed of how completely the synthesis questions and corresponding PICO for each synthesis are reported in SRs. Therefore, we aimed to examine the characteristics of population, intervention and outcome groups and the extent to which they were completely reported for each synthesis in a sample of SRs of interventions., and which elements of information may not be completely reported.

2. Study design and setting

Deviations from the published protocol [7] are reported in Supplementary File 1, Section A. Results from a companion study examining the use of synthesis methods other than meta-analysis in the same sample of SRs are published elsewhere [8].

2.1. Eligibility criteria

We aimed to include SRs that were likely to address multiple synthesis questions and include primary studies with diverse characteristics. We therefore sought SRs of public health and health systems interventions, which are likely to include diverse populations and interventions and a range of study designs [9], and consequently address multiple synthesis questions.

We included SRs that

1. aimed to synthesize the results of primary studies, stated eligibility criteria and reported a search strategy;
2. examined the quantitative effects of any public health or health systems intervention, including policies, programs or strategies, treatments or elements of care;
3. included at least one comparison with at least two studies; and
4. were published in English.

We excluded SRs that

5. synthesized the results of other SRs (e.g., overviews of SRs);
6. used network meta-analysis; or
7. addressed questions other than intervention effectiveness (see protocol [7] and Supplementary File 1, Section A).

2.2. Search strategy

Records of all SRs published during 2018 were obtained from two databases: *Health Evidence,* [10] indexing SRs in public health (searched 29 November 2019), and *Health Systems Evidence,* [11] indexing SRs of health systems interventions (searched 10 October 2019).

2.3. Sample size and study selection

For reasons of feasibility and precision, our sample was restricted to 100 SRs (see protocol for sample size justification [7]). From the search results, we randomly sampled and screened records until our target sample size was met. Random sampling, selection and data extraction were conducted in EPPI-Reviewer [12]. Two authors (MC, SB) independently screened titles and abstracts and excluded clearly ineligible records. SRs were then assessed for eligibility using the full text by one author (MC). A second author (one of SB,

Table 1. Examples of data extraction and coding items

Category	Examples of data extraction and coding items[a]
SR characteristics	• Presence of diversity in eligibility criteria (populations, interventions, settings, study designs). • No. of included studies. • Availability of a protocol or register entry. • Cited sources of general guidance on SR methods. • Methodological quality indicators.
Specification of groups for comparisons	• Identification (i.e., any basis for groups or naming of groups, or both) in the methods description (irrespective of which section of the SR this was reported in). • Definition in sufficient detail to allow replication of decisions about which studies are eligible for each group. • Stated role of groups in synthesis (i.e., for comparisons or for a secondary role [including subgroup or sensitivity analyses]). • Rationale for groups (including use of taxonomies).
Basis of groups for comparisons	• Population: e.g., intended recipient of intervention; disease/condition; participant characteristics (PROGRESS-Plus framework [13]); setting. • Intervention/comparator: e.g., intervention characteristics (Template for Intervention Description and Replication framework [14]); handling of inactive controls; multicomponent interventions; cointerventions. • Outcome: e.g., outcomes within a domain, measurement tools/methods, time points. • Methodological characteristics: e.g., study design; design features; risk of bias.
Use of groups in practice	• Use in synthesis of groups for comparisons: o Groups identified in the SR methods description. o Groups used that were not identified in the methods description. • Statements about changes to planned methods.

[a] See complete Data Dictionary in Supplementary File 1, Section B.

JM, RR) independently assessed eligibility based on the full text for 21 SRs. Disagreements were resolved by consensus.

2.4. Data extraction and management

Four authors (MC, SB, JM, RR) piloted the data extraction form and coding on five SRs to refine the items and coding guidance, and achieve a shared understanding of the form. Two authors (MC and one of SB, JM, RR) independently coded and extracted data from a further 12 SRs, and met to resolve disagreements. Data were coded and extracted from the remaining SRs by one author (MC), for reasons of feasibility. Instead authors met weekly (usually) to discuss any uncertainty arising during single data extraction and coding. Amendments to the data extraction form and coding guidance were made following these discussions. Changes to coding were applied to all SRs. See Table 1 for an overview of extracted data and complete Data Dictionary in Supplementary File 1, Section B.

We aimed to identify the groups intended for and/or used in comparisons, irrespective of whether the groups were fully specified. Definitions and our approach to coding are presented in Table 2. We coded groups based on each PICO component separately (populations, interventions, outcomes and methodological characteristics). For any particular PICO component, there could be multiple ways of grouping (e.g., grouping population by disease severity in one synthesis and age in another). Given the groups could be specified with different levels of completeness, we coded all levels that applied. e.g., suppose an SR had three ways of grouping population: if two had been *named but not defined*, and one *defined in enough detail to replicate*, we coded both levels of completeness. For outcomes, we took a different approach to coding as most SRs include multiple outcomes, each of which is intended for separate synthesis and could therefore be considered a 'group'. For feasibility, we limited coding to the first primary outcome in each SR. Within this outcome, we coded groups that were subsets based on the time point of measurement (e.g., short vs. long-term outcomes), and groups that included more than one outcome, measure or tool within a broader outcome or domain for synthesis (whether this brought together all outcomes in a single synthesis, or created subsets). We did not code groups if *all* populations, interventions, methodological characteristics or time points of outcome measurement eligible for the review were grouped together for synthesis, nor did we code groups where the first primary outcome was used for synthesis without including more than one more specific outcome or measure (most likely when the outcome used for synthesis was very specific, such as weight in kilograms).

We calculated descriptive summary statistics using Stata [15]. The dataset [16] and analysis code [17] are available in a public repository.

Table 2. Definitions used for coding and coding approach

Term	Definition	Coding approach[a]
Group	A subset of the populations (e.g., children in an SR examining the effects of an intervention for all age groups), interventions (e.g., a specific type of exercise in an SR examining the effects of exercise), or methodological characteristics (e.g., studies judged to be at low risk of bias).	For each PICO[b] component, groups were coded as: • Identified: If any reference was made to groups (i.e., basis for groups, naming of groups or both) in the methods description (irrespective of which section this was reported in). For example, a statement that studies were ''grouped by age'' would be coded as identified because the basis of groups is mentioned. • Defined: If the groups were defined in enough detail to replicate decisions about which included studies are eligible for each group. Where enough detail to replicate decisions was not provided, the extent of specification was coded (e.g., only basis for grouping stated, groups named but not defined). • Used in practice: If the groups appeared in the synthesis (i.e., defining a comparison for meta-analysis or other synthesis methods). Where groups were identified in the SR (e.g., in the Background or eligibility criteria), but there was no evidence that the authors intended to use the groups in synthesis (see 'Role of groups in synthesis' below), groups were not coded.
Outcome group	Outcome groups were defined (for our purpose) as either: • a time period for outcome measurement (e.g., immediate postintervention, longest follow-up) that defined a subset of possible measurements; or • an outcome domain within which more than one outcome (e.g., physical activity and smoking within a 'health behavior' domain), or measurement method (e.g., self-reported and biochemically validated smoking) was included, whether this brought together all included studies or defined a subset.	For reasons of feasibility, coding was limited to the first reported primary outcome of each SR (see Supplementary File 1, Section A, for decision rules to select this outcome). Outcome groups based on domains were coded if there was evidence that multiple measures or more specific outcomes were grouped together in a comparison. Outcome groups were then coded as for other PICO[b] components above.
(Pairwise) comparison	The comparison of an outcome between two interventions (e.g., exercise vs. counseling, or exercise vs. no intervention) for a particular population. The comparison may be further restricted based on methodological characteristics.	We coded comparisons regardless of the method of synthesis, including where studies were grouped for summary without the use of a statistical synthesis method.
Role of groups in synthesis	The purpose for which groups were intended or used in the synthesis or summary of results. Synthesis was defined as a process of combining results from a set of included studies with the aim of drawing conclusions about a body of evidence [5]. This could include meta-analysis, other synthesis methods (e.g., vote counting), or summaries of individual study results when statistical synthesis was not used.	For each PICO[b] component, the role of groups was coded as: 1. For comparisons: • If there was evidence (either from an explicit statement or use of the groups was identifiable in practice) that the authors intended to use the groups for comparing the effect of interventions. • These comparisons may be presented separately (e.g., meta-analysis of 'exercise vs. control' and 'counseling vs. control', each in a separate forest plot) or together (e.g., the groups 'exercise' and 'counseling' would be coded as 'for comparisons' in a meta-analysis stratified by the comparisons 'exercise vs.

(*Continued*)

Table 2. Continued

Term	Definition	Coding approach[a]
		control' and 'counseling vs. control' if the effect of both exercise and counseling was of interest and the purpose of these groups was not solely to investigate statistical heterogeneity). 2. For a secondary purpose: • If there was evidence that the authors intended to use the groups to investigate statistical heterogeneity (subgroup analysis, meta-regression) or examine the robustness of results to different assumptions (i.e., sensitivity analysis). Note that where there was not an explicit statement, we made a judgment about the role based on whether the effect estimate for each group was of interest (coding as 'for comparisons') or whether a secondary purpose was of interest (coding as 'for secondary purpose').

[a] Detailed coding definitions and examples are available in the data dictionary in Supplementary File 1, Section B.
[b] For this study, we use PICO to include population(s), intervention(s) (and comparators), outcome(s) and methodological characteristics.

3. Results

3.1. Search results

We identified 865 unique records and randomly selected 166 records for abstract screening, oversampling to allow for exclusions (15 were excluded). We then assessed the full text of 108 SRs (in random order) until we reached our sample size of 100. See Supplementary File 1, Section C, for included and excluded SRs, and Supplementary File 2, Figure S1: PRISMA flow diagram.

3.2. Characteristics of included SRs

Characteristics of the included SRs are presented in Table 3, with additional characteristics in Supplementary File 2, Table S1.

3.3. Specifying comparisons for synthesis: population groups

Eleven SRs (11/100) identified and/or used population groups for comparisons (Fig. 2, Table 4). Of these, all 11 identified the groups in the methods description. Five SRs (5/11, 45%) defined groups in enough detail to replicate decisions about which included studies were eligible for each group, four (4/11, 36%) provided a rationale, and eight (8/11, 73%) stated that the groups would be used for comparisons. Ten SRs (10/11, 91%) used population groups in practice for comparisons, while 6 SRs (6/11, 55%) did not use identified groups in practice. Three SRs specified both a broader synthesis combining all populations, as well as separate comparisons for specific population groups.

Of the 11 SRs that identified population groups for comparisons, the most common basis for grouping was by disease or condition (7/11, 64%, Table 5). The most common rationale for considering populations separately was anticipated clinical differences (3/11, 27%).

3.4. Specifying comparisons for synthesis: intervention groups

Forty-one SRs (41/100) identified and/or used intervention groups for comparisons (Fig. 2, Table 4). The remainder (59/100) grouped all active interventions together and compared them against one inactive comparator group (e.g., a review examining the effects of group walking interventions to promote physical activity compared any group walking intervention against any inactive comparator, and did not examine the effects of more specific interventions). Of the 41 SRs with intervention groups, 29 SRs identified the groups in the methods description (29/41, 71%). Twelve SRs (12/41, 29%) defined groups in enough detail to replicate decisions about which included studies were eligible for each group, six (6/41, 15%) provided a rationale, and 24 (24/41, 59%) stated that the groups would be used for comparisons. Thirty-seven SRs (37/41, 90%) used intervention groups in practice for comparisons, including 16 (16/41, 39%) that introduced groups for the first time in the synthesis that had not been identified in the methods description. In five SRs (5/41, 12%), groups with a stated role in comparisons were instead used in a secondary role (such as subgroup or sensitivity analysis), and in seven (7/41, 17%) identified groups with a stated role in comparisons were not used in practice. Seven SRs (7/41, 17%) specified more than one level of grouping interventions for synthesis, such as a broad analysis including all interventions, as well as separate syntheses assessing specific interventions.

Table 3. Characteristics of included SRs

Characteristics	*n*[a] or median (IQR)[b]
Review eligibility criteria	*n* = 100
Participants with different health conditions or characteristics (e.g., patients and carers) included	64
Intervention types[c]	
Consumer communication and participation	37
Prevention, treatment or management of a health condition	35
Health systems and delivery of care	17
Population health	15
Intervention setting[c]	
Community	40
Any	16
Hospital	15
Multiple settings	14
Primary care	11
Other	10
Unclear or no information	6
Diversity in included interventions[c]	
Important variations on a single intervention included	98
More than one distinct intervention type included	48
Multicomponent or cointerventions included	85
Both randomized trials and nonrandomized studies included	50
Methods	
No. of included studies	15 (10 to 26)
Synthesis methods used[c,d] [8]	
Meta-analysis of effect estimates	58
Other statistical synthesis methods	60
No synthesis for at least one outcome	19

[a] The denominator is 100. Where *n* is reported, this is equivalent to a percentage. Illustrative exact binomial confidence intervals for different percentages when the sample size is 100: 1% (0–5%); 5% (2–11%); 10% (5–18%); 20% (13–29%); 30% (21–40%); 50% (40–60%). Confidence interval widths for percentages greater than 50% are the same as for 100% - percentage.
[b] IQR: interquartile range.
[c] For this item, SRs may be coded into more than one category, thus the sum of SRs across categories may be greater than 100.
[d] The methods used to synthesize results for each outcome were coded for every outcome reported in the SR.

Of the 29 SRs that identified intervention groups for comparisons, the most common basis for grouping was what was delivered (i.e., the content or components of the intervention, classified using the Template for Intervention Description and Replication reporting guideline [14] 20/29, 69%, Table 5). The most common rationale provided for the groups used was anticipated clinical difference between the interventions (3/29, 10%).

Most SRs grouped all inactive controls (e.g., no intervention, placebo, minimal attention controls, wait list controls or 'usual care') together for synthesis without providing a rationale for doing so (51/100, Table 5). Fifty-four SRs grouped multicomponent interventions (in which all components would be eligible for inclusion in the SR, e.g., a package including education, exercise and cooking classes in an SR of nonpharmaceutical interventions for weight loss) with single-component interventions (54/100). Thirty-four (34/100) grouped studies with and without cointerventions (components that would not be eligible for inclusion alone, e.g., a weight-loss drug provided alongside education in the same SR of nonpharmaceutical interventions) together for synthesis.

3.5. Specifying comparisons for synthesis: outcome groups

Sixty-seven SRs (67/100) identified and/or used outcome groups for comparisons (Fig. 2, Table 4). The remainder (33/100) presented a synthesis for each named outcome in the review, and did not combine outcomes, measures or tools together for synthesis, or create outcome groups based on time periods. Of the 67 SRs with outcome groups, 56 SRs identified the groups in the methods description (56/67, 84%). Twenty-eight SRs (28/67, 42%) defined groups in enough detail to replicate decisions about which included studies were eligible for each group, ten (10/67, 15%) provided a rationale, and 36 (36/67, 54%) stated that the groups would be used for comparisons. Sixty-six SRs (66/67, 99%) used outcome groups in practice for comparisons in the synthesis, including 19 (19/67,

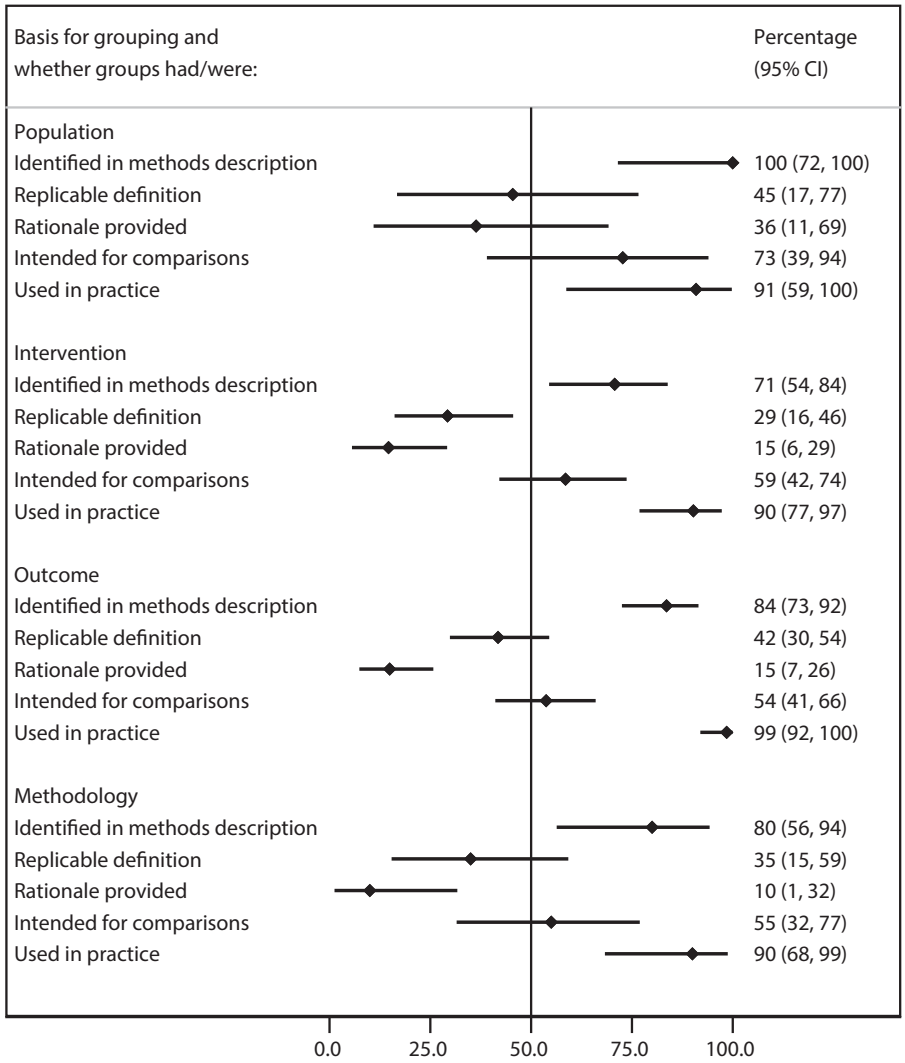

Fig. 2. Specification of groups for comparisons based on each PICO component.

28%) that used groups that were not identified in the methods description. In one SR (1/67, 1%), identified groups were not used in practice.

Two SRs (2/67, 3%) specified more than one level of grouping outcomes for synthesis, such as a broad synthesis of physical activity, as well as separate syntheses of moderate or vigorous activity. Additional characteristics of outcome groups are presented in Table 5.

3.6. Defining comparisons for synthesis: methodological groups

Twenty SRs (20/100) identified and/or used methodological groups for comparisons (Fig. 2, see Table 4 for extent of specification). Of the 16 SRs that identified methodological groups for comparisons, the most common basis for grouping was study design (e.g., randomized and non-randomized studies (14/16, 88%, see Table 5 for additional characteristics).

3.7. Changes to planned methods

One SR specified a contingency plan to switch from a broader to a narrower analysis depending on the data found (planning to synthesize studies of vaccine efficacy across all vaccines, with a contingency plan to separate syntheses by vaccine types should high statistical heterogeneity be found).

Although many SRs did not use the specified groups in practice (Table 4), only two explicitly stated that the included studies could not be grouped as planned (because the included studies all fell into a single group, or due to the degree of diversity observed in the included study designs).

4. Discussion

4.1. Key findings

We examined the extent to which PICO groups for each synthesis were completely specified in a random sample of

Table 4. Specification and use of groups for comparisons

Methods	PICO[a] components			
	P *n* (%)	I/C *n* (%)	O[b] *n* (%)	M *n* (%)
Did groups have any role in comparisons (specified and/or used)?	*n* = 100	*n* = 100	*n* = 100	*n* = 100
Yes[c]	11 (11)	41 (41)	67 (67)	20 (20)
No	89 (89)	59 (59)	33 (33)	80 (80)
Specification of groups for comparisons[d]	*n* = 11	*n* = 41	*n* = 67	*n* = 20
Were the groups identified in the methods description?[e]				
Yes	11 (100)	29 (71)	56 (84)	16 (80)
No	0 (0)	16 (39)	19 (28)	5 (25)
Were the groups defined?[e]				
Yes, in enough detail to replicate decisions	5 (45)	12 (29)	28 (42)	7 (35)
Yes, but not enough detail to replicate	3 (27)	2 (5)	7 (10)	2 (10)
No, groups were named but not defined	1 (9)	6 (15)	16 (24)	6 (30)
No, only basis for grouping stated	2 (18)	9 (22)	5 (7)	1 (5)
Not identified in the methods description	0 (0)	16 (39)	19 (28)	5 (25)
Was a rationale provided for the groups?[e]				
Yes	4 (36)	6 (15)	10 (15)	2 (10)
No	7 (64)	24 (59)	46 (69)	14 (70)
Not identified in the methods description	0 (0)	16 (39)	19 (28)	5 (25)
Was the role in the synthesis stated?[e] (i.e., intention to use the groups for comparisons)				
Yes, to define comparisons	8 (73)	24 (59)	36 (54)	11 (55)
No role specified	4 (36)	5 (12)	20 (30)	5 (25)
Not identified in the methods description	0 (0)	16 (39)	19 (28)	5 (25)
More than one level of grouping was specified or used (broader and narrower)?[f]	3 (27)	7 (17)	2 (3)	0 (0)
Use in practice of groups for comparisons	*n* = 11	*n* = 41	*n* = 67	*n* = 20
Were the groups used in practice?[e]				
Yes	10 (91)	37 (90)	66 (99)	18 (90)
No – identified for comparisons, but used in a secondary role	0 (0)	5 (12)	0 (0)	1 (5)
No – identified but not used	6 (55)	7 (17)	1 (1)	2 (10)

[a] P = population groups. I = intervention groups. O = outcome groups. M = methodological groups. For P, I, M: groups were counted if they divided the set of included studies into smaller groups.

[b] For O: only the first mentioned primary outcome was considered (see Supplementary File 1, Section A for definition). An outcome group was coded if groups based on time periods divided the set of included studies into smaller groups, or if there was evidence that multiple specific measures or outcomes were grouped together for comparison.

[c] Groups were specified and/or used in practice for comparisons. Groups for secondary roles such as subgroup or sensitivity analysis were not counted.

[d] For specification and use (beneath gray header rows), the denominator is the number of SRs with groups for comparisons for each PICO component.

[e] For this item, for each PICO component, there could be multiple ways of grouping within the SR (e.g., grouping population by disease severity and gender in the same SR), and groups could be specified with different levels of completeness, thus multiple response options might apply. Therefore, the sum of percentages across the response options may be greater than 100. Each selected response option applied to at least one of the groups for the particular PICO component. e.g., the response options 'Yes' and 'No' for 'Were the groups identified in in the methods description?' should be interpreted as 'Yes, for at least one group' and 'No, for at least one group'.

[f] Both a broader and narrower level of grouping was specified or used in practice on the same basis. e.g., one comparison assessed the effect of the intervention in all populations, and separate comparisons also assessed the effect in specific age groups (e.g., in babies, children, adults) within a single SR.

100 SRs of interventions, focusing on groups that were intended or used for comparisons (i.e., not for subgroup or sensitivity analyses). We identified examples of all the variations of reporting in Fig. 1, including where groups were identified but not used (e.g., as expected if studies were unavailable for inclusion in the synthesis) and where groups were used but not identified in the methods description (an unacceptable practice). We found that while groups were commonly identified for populations, interventions and outcomes in the description of methods, less than half of the identified groups were defined in enough detail to replicate decisions about which studies in the review were eligible for each group. Groups were commonly identified without a rationale or any explicit statement that they would be used for comparisons. The most extreme cases of incomplete reporting were for intervention groups, with

Table 5. Characteristics of identified groups for comparisons

Characteristics	*n* (%)
Population groups[a]	*n* = 11
Basis for groups[b,c]	
Disease/condition	7 (64)
Specific disease or condition	6 (55)
Type/severity/level within a single disease or condition	2 (18)
Equity characteristics (PROGRESS-Plus framework [13])	6 (55)
Place of residence	1 (9)
Gender/sex	1 (9)
Education	1 (9)
Personal characteristics associated with discrimination (e.g., age)	4 (36)
Other	1 (9)
Rationale for groups[c]	
Clinical difference	3 (27)
Equity	1 (9)
No rationale reported	7 (64)
Intervention groups[a]	*n* = 29
Basis for groups (classified using Template for Intervention Description and Replication)[b,c] [14]	
What? (e.g., materials or procedures)	20 (69)
How? (e.g., mode of delivery)	4 (14)
When and how much? (e.g., number, duration, frequency, intensity)	3 (10)
Why? (e.g., rationale or goal)	3 (10)
Who? (e.g., provider)	2 (7)
Where? (e.g., type of location, features)	2 (7)
Unclear	3 (10)
Rationale for groups[b,c]	
Clinical difference	3 (10)
Other	2 (7)
Unclear	1 (3)
No rationale reported	24 (83)
Used taxonomy or framework?	
Yes — developed for this SR	1 (3)
Yes — existing	1 (3)
No information	27 (93)
Handling of inactive controls, multicomponent interventions and cointerventions[d]	*n* = 100
Handling diverse inactive controls[b]	
Grouped with other inactive 'controls' with no rationale	51 (51)
Grouped with other inactive 'controls' with rationale	11 (11)
Separated	7 (7)
Other (e.g., combining active comparators with inactive controls)	15 (15)
Unclear or no information	18 (18)
Not applicable[e]	1 (1)
Handling studies with multicomponent interventions[b]	
Grouped together with single-component interventions	54 (54)
Separate group for multicomponent interventions	11 (11)
Individual components analyzed using meta-regression	1 (1)
Allocated to groups based on 'main' component	0 (0)
Excluded from synthesis or the SR[f]	0 (0)
Other	10 (10)
Unclear or no information	16 (16)
Not applicable[e]	13 (13)
Handling studies with cointerventions[b]	
Grouped together with others	34 (34)
Excluded from synthesis or the SR[f]	1 (1)
Separate group for studies with cointerventions	0 (0)
Other	4 (4)
Unclear or no information	51 (51)
Not applicable[e]	13 (13)
Outcome groups[a]	*n* = 56
Type of group[b]	
Narrow outcome groups (e.g., specific measures of exercise)	31 (55)
Broad outcome groups (e.g., different health behaviors)	24 (43)
Time periods (e.g., short and long term follow-up)	5 (9)
Rationale for groups[c]	
Existing core outcome set/taxonomy	5 (9)
Hypothesised effects	2 (4)
Clinical/conceptual similarity	1 (2)
Other	2 (4)
No rationale given	46 (82)
Used taxonomy or framework?	
Yes — developed for this SR	2 (4)
Yes — existing	6 (11)
No information	48 (86)
Methodological groups[a]	*n* = 16
Basis for groups[c]	
Study designs	14 (88)
Unclear	2 (13)
Rationale for groups[c]	
Risk of bias	1 (6)
Other	1 (6)
No rationale provided	14 (88)
Used taxonomy or framework?	

(*Continued*)

Table 5. Continued

Characteristics	*n* (%)
Yes – existing	2 (13)
Yes – developed for this SR	0 (0)
No information	14 (88)

[a] In this section, the denominator is the number of SRs for which groups were identified in the methods description, and either specified and/or used in practice for comparisons. Characteristics were not collected for groups used in practice for comparisons where they had not been identified in the methods description.

[b] For this item, there could be multiple ways of grouping within the SR (e.g., grouping population by disease severity and gender in the same SR), thus multiple response options might apply. Therefore, the sum of percentages across the response options may be greater than 100. Each selected response option applied to at least one of the groups for the particular PICO component. For example, the response option 'Disease/condition' for 'Basis for groups' should be interpreted as 'Yes, for at least one group'.

[c] For this item, additional coding options with zero responses are not displayed (see complete data dictionary in Supplementary File 1, Section B).

[d] In this section, the denominator is the total number of SRs (n = 100), coded on the basis of both the text of the SR and the observed synthesis. SRs may be coded into more than one category, thus the sum of SRs across categories may be greater than the denominator.

[e] Interventions of this type were not included in the SR.

[f] Interventions of this type were eligible for the SR, but were excluded from synthesis, or some interventions of this type were excluded from the SR while others were included.

39% of SRs using groups in practice that were not mentioned in the methods description (Fig. 1, orange pathways stemming from 'no groups identified'). Finally, we found that it was rare for a contingency plan to be reported for an alternative synthesis structure if the preferred comparisons could not be implemented (e.g., grouping more broadly if there were too few studies).

4.2. What this study adds to what is already known

We are unaware of any studies that have investigated in detail the completeness of reporting and characteristics of groups used to define comparisons in SRs. Others studies have examined limited aspects of the use of groups for comparisons (such as in Campbell et al. [18], which found intervention characteristics were the most common basis for grouping studies) and subgroup analyses (Paquette et al. [19] examined the frequency and reporting of participant and intervention characteristics used in subgroup analyses in SRs of atrial fibrillation).

Our findings indicate that incomplete reporting of synthesis questions may be common, which is perhaps unsurprising. Before 2019 (when the concept of 'PICO for each synthesis' was introduced [3]), methods guidance and appraisal tools for SRs had only focused on review questions and criteria (i.e., 'PICO for the review'). This is despite the potentially critical impact that decisions about the PICO questions and criteria for each synthesis can have on the findings of a review. A concerted effort is likely to be needed to improve reporting practice, which could include development of a practical tool to help systematic reviewers understand the distinction between the PICO for the review and for each synthesis, and to help in the planning and reporting of synthesis questions (work which our team has undertaken, see InSynQ checklist and guide [20,21]). In parallel efforts, the need for more clarity in specifying objectives in randomized trials has been recognised (i.e., specifying the precise definition of the treatment effect to be estimated), leading to the development of the estimand framework [22,23] (which requires specification of the population, treatment groups and outcome, along with other attributes).

4.3. Strengths and limitations

Our sample (and thus findings) is likely to be representative of SRs of public health and health systems interventions, because the SRs were randomly selected and there were minimal restrictions on eligible characteristics. Although public health and health systems research can address diverse questions, our eligible SRs were all reviews of interventions, for which the 'PICO' framework (or one of its variants) is appropriate for both the review question and synthesis questions.

Duplicate full text screening, data extraction and coding were not conducted for all SRs, leaving some risk of missed, misclassified or miscoded data. Coding required judgment, especially where there was little or no description of the planned comparisons, or where meta-analysis was not used, and inferences were often required about the groups identified and their intended role. In practice, this indicates the extremely limited information about the synthesis questions and structure in many included reviews. However, we opted to code in an overly inclusive way, accepting often very limited and fragmented information as evidence that a group had been identified or defined. Although other teams might code differently, taking this conservative approach meant that we were likely to underrather than overestimate the extent of incomplete reporting. Any risk of misclassification was further mitigated by a detailed piloting process and regular discussion of specific cases amongst members of the study team, who had extensive SR experience.

4.4. Implications of this research

In SRs, clear and complete reporting of the PICO for each synthesis underpins the conduct of the synthesis and interpretation of findings, with benefits for those producing and using SRs. Authors, and others involved in producing reviews, are encouraged to specify the PICO for each

synthesis when reporting their reviews, so that the synthesis questions and structure are clear, and decisions about which studies were eligible for each synthesis are transparent and replicable. Doing so may enhance readers' understanding of the findings, allow critique of the methods used and decisions taken, and assist in applying the results to decision making. As with any method, changes to planned groups should be clearly reported and a rationale provided.

Ideally, the PICO for each synthesis should be planned and reported in protocols. Such planning requires a shift that may pose challenges in some reviews, but can streamline decision-making at review stage and lead to more focused and useful reviews. At review stage, authors often make decisions about how to handle unexpected diversity in study characteristics or reported data by modifying their planned methods [24]. Some of this could be prevented by planning for the possibility that the preferred groups cannot be applied in practice, for example by outlining an alternative synthesis plan should sufficient data not be available.

5. Conclusion

SRs of public health and health systems interventions inform critical policy and public health decisions. In a sample of such SRs, we found that multiple research questions within an SR are often addressed by grouping studies for synthesis based on population, intervention, outcome and methodological characteristics, but these groups are often incompletely specified. Improved reporting of how studies are grouped for synthesis would improve transparency and replicability, and help ensure that the synthesis is not driven by what is reported in the studies. This should lead to more focused synthesis questions and, ultimately, more useful SRs for health decision makers.

CRediT authorship contribution statement

Miranda S. Cumpston: Methodology, Formal analysis, Investigation, Data curation, Writing − original draft, Writing − review & editing, Project administration. **Joanne E. McKenzie:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing − review & editing, Supervision. **Rebecca Ryan:** Investigation, Validation, Writing − review & editing. **James Thomas:** Investigation, Writing − review & editing. **Sue E. Brennan:** Conceptualization, Methodology, Validation, Investigation, Writing − review & editing, Supervision.

Data availability

Data are available in a public repository and cited in the paper (https://doi.org/10.26180/23597031). The data dictionary is provided in Supplementary File 1, Section B. The extracted dataset [16] and Stata code [17] used in the analysis are available online in a public repository, Bridges.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

We acknowledge Elizabeth Korevaar at the School of Public Health and Preventive Medicine, Monash University, for assistance with STATA code, and Kristin Read, Research Coordinator at Health Evidence™, McMaster University, for providing access to search results from that database. MSC receives funding from the Australian Government Research Training Program. JEM is supported by a National Health and Medical Research Council Investigator Grant (GNT2009612). SEB's position at Cochrane Australia is funded by the Australian Government through the National Health and Medical Research Council. RR's position at Cochrane Consumers and Communication is supported by funding from the Australian Government through the National Health and Medical Research Council. Funding organisations had no role in the conduct or reporting of this study. JT is supported in part by the National Institute for Health and Care Research. This report is independent research supported by the National Institute for Health and Care Research ARC North Thames. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.09.013.

References

[1] McKenzie JE, Beller EM, Forbes AB. Introduction to systematic reviews and meta-analysis. Respirology 2016;21(4):626−37.

[2] Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Chapter 2: determining the scope of the review and the questions it will address. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane handbook for systematic reviews of interventions. Cochrane; 2023. Version 6.4 (updated August 2023). Available at www.training.cochrane.org/handbook. Accessed October 24, 2023.

[3] McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane handbook for systematic reviews of interventions. Cochrane; 2023. Version 6.4 (updated August 2023). Available at www.training.cochrane.org/handbook. Accessed October 24, 2023.

[4] Chamberlain C, O'Mara-Eves A, Porter J, Coleman T, Perlen SM, Thomas J, et al. Psychosocial interventions for supporting women to stop smoking in pregnancy. Cochrane Database Syst Rev 2017; 2:CD001055.

[5] McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV. Chapter 9: summarizing study characteristics and preparing for synthesis. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane handbook for systematic reviews of interventions. Cochrane; 2023. Version 6.4 (updated August 2023). Available at www.training.cochrane.org/handbook. Accessed October 24, 2023.

[6] Thomas J, Petticrew M, Noyes J, Chandler J, Rehfuess E, Tugwell P, et al. Chapter 17: intervention complexity. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane handbook for systematic reviews of interventions. Cochrane; 2023. Version 6.4 (updated August 2023). Available at www.training.cochrane.org/handbook. Accessed October 24, 2023.

[7] Cumpston M, McKenzie J, Thomas J, Brennan S. The use of 'PICO for synthesis' and methods for synthesis without meta-analysis: protocol for a survey of current practice in systematic reviews of health interventions [version 2; peer review: 2 approved]. F1000Res 2021;9:678.

[8] Cumpston MS, Brennan SE, Ryan R, McKenzie JE. Synthesis methods other than meta-analysis were commonly used but seldom specified: survey of systematic reviews. J Clin Epidemiol 2023;156:42−52.

[9] Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al. Complex interventions and their implications for systematic reviews: a pragmatic approach. J Clin Epidemiol 2013;66:1209−14.

[10] Health Evidence. Health evidence. McMaster University. Available at www.healthevidence.org. Accessed November 19, 2019.

[11] Health Systems Evidence. About HSE. McMaster University. Available at www.healthsystemsevidence.org/about?lang=en. Accessed November 19, 2019.

[12] EPPI-Reviewer web: software for research synthesis. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education; 2020.

[13] O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol 2014;67:56−64.

[14] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687.

[15] Stata statistical software: release 17. College Station, TX: StataCorp LLC; 2021.

[16] Cumpston M, McKenzie J, Ryan R, Thomas J, Brennan S. PICO for synthesis data file: cumpston_et_al_2023_PICO.xlsx. Melbourne: Monash University; 2023.

[17] Cumpston M, McKenzie J, Ryan R, Thomas J, Brennan S. PICO for synthesis Stata code: cumpston_et_al_2023_PICO.do. Melbourne: Monash University; 2023.

[18] Campbell M, Katikireddi SV, Sowden A, Thomson H. Lack of transparency in reporting narrative synthesis of quantitative data: a methodological assessment of systematic reviews. J Clin Epidemiol 2019; 105:1−9.

[19] Paquette M, Alotaibi AM, Nieuwlaat R, Santesso N, Mbuagbaw L. A meta-epidemiological study of subgroup analyses in Cochrane systematic reviews of atrial fibrillation. Syst Rev 2019;8(1):241.

[20] Brennan SE, Cumpston MS, Ryan R, McKenzie JE. InSynQ (Intervention Synthesis Questions) checklist and guide for developing and reporting the questions addressed in systematic reviews of interventions 2023: Version 1.0, updated 29 April 2023. Available at https://InSynQ.info. Accessed October 24, 2023.

[21] Cumpston M, McKenzie J, Ryan R, Flemyng E, Thomas J, Brennan S. Development of the InSynQ checklist: a tool for planning and reporting the synthesis questions in systematic reviews of interventions. Authorea June 27, 2023. Available at https://doi.org/10.22541/au.168786964.46845500/v1. Accessed October 24, 2023.

[22] Clark TP, Kahan BC, Phillips A, White I, Carpenter JR. Estimands: bringing clarity and focus to research questions in clinical trials. BMJ Open 2022;12(1):e052953.

[23] International Council for harmonisation of technical requirements for pharmaceuticals for human use (ICH). ICH harmonised guideline: Addendum on estimands and sensitivity analysis in clinical trials. ICH. Available at https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. Accessed June 16, 2023.

[24] Cumpston M, Lasserson T, Flemyng E, Page MJ. Chapter III: reporting the review. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane handbook for systematic reviews of interventions. version 6.4 (updated August 2023). Cochrane. Available at www.training.cochrane.org/handbook. Accessed October 24, 2023.