Defining and determining which quantitative study designs to include in your systematic review of effects of a healthcare intervention

Date of issue: April 2024

Cite as: Lindson N*, Hartmann-Boyce J*, Bero L, Dumville J, Flemyng E, Livingstone N, Reeves B, Tugwell P, Wilson DB, Sharma Waddington H. Defining and determining which quantitative study designs to include in your systematic review of effects of a healthcare intervention. Cochrane.

* Joint first authors.

BOX BELOW

Please note:

- Randomised controlled trials (RCTs) may not be available or may only partly answer an overall review question.
- Non-randomised studies of interventions (NRSI) may be particularly useful in answering specific synthesis questions within a systematic review, for example questions about long-term or adverse events (harms), or to address important equity considerations such as effects for specific populations.
- NRSI may also be important in syntheses that examine the effects of interventions, programmes and policies delivered under real-world settings, using routine methods of assignment of patients or other groups to a health care intervention.
- Including NRSI in systematic reviews is complex. Some types of NRSI design are thought more credible than others in estimating intervention effects in particular circumstances, hence it is important to ascertain how a study was carried out (study design). This guidance offers insights into some key issues to consider. If you do decide to include NRSI, your author team should include people with expert knowledge of NRSI methods as Cochrane editors will reject systematic reviews if there are problems with how they have been applied.

About this guidance

This guidance aims to help Cochrane authors consider two key questions:

1) should non-randomised studies of interventions (NRSI) be included in your systematic review of effects of a health care intervention; and, if so,
2) what types of NRSI design may help answer review questions about intervention effects?

It has been developed primarily for systematic reviews of the effects of healthcare interventions in public health. NRSI are defined in the *Cochrane Handbook* as any quantitative study estimating the effects of an intervention (benefit or harm) that does not use randomisation to allocate individuals (or groups of individuals) to intervention groups. It is increasingly recognised that NRSI have an important role to play in systematic reviews of health care interventions (e.g., Saldanha et al., 2022). Inclusion of NRSI in systematic reviews is an area where methods are evolving quickly, hence authors should also refer to relevant methods guidance referenced in this document, including Chapter 24 of the *Cochrane Handbook,* which authors are advised to consider and should adhere to in their systematic review. The Grading of Recommendation, Assessment, Development and Evaluation (GRADE) Working Group have also developed useful guidance on incorporating NRSI in systematic reviews (Schünemann et al., 2013; Cuello Garcia et al., 2022).

This guidance assumes that authors have defined their systematic review aim, articulated the review populations, interventions, comparators and outcomes (PICO) criteria, together with the PICO for each synthesis (e.g. for a particular population or outcome) addressed in the systematic review (see **Box 1**). Authors should also have conducted a scoping exercise of the randomised controlled trial (RCT) and NRSI evidence, and/or have sufficient knowledge of the subject area to answer question 1 below.

Box 1. PICO for each synthesis addressed in the systematic review

PICO is an acronym used to help shape clinical questions and inclusion criteria for systematic reviews of interventions. It stands for:

- **P**opulations: the individuals or groups of interest
- **I**nterventions: the interventions, programmes or policies of interest
- **C**omparators: the conditions to which the interventions will be compared
- **O**utcomes: the variables that assess the effects.

Systematic reviews usually include multiple PICOs. For example, important outcomes may include both intended and unintended consequences (including adverse events), be short term or measured over a long period, sometimes many years later; the population may include all eligible individuals, but specific sub-questions may relate to the most vulnerable groups such as those with co-morbid health issues; the intervention may have been investigated in explanatory (proof-of-concept) and routine care settings; the comparator may include those receiving standard care (practice as usual), a wait-list, an alternate intervention, or nothing at all.

The purpose of this document is to help Cochrane authors define eligible study designs for each PICO at the protocol stage of their review. See also Chapter 2 of the Cochrane Handbook.

Question 1: Should you include non-randomised studies of interventions?

NRSI use a variety of designs and methods to estimate and quantify the causal effect on the outcome(s) of interest of a health care intervention, programme or policy that has not been assigned to study participants randomly.[1] In an NRSI the health care intervention may be assigned using a forcing variable like a threshold on a pre-test measure, selection by planners or clinicians, and/or self-selection by patients (see the glossary in **Annex 1**).

Randomised controlled trials (RCTs), when designed and conducted rigorously, are widely considered to be the least biased form of primary evidence when assessing the effects of healthcare interventions (OCEBM Levels of Evidence Working Group, 2011; Higgins et al., 2019). Therefore, when conducting a Cochrane systematic review of the effects of a healthcare intervention, RCTs should always be included when available. However, NRSI may also be acceptable sources of evidence in the following circumstances:

1) For evidence about unintended consequences or adverse events (harms), which RCTs may not measure, or only measure imprecisely (that is, they are underpowered). NRSI may also be at less risk of bias in measuring unintended consequences, than intended consequences, in some circumstances because initial allocation to treatment conditions by practitioners or researchers is less likely to be associated with the probability of an outcome that is not anticipated (Golder et al., 2011). An example is a review of the effects of infant sleeping position on mortality (Gilbert et al., 2005);
2) If the PICO relates to the effects of interventions conducted in real-world settings rather than under highly controlled (explanatory or proof-of-concept) settings, especially in settings or fields where the conduct of an RCT would be challenging or uncommon; for example, when estimating the effects of juvenile curfews on offending (Wilson et al., 2016);
3) For evidence about long-term effects, where RCTs may be impractical, for example, due to the need to isolate control groups from the health care intervention over long periods. To take one example, a review of the effects of mass deworming interventions incorporated NRSI to measure long-term effects on education and work (Welch et al., 2016);
4) For outcomes measured among particular groups of participants who are not well represented in other studies, such as those included under PROGRESS-Plus (O'Neill et al., 2014).[2] For example, reviews of the effects of corticosteroids for pre-term deliveries (Crowley et al., 1990) and of the effects of treatments for dislocated or shallow hips in infants (Dwan et al., 2022) incorporated NRSI;
5) For rare primary outcomes, which are not necessarily unintended or adverse, for which evidence from RCTs may be underpowered. For example, a review of the effects of water, sanitation and hygiene interventions on mortality in childhood incorporated both RCTs and NRSI (Sharma Waddington et al., 2023);

[1] In the social sciences, the term "quasi-experimental design" (QED) is used to describe non-randomised studies of intervention effects (Shadish et al., 2002). NRSI that are conducted retrospectively are sometimes also called "natural experiments" (Craig et al., 2017).

[2] PROGRESS stands for place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status and social capital; Plus refers to personal characteristics associated with discrimination (e.g., disability), features of relationships (e.g., parents who smoked) and instances when a person is temporarily disadvantaged (e.g., respite care).

6) For emergent health conditions, if RCTs are likely to take longer to design and conduct than the decision-making cycle requires; for example, a review on the effects of personal protective equipment in reducing infection conducted in response to the COVID-19 pandemic (Verbeek et al., 2020).

Some publications reporting NRSI may also be known and used in policy and practice but be ineligible for a systematic review if it only plans to include RCTs. In these cases, reviewing this evidence alongside systematically identified eligible studies can help raise awareness among decision makers about any concerns regarding these studies (drawing on appropriate risk of bias assessment[3]).

The questions below (**Box 2**) can help you decide whether to include NRSI in your review of effects of a health care intervention. Your initial scoping of the literature (see page 1) should provide an indication of the evidence available for each review PICO/synthesis and facilitate the answering of these questions.

[3] LINK TO PROJECT 1, PROJECT 4..

Box 2. Should you consider including evidence from NRSI in your systematic review of effects?
These questions should be considered for each proposed PICO in your systematic review protocol [adapted from Section 24.1.1. of the *Cochrane Handbook* and using Schünemann et al. (2013) and Cuello-Garcia et al (2022)].

Consider including NRSI evidence if (tick all that apply):

RCTs are not available at all. For example, an RCT might be unfeasible, unethical, or unavailable for an emergent health condition.[a]	
RCT evidence is available but judged to be sparse (of limited quantity), uncertain (imprecise) or seriously indirect (meaning the evidence does not map directly onto your PICO).[a]	
PICO outcomes concern unintended or adverse events (harms).[b]	
PICO outcomes are likely only to occur in the long-term and/or are rare.[b]	
Inclusion may facilitate investigation of important equity or generalisability considerations that are considered as high priorities for the review. For example, if RCTs have only been conducted in high-income contexts or socioeconomically advantaged groups, or another relevant characteristic, where there is reason to think that there may be differences in what the evidence shows in these different groups.[c]	
Evidence from NRSI has been used for clinical or policy decision-making in your area of interest. In this case, end users of your review may benefit from you formally assessing this evidence.	

→If you have ticked one or more of these boxes, then you should consider including NRSI.

Other reasons to consider including NRSI recommended by GRADE, the Campbell Collaboration and others, but not necessarily recommended in the context of a Cochrane review, are:

- Effects differ between NRSI and RCT evidence in your area of interest and you want to explore this further.
- It will enhance the ability of your review to address questions about moderator effects (e.g., in meta-regression analysis).
- It can inform your baseline absolute risk figures used in summary of findings tables, to add context in your introduction or discussion sections, or to investigate correlation between surrogate outcomes and patient important outcomes (Gallo et al., 2017).

***Notes**: a) GRADE refers to this as 'replacement evidence'; b) GRADE refers to this as 'sequential evidence'; c) GRADE refers to this as 'complementary evidence'. NRSI may provide less certain evidence than well-conducted RCTs so including them does not necessarily result in higher quality or more consistent findings. For Cochrane reviews it is advised that NRSI at critical risk of bias are not included in the synthesis and therefore do not contribute to your review findings.*
***Abbr**: NRSI = non-randomised studies of interventions; RCT = randomised controlled trial; PICO = Population, Intervention, Comparator, Outcome.*

Decision 2: Which types of NRSI should you consider including and how should you describe them?

If you decide to include NRSI in your systematic review of effects of a health care intervention, the question then becomes which types of NRSI to include. These decisions should be driven by which NRSI study design feature(s) are most likely to address your research questions with the least bias (see Higgins et al., 2013, **Table 1** and **Annex 1**). The NRSI study design feature(s) most relevant to a review can differ depending on the PICO, with different design features yielding evidence with different risk of bias in different contexts. Decisions about inclusion will be guided by several

considerations, such as your initial scoping of the literature and/or prior knowledge of the NRSI evidence in the area. The design features that make NRSI eligible for inclusion in your review should be pre-specified and justified in your protocol.

It is not sufficient for authors to simply indicate they will include "NRSI" (or, for example, "quasi-experimental designs" or "natural experiments"). More information is needed to classify eligible study designs by their defining characteristics. Using NRSI study design labels alone, such as "controlled before-after" or "interrupted time series", can also be risky because they are inconsistently used (Polus et al., 2017; Vigneri et al., 2022). Therefore, where labels are used – for example, because they are commonly used short-hand names for more complex designs – they should be specified with respect to the key design features. **Table 1** presents four types of design feature that can be used to characterise NRSI. Authors should also read Section 24.2.1.3 of the *Cochrane Handbook*.

Annex 1 provides a glossary of study design labels and the commonly associated features. Cochrane's Effective Practice and Organisation of Care Review Group also has working definitions of these labels in the resource 'What study designs can be considered for inclusion in an EPOC review and what should they be called?'. Reeves et al. (2017) also sets out features associated with study design labels commonly used by systematic review authors.

Further guidance may be useful for implementing these approaches, including: 1) supporting the use of design features in the reporting of NRSI; and 2) supporting authors in recognising and selecting different NRSI based on their design features.

Table 1: Criteria to help define NRSI study design features for inclusion in systematic reviews of interventions of healthcare interventions (developed from Section 24.2.2 of the *Cochrane Handbook*)

Potential feature for eligibility criteria	Example reviews that used this to define NRSI eligibility criteria	Notes
1) Based on how the intervention effect was measured It should be clear whether data from a comparison group (i.e. the group that received another intervention) are collected: (a) contemporaneously (i.e., the intervention effect was measured as the difference between groups); (b) historically (i.e., the intervention effect was measured as within-group difference over time), or; (c) contemporaneously and historically (also called the "double-difference") (i.e., the intervention effect was measured as the difference between groups in the within-group change over time). **See *Effect measurement* in Annex 1.**	**Example review**: Pool fencing for preventing drowning of children (restricts eligibility to studies with a contemporaneous or historical comparison group). **Example review:** Smoking cessation for secondary prevention of cardiovascular disease (restricts eligibility to cohort studies that measure outcomes with at least a six-month follow-up from baseline).	Regarding the use of historical comparisons, the single-group (uncontrolled) pre-test post-test design is usually excluded from systematic reviews. With very few exceptions – specifically, where the effect happens immediately following the intervention, is large and is unlikely to be confounded (see Victora et al., 2004) – usually, a critically high risk of confounding will be present in these types of study, necessitating the incorporation of studies with contemporaneous (controlled) comparisons or, in the case of historical comparisons, interrupted time-series design (see also **Annex 1**).
2) Based on how groups were formed Participants' allocation to groups (e.g., treatment and comparison) may be determined by an independent third party, a researcher, a policymaker, the patient's clinician or the patient themselves. The most rigorous NRSI are usually those where group formation is determined by someone completely independent from the research.	**Example review:** Farmer field schools for improving farming practices and farmer outcomes (restricts eligibility to studies using known rules to allocate treatment, such as a threshold on a scaled baseline measure as in the case of discontinuity designs, or other methods of allocation,	We are usually more confident about NRSI when the allocation method is known and can be modelled appropriately. In contrast, studies where the allocation method is unknown – for example, where groups are formed purely by self-selection of patients – may produce treatment effect estimates that are critically biased. Therefore, systematic reviews that are very inclusive according to this characteristic should carefully assess eligible studies using appropriate risk of bias tools (e.g., tools that distinguish studies with 'critical risk of bias' that are subsequently excluded from evidence synthesis).[4]

[4] Links to PROJECT 1, PROJECT 4.

See *Allocation method* in Annex 1.	including selection by planners, practitioners and participants).	
3) Based on when outcome data were collected and from which participants In NRSI that require one or more pre-tests (also called baseline measures), it is important that pre-test measurement is done before the treatment has been implemented. In some longitudinal studies, the same participants are followed up in successive rounds of data collection (also called "panel data"), whereas in others, data from repeated cross-sections or health episodes are collected from different (or some different and some same) participants in each round. Data may be collected or analysed at individual patient or group levels (e.g., health facility, district or other). **See *Data collection* in Annex 1.**	**Example review**: Reducing medication errors for adults in hospital settings (restricts eligibility to studies with outcome data available at multiple times before and multiple times after the intervention, including controlled studies with the same measurements in a comparison group).	If inclusion in a review requires the same participants to be followed up throughout the study, cross-sectional and interrupted time series designs that measure outcomes in different individuals at different time points would not be eligible.
4) Based on the ability of the study to address different forms of confounding Observable confounders are factors that are measured and considered in analysis in primary studies, such as prognostic factors like sociodemographic characteristics and geographical location. Unobservable confounders are factors that are unmeasurable (or not typically measured) in primary studies, such as individual participant or group motivation and aptitudes. NRSIs that can potentially address unobservable sources of confounding by design include	**Example review:** Paying for performance to improve the delivery of health interventions in low- and middle-income countries (restricts eligibility to NRSI studies where the choice of the comparison site was appropriate, i.e., similar socioeconomic characteristics, no major differences evident in the baseline groups, or both).	Confounding occurs when there are common causes of intervention group assignment and outcome (a particularly important issue, which can affect different types of NRSI in different ways). Understanding and accounting for confounders, or controlling for them, is important for inferring causality. For studies addressing observable confounding only, you may want to pre-specify that you will only include NRSI *if* they take specific confounders (e.g., age, sex, location, socio-economic status) into account, when it would be relevant to do so. These confounders should be discussed within the author team and listed in the study protocol.

discontinuity designs, and studies designed or analysed using instrumental variables (e.g., Mendelian randomisation). In addition, some studies analysing controlled before and after data can address sources of unobservable confounding that are fixed over the course of a study, such as innate ability, at the unit of analysis (e.g., the patient or health episode); however, these types of studies are not able to address sources of unobservable confounding that might vary over the course of a study, including participant motivation. Other types of study (and otherwise rigorous designs with problems in conduct) are likely to address observable sources of confounding only. **See also *Type of confounding addressed* in Annex 1.**		Bias due to confounding, selection bias, information bias, reporting bias and other sources of bias will also be systematically evaluated in risk-of-bias assessments.[5]

[5] Links to PROJECT 1, PROJECT 4.

Annex 1: Glossary of study designs and their main features

Approach	*Definition*	*Effect measurement*	*Allocation to treatment*	*Data collection*	*Type of confounding addressed*
Randomised controlled trial (RCT) design	RCTs have the essential features of random assignment of study participants to two or more treatment conditions, such as treatment and control. RCTs use a range of design types, including simple intervention and control (no intervention) group designs, multiple interventions head-to-head (with or without control group) designs, and factorial (A, B, A+B) designs. Variations include cluster-randomised designs that randomly assign groups of study participants or other units such as health facilities, block-randomised designs that stratify by characteristics of interest, randomised cross-over designs that swap treatment regimens across groups over time, and so on (see Higgins et al., 2019).	Contemporaneous control[1]	Randomisation	At least one post-test[5] and often one pre-test[6]	Unobservable[7] and observable[8]
Quasi-randomised controlled trial	A study that uses a known, systematic method of assignment to treatment, such as alternation of participants ordered alphabetically to treatment groups, Some natural experiments also use this approach, for example where groups are created through administrative targeting errors.	Contemporaneous control	Forcing variable[4] (e.g., alternation)	At least one post-test and often one pre-test	Unobservable and observable
Discontinuity design, also called regression discontinuity design (RDD)	A design where treatment assignment is explicitly known to the researcher and based on an explicit threshold on a scaled baseline measure. This may be a pre-test for the outcome measure or another measure suitable for determining who receives treatment; common assignment variables include participant age, income, location with respect to an administrative boundary, or performance on a baseline assessment. Fuzzy regression discontinuity is a subtype that exploits a natural discontinuity in the probability or rate of participation in a treatment program at some cut-point on a baseline measure. In this case the treatment assignment method is not under the control of, or	Contemporaneous comparison[2]	Forcing variable[4] (scale threshold)	At least one post-test and sometimes a pre-test	Unobservable and observable

Approach	*Definition*	*Effect measurement*	*Allocation to treatment*	*Data collection*	*Type of confounding addressed*
	perfectly known to, the researcher. Fuzzy regression discontinuity is often done retrospectively using instrumental variable analysis methods. For a systematic review on the correspondence of effect size estimates in RCTs and discontinuity designs of the same interventions and populations, see Sharma Waddington et al. (2022).				
Instrumental variables (IV) design	A statistical method often applied to cross-section data, which uses a pre-treatment variable, called an instrument, that is predictive of who gets treatment but has no direct causal effect on the outcome. This is often misunderstood to mean that the instrumental variable cannot be correlated with the outcome; ideally it will be strongly correlated with the outcome but logically cannot be a causal driver of the outcome. That is, the instrumental variable's effect on the outcome is only indirect and occurs through the treatment variable. A good example of this design is "Mendelian randomisation" which uses a genetic marker measured at pre-test as the instrument. Although often done in retrospective study designs, instrumental variable analysis is also applied in the context of RCTs (where the instrument is the random assignment variable) to estimate the unbiased effect of adhering to treatment. Where intervention eligibility is universal, but knowledge about eligibility is not, an instrumental variable is sometimes used prospectively, using "randomised encouragement design". This is done by randomly assigning information about the intervention, which is then analysed using IV estimation.	Contemporaneous comparison	Forcing variable[4] (instrument)	At least one post-test	Unobservable and observable
Interrupted time series (ITS) design	A design that usually relies on aggregate data, such as the monthly mortality examined for some number of time periods prior to the start of the intervention and some number of time periods after the start of the	Historical comparison[3]	Forcing variable[4] (time)	Multiple pre-tests and multiple post-tests	Unobservable and observable

Approach	*Definition*	*Effect measurement*	*Allocation to treatment*	*Data collection*	*Type of confounding addressed*
	intervention. There is often only one observation per observational time point, but there can be many (e.g., multiple patients or health episodes per observational time point). There is discussion about the minimum number of observations before and after treatment for valid inference. For example, EPOC (n.d.) specified there should be at least three pre-tests and three post-tests sufficient to establish a trend in outcomes before and after treatment. However, it is likely that more observations are needed in valid interrupted time-series design (see Fretheim et al., 2015).				
Controlled-ITS design	This design is an augmentation of ITS with one or more comparison series. This helps establish that any treatment effect observed in the treatment series is not the result of broader historical changes.	Contemporaneous and historical comparison	Forcing variable (time), with treatment group selection by planners, practitioners or patients	Multiple pre-tests and multiple post-tests	Unobservable and observable
Non-equivalent groups with pre-test and post-test design (including controlled-before and after (CBA) and non-randomised controlled trial (NRCT) designs)	This design includes a program or treatment group and a non-randomly created comparison group. Its essential feature is a baseline measure of the outcome (a pretest), and it may also have other baseline measures, such as sociodemographic factors. These baseline measures can be used in statistical models to adjust for observable confounding, as well as unobservable confounders that are typically fixed over the course of a study (that is, they are "time-invariant"), such as intellectual ability. This design type is commonly used in impact evaluations of public health, social, economic, and educational interventions and policies. The design may also be combined with matching of groups at pre-test using statistical methods (e.g., propensity score matching). Data may be collected from panels of the same units (individual participants or groups) over time, or from repeated cross-sections of different units (or	Contemporaneous and historical comparison	Selection by planners, practitioners or patients	At least one pre-test and one post-test	Time-invariant unobservable and observable

Approach	*Definition*	*Effect measurement*	*Allocation to treatment*	*Data collection*	*Type of confounding addressed*
	some different and some same) over time. Data may be collected and analysed at individual participant level (sometimes called NRCT designs) or aggregated (e.g., health facility) levels (sometimes called CBA designs).				
Non-equivalent groups, post-test only design (sometimes called cross-section design)	This design includes a program or treatment group and a non-randomly created comparison group. In its simplest form, this design only has a single assessment of the outcome following program participation, although variations on this design might include follow-up assessments. A variant is the cross-sectional retrospective design, which is based solely on existing data where the temporal precedence between the receipt of treatment and the outcome may be ambiguous.	Contemporaneous comparison	Selection by planners, practitioners or patients	One post-test	Observable only
Case-control design	The typical case-control design selects participants based on the outcome; that is, the researcher identifies a set of cases and a set of comparators. Cases are individuals who exhibit the outcome (generally undesired, such as having cancer or another disease, or dying) and comparison individuals are those who do not exhibit the outcome. Exposure to a prior variable of interest, such as a health care intervention, is assessed. This design is widely used in epidemiological studies to identify potential causal factors for various diseases. However, it is also used in the context of treatment effectiveness research to examine potential negative long-term outcomes of treatment.	Contemporaneous comparison	Selection by planners, practitioners or patients	One post-test	Observable only
Synthetic control design	A method usually applied to instances where there is no natural comparison group (e.g., because of the universality of access or measurement at an aggregated level such as the country), where a comparison group is created statistically from multiple external groups, usually from outside of the sample population (e.g., a variety of external administrative units or countries).	Contemporaneous comparison	Selection by planners, practitioners or patients	At least one post-test and sometimes a pre-test	Observable only

Approach	*Definition*	*Effect measurement*	*Allocation to treatment*	*Data collection*	*Type of confounding addressed*
Cohort design (sometimes called non-equivalent group cohort design)	A method where a treated and an untreated group are followed up over time, and where the data collection points by group may or may not be contemporaneous. Non-randomised cross-over studies may also use cohort design data structures.	Contemporaneous comparison or historical comparison	Selection by planners, practitioners or patients	At least one post-test and sometimes a pre-test	Observable only
Single group pre-test post-test design (also called uncontrolled before versus after (BA) design or reflexive control design)	The typical BA design selects study participants that receive the treatment, and their status on one or more outcome variables is assessed before and after treatment participation. This design might also include follow-up assessments. Owing to the absence of a contemporaneous comparison, the design is often considered limited in its ability to address sources of confounding like another intervention happening concurrently or any changes over time due to the `state-of-the-world' that may be mistaken for a treatment effect (also called a maturation effect).	Historical comparison	Selection by planners, practitioners or patients	One pre-test and at least one post-test	Observable in certain circumstances, otherwise none

Sources: authors drawing on Reeves et al. (2017) and the Campbell Collaboration (Wilson et al., 2024).

Definitional notes:
1 Contemporaneous control: measurement of a group that does not receive the intervention of interest (or receives something else) at the same time as the treatment group in randomised controlled trials.
2 Contemporaneous comparison: measurement of a group that does not receive the intervention of interest (or receives something else) at the same time as the treatment group in non-randomised studies of interventions.
3 Historical comparison: measurement at a time before the intervention has occurred in non-randomised studies of interventions.
4 Forcing variable: a variable that determines treatment allocation at a particular value or threshold in non-randomised studies of interventions.
5 Post-test: measurement after the intervention has been implemented.
6 Pre-test: measurement before the intervention has been implemented.
7 Unobservable confounding: factors affecting both treatment allocation and the outcome which cannot be (or are not) measured.
8 Observable confounding: measured factors affecting both treatment allocation and the outcome.

Document version control:

Document created:	March 2022
Document last updated:	April 2024
Document version:	1.2
Author details	**Lindson N** (co-lead), Nuffield Department of Primary Care Health Sciences, University of Oxford, UK **Hartmann-Boyce J** (co-lead), Department of Health Promotion and Policy, University of Massachusetts, Amherst, USA; Nuffield Department of Primary Care Health Sciences, University of Oxford, UK **Bero L**, Center for Bioethics and Humanities, University of Colorado—Anschutz Medical Campus, Denver, Colorado, USA **Dumville J**, School of Health Sciences, University of Manchester, UK **Livingstone N**, Cochrane Central Executive, UK **Reeves B,** Bristol Medical School, University of Bristol, UK **Tugwell P**, University of Ottawa, Department of Medicine, Faculty of Medicine, Ottawa, Canada; Bruyère Research Institute, Ottawa, Canada; Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, Canada; University of Ottawa, School of Epidemiology and Public Health, Faculty of Medicine, Ottawa, Canada; WHO Collaborating Centre for Knowledge Translation and Health Technology Assessment in Health Equity, Bruyère Research Institute, Ottawa, Canada **Flemyng E,** Cochrane Central Executive, UK **Wilson DB,** George Mason University, USA **Sharma Waddington H** (co-lead)
Declarations	**Nicola Lindson** was previously a Managing Editor for the Cochrane Tobacco Addiction. **Jamie Hartmann-Boyce** was previously an editor for the Cochrane Tobacco Addiction. **Jo Dumville** was previously Coordinating Editor for Cochrane Wounds. **Nuala Livingstone** and **Ella Flemyng** are employed by Cochrane. **Lisa Bero** is Cochrane's Senior Research Integrity Editor and a Cochrane Editorial Board member. **David B Wilson** was previously a Campbell Methods Editor and is Member of the Steering Group of the Campbell Crime and Justice Coordinating Group. **Peter Tugwell** is a Cochrane editor, Cochrane Editorial Board member, Co-Convenor of Cochrane Equity Methods Group and Co-Chair of the Campbell International Development Coordinating Group (IDCG). **Barney Reeves** is a Cochrane editor and Co-Convenor of Cochrane's NRSI Methods Group and co-author of the *Cochrane Handbook for Systematic Reviews of Interventions* chapter on using non-randomised studies of interventions. **Hugh Sharma Waddington** is Co-Convenor of Cochrane's NRSI Methods Group, a Campbell Methods Editor, and was previously Co-Chair and Editor for the Campbell Collaboration IDCG. This project was funded in part by Cochrane as part of the "Improving the quality of Cochrane reviews in public health" program. The authors declare no other conflicts of interest.
Contact for queries:	Hugh Sharma Waddington: hugh.waddington@lshtm.ac.uk Jamie Hartmann-Boyce: jamie.hartmann-boyce@phc.ox.ac.uk Nicola Lindson: nicola.lindson@phc.ox.ac.uk

Supporting references and further reading

**Cited in this document

**Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomised studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Determining study design eligibility

Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annual review of public health. 2017 Mar 20;38:39-56.

**Crowley P, Chalmers I, Keirse MJ. The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. Br J Obstet Gynaecol. 1990 Jan;97(1):11-25. doi: 10.1111/j.1471-0528.1990.tb01711.x. PMID: 2137711.

**Cuello-Garcia CA, Santesso N, Morgan RL, Verbeek J, Thayer K, Ansari MT, Meerpohl J, Schwingshackl L, Katikireddi SV, Brozek JL, Reeves B, Murad MH, Falavigna M, Mustafa R, Deborah L, Regidor DL, Alexander PE, Garner P, Akl EA, Guyatt G, Schünemann HJ. GRADE guidance 24 optimizing the integration of randomised and non-randomised studies of interventions in evidence syntheses and health guidelines. J Clin Epidemiol. 2022; 142: 200-208. Doi: 10.1016/j.jclinepi.2021.11.026. ISSN 0895-4356.

de Vocht F, Katikireddi SV, McQuire C, Tilling K, Hickman M, Craig P. Conceptualising natural and quasi experiments in public health. BMC medical research methodology. 2021 Dec;21(1):1-8.

Detterbeck FC, Gould MK, Lewis SZ, Patel S. Extending the reach of evidence-based medicine: A proposed categorization of lower-level evidence. Chest. 2018 Feb 1;153(2):498-506.

**Dwan K, Kirkham J, Paton RW, Morley E, Newton AW, Perry DC. Splinting for the non-operative management of developmental dysplasia of the hip (DDH) in children under six months of age. Cochrane Database of Systematic Reviews 2022, Issue 10. Art. No.: CD012717. DOI: 10.1002/14651858.CD012717.pub2.

**Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, Goodacre S, Herrin J, Kerse N, McKinlay RJ, Wright A, Soumerai SB. A reanalysis of cluster randomised trials showed interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2015 Mar;68(3):324-33. doi: 10.1016/j.jclinepi.2014.10.003. Epub 2014 Dec 11. PMID: 25499983.

**Gallo L, Eskicioglu C, Braga LH, Farrokhyar F, Thoma A. Users' guide to the surgical literature: how to assess an article using surrogate end points. *Canadian Journal of Surgery*. 2017, 60 (4) 280-287; DOI: 10.1503/cjs.002217

** Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. Int J Epidemiol. 2005 Aug;34(4):874-87. doi: 10.1093/ije/dyi088. Epub 2005 Apr 20. PMID: 15843394.

** Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011

May;8(5):e1001026. doi: 10.1371/journal.pmed.1001026. Epub 2011 May 3. PMID: 21559325; PMCID: PMC3086872.

Handley MA, Lyles CR, McCulloch C, Cattamanchi A. Selecting and improving quasi-experimental designs in effectiveness and implementation research. Annual review of public health. 2018 Apr 1;39:5-25.

Hernán MA, Robins JM. Using big data to emulate a target trial when a randomised trial is not available. American journal of epidemiology. 2016 Apr 15;183(8):758-64.

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons; 2019 Sep 23. *In particular, readers may wish to visit Chapters 2 (*https://training.cochrane.org/handbook/current/chapter-02) and *3 (section 3.3 in particular:* https://training.cochrane.org/handbook/current/chapter-03#a-33-determining-which-study-designs-to-include)*.*

Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. Journal of Clinical Epidemiology. 2015;68 (2):132-143. https://doi.org/10.1016/j.jclinepi.2014.06.021.

**OCEBM Levels of Evidence Working Group*. "The Oxford Levels of Evidence 2". Oxford Centre for Evidence-Based Medicine. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JPT, Mathes T, Pfadenhauer LM, Rehfuess EA. Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews. Journal of clinical epidemiology. 2017;91:56-69. https://doi.org/10.1016/j.jclinepi.2017.07.008.

**Saldanha IJ, Skelly AC, Ley KV, Wang Z, Berliner E, Bass EB, Devine B, Hammarlund N, Adam GP, Duan-Porter D, Bañez LL, Jain A, Norris SL, Wilt TJ, Leas B, Siddique SM, Fiordalisi CV, Patino-Sutton C, Viswanathan M. Inclusion of Nonrandomized Studies of Interventions in Systematic Reviews of Intervention Effectiveness: An Update [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2022 Sep. Report No.: 22-EHC033. PMID: 36153934.

Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, Shea B, Wells G, Helfand M. Non-randomised studies as a source of complementary, sequential or replacement evidence for randomised controlled trials in systematic reviews on the effects of interventions. Res Synth Methods. 2013 Mar;4(1):49-62. doi: 10.1002/jrsm.1078.

**Sharma Waddington H, Fenton Villar P, Valentine J. Can Non-Randomised Studies of Interventions Provide Unbiased Effect Estimates? A Systematic Review of Internal Replication Studies. Evaluation Review. 2022;47(3):563-593, https://doi.org/10.1177/0193841X221116721.

**Sharma Waddington H, Masset E, Bick S, Cairncross S. Impact on childhood mortality of interventions to improve drinking water, sanitation, and hygiene (WASH) to households: Systematic review and meta-analysis. PLOS Medicine. 2023 April;20(4):e1004215. https://doi.org/10.1371/journal.pmed.1004215.

**Shekelle PG, Maglione MA, Luoto J, et al. Global Health Evidence Evaluation Framework [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan. Table B.9, NHMRC Evidence Hierarchy: designations of 'levels of evidence' according to type of research question

(including explanatory notes) Available from: https://www.ncbi.nlm.nih.gov/books/NBK121300/table/appb.t21/

**Verbeek JH, Rajamaki B, Ijaz S, Sauni R, Toomey E, Blackwood B, Tikka C, Ruotsalainen JH, Kilinc Balci FS. Personal protective equipment for preventing highly infectious diseases due to exposure to contaminated body fluids in healthcare staff. Cochrane Database of Systematic Reviews 2020, Issue 5. Art. No.: CD011621. DOI: 10.1002/14651858.CD011621.pub5.

**Victora CG, Habicht J-P and Bryce J. Evidence-based public health: moving beyond randomised trials. American Journal of Public Health. 2004;94(3):400-405.

**Welch, VA et al. Deworming and adjuvant interventions for improving the developmental health and well-being of children in low- and middle-income countries: a systematic review and network meta-analysis. Campbell Systematic Reviews. 2016;12(1):1-383. https://onlinelibrary.wiley.com/doi/full/10.4073/csr.2016.7

Defining study designs

Aquilina J, Neves JB, Tran MG. An overview of study designs. British Journal of Hospital Medicine. 2020 May 2;81(5):1-6.

Brown EM. Optimal designs of clinical studies in the context of their inclusion in systematic reviews. Journal of Antimicrobial Chemotherapy. 2021 Oct;76(10):2498-500.

**Cochrane's COVID-19 Study Register. About page: https://community.cochrane.org/about-covid-19-study-register (How do we define our study characteristics?)

**EPOC Study designs guidance. https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/EPOC%20Study%20Designs%20About.pdf

**Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. The Lancet. 2002 Jan 5;359(9300):57-61.

**Higgins JP, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, Tugwell P and Wells G. Issues relating to study design and risk of bias when including non-randomised studies in systematic reviews on the effects of interventions. Res. Syn. Meth. 2013;4:12-25. https://doi.org/10.1002/jrsm.1056

**Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons; 2019 Sep 23. *In particular, Chapters 24 (*https://training.cochrane.org/handbook/current/chapter-24) and *23* (https://training.cochrane.org/handbook/current/chapter-23).

Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. Journal of Clinical Epidemiology. 2017 Sep 1;89:30-42.

**Shadish W, Cook T, and Campbell D. Experimental and quasi-experimental designs for generalized causal inference. 2002. BROOKS/COLE CENGAGE Learning.

**Vigneri M, Clarke M, Exley J, Tugwell P, Welch V, White H. Epidemiology and development economics two sides of the same coin in impact evaluation. Journal of Clinical Epidemiology. 2022 April;144:16-21. https://doi.org/10.1016/j.jclinepi.2021.11.029.

**Wilson D, Aloe AM, Pigott TD, Stewart G, Tugwell P, Welch V, Sharma Waddington H. (2024). A study design nomenclature for systematic reviews assessing the effectiveness of interventions. The Campbell Collaboration (under review).

**Wilson DB, Gill C, Olaghere A, McClure D. Juvenile Curfew Effects on Criminal Behavior and Victimization: A Systematic Review. Campbell Systematic Reviews. (2016);12(1):1-97. https://doi.org/10.4073/csr.2016.3

Equity considerations

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons; 2019 Sep 23. *In particular, Chapter 16 (*https://training.cochrane.org/handbook/current/chapter-16).

Main C, Mather L, Ogilvie D, Petticrew M, Sowden A, Thomas S, Whitehead M, Egan M. The inverse evidence law in tobacco control: findings from a systematic review of reviews. Journal Of Epidemiology & Community Health. 2005 Sep;59.

Ogilvie D, Egan M, Hamilton V, Petticrew M. Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go?. J Epidemiol Community Health. 2005;59(10):886-892. doi:10.1136/jech.2005.034199

**O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, Evans T, Pardo Pardo J, Waters E, White H, Tugwell P. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. Journal of Clinical Epidemiology. 2014; 67(1):56-64. doi:10.1016/j.jclinepi.2013.08.005

Tugwell P, Petticrew M, Kristjansson E, Welch V, Ueffing E, Waters E et al. Assessing equity in systematic reviews: realising the recommendations of the Commission on Social Determinants of Health BMJ 2010; 341:c4739. doi:10.1136/bmj.c4739.

Welch VA, Petticrew M, O'Neill J et al. Health equity: evidence synthesis and knowledge translation methods. Syst Rev 2, 43 (2013). https://doi.org/10.1186/2046-4053-2-43.