Appendix A:
Approaches to Assessing the Risk of Bias in Studies

Approaches to critical appraisal of study methodology and related terminology has varied and is evolving. Overlapping terms include quality, internal validity, risk of bias, or study limitations, but a central goal is an assessment of the validity of the findings. We use the phrase "assessment of risk of bias" as the most representative of the goal of evaluating the degree to which the effects reported by a study represent the "true" causal relationship between exposure and outcome.

A valid estimate requires the absence of bias or systematic error in selection, performance, detection, measurement, attrition, and reporting and adequacy in addressing potential confounding. The interpretation of the effect of an estimate also requires the evaluation of precision (the absence of random error through adequate study size and study efficiency).[1] For studies that do not lend themselves to quantitative pooling of estimates, reviewers will likely be making assessments of individual study risk of bias and precision at the same time, supporting the broader notion of evaluating study quality. Thorough assessment of these threats to the validity of an estimate is critical to understanding the believability and interpretation of a study.

Table 1 builds on and revises, drawing on numerous sources,[1-3] the Cochrane Collaboration taxonomy[4] of threats to validity and precision to expand the discussion of confounding and selection bias.

The inclusion of observational studies considerably expands the challenges in establishing causal inference in systematic reviews. Observational studies cannot, by design, offer establish causality through features such as randomization and concealment of allocation. They are therefore at greater risk than RCTs for confounding by indication and selection bias.

In contrast, threats to validity and precision from performance bias, detection bias, inadequate sample size, and lack of study efficiency do not differ markedly in theory between RCTs and observational studies (although some features such as blinding of assessors that protect against detection bias are more likely in experimental designs than in observational studies). For both designs, these risks of bias can threaten the validity of results. Performance bias and detection of effects have the potential to alter effect sizes unpredictably and need to be evaluated as well in observational studies.[4, 5] These sources of bias can invalidate the results of observational studies, as can problems with confounding and selection bias. In relation to risks from confounding, Greenland and Morgenstern note that "[b]iases of comparable or even greater magnitude can arise from measurement errors, selection (sampling) biases, and systematically missing data, as well as from model-specification errors. Even when confounding and other systematic errors are absent, individual causal effects will remain unidentified by statistical observations."[6, p. 208]

Table 1. Threats to validity and precision[a]

Threats	Definition
Threats to validity (systematic error)	
Confounding by indication or allocation bias	Systematic differences between baseline characteristics of the groups that arise when patient prognostic characteristics, such as disease severity or comorbidity, influence both treatment source and outcomes. Confounders are the common cause for intervention and exposure; they occur before exposure. Confounding by indication can occur from self-selection of treatments or physician-directed selection of treatments.
Selection bias	Selection bias occurs when studies are conditioned on (that is, they differentially select for) common effects of the exposure and the outcome. Selection bias occurs after exposure and arises when the association between exposure and outcome is different for those who participate compared with those who do not participate in a study (i.e., all those who are theoretically eligible). Includes inappropriate selection of controls in a case-control study, differential loss to follow-up for groups being compared (attrition bias), incidence-prevalence bias, nonresponse bias, and in- or exclusion of specific groups for study.
Performance bias	Systematic differences in the care provided to participants and protocol deviation. Examples include contamination of the control group with the exposure or intervention, unbalanced provision of additional interventions or co-interventions, difference in co-interventions, and inadequate blinding of providers and participants.
Detection bias	Systematic differences in outcomes assessment among groups being compared, including misclassification of the exposure or intervention, covariates, or outcomes because of variable definitions and timings, diagnostic thresholds, recall from memory, inadequate assessor blinding, and faulty measurement techniques. Erroneous statistical analysis might also affect the validity of effect estimates.
Reporting bias	Systematic differences between reported and unreported findings (e.g., differential reporting of outcomes or harms, incomplete reporting of study findings, potential for bias in reporting through source of funding).
Threats to precision (random error)	
Inadequate study size	Not powered to test study hypothesis for an individual study. (Multiple underpowered small studies can result in a precise meta-analysis estimate)
Lack of study efficiency	Absence of needed stratification in design. When confounding and effect modifiers do not exist, an equal apportionment ratio between exposed and unexposed is the most efficient design. Comparisons within strata may be required to account for known confounders and effect modifiers. Matching on stratification variables allows for an efficient design.

[a] From Hernan et al., 2004,[2] Rothman et al., 2003,[1] Higgins et al., 2006,[4] and Viswanathan et al., 2012.[3]

Confounding

Although confounding is more of a concern for observational studies than RCTs, RCTs are not entirely free of these concerns. Confounding that is random (by chance) is expected to be equally distributed between arms by randomization, but RCTs may not always successfully randomize such potential confounders. Confounding by chance (that is confounding that is unknown, unmeasured, or poorly measured, but expected to be equally distributed) should occur with the same probability in RCTs and observational studies, because, it is, by definition, occurring by chance.[7] Confounding by indication or contraindication, on the other hand, occurs when both treatment and outcome are influenced by a third factor, namely prognosis. Confounding by indication can occur when the expectation of prognosis influences the patient or provider's selection of treatment as well as the potential outcome. Trials, by virtue of concealed randomization, avoid this source of confounding, by "breaking the link between prognosis and prescription."[8, p. e67] Observational studies of benefits that cannot address source of confounding by design (through, for instance, restriction to patients with the same prognosis) must account for it in analysis to the extent possible. Vandenbroucke notes that confounding by contraindication is not always a concern for observational studies of harms: harms are often unanticipated outcomes, so an expectation of prognosis for harms is unlikely to have influenced the selection of treatment.[8, p. e67] The extent to which confounding by indication may occur in observational

studies of harms lies on a gradient, from the completely unanticipated (as with ACE inhibitors and angioneurotic edema) to the potentially likely (as with cardiac arrhythmias induced by anti-arrhythmic drugs).

Confounding by indication may also occur when the person allocating treatment is influenced by factors other than prognosis. Shrier et al. offer the example of the appearance of fatigue causing the physician to select one diabetes treatment over another.[7] When confounding by indication can be controlled, for instance, by the inclusion of physician-rated appearance of patient fatigue as a covariate in modeling, the effect of the potential confounder can be accounted for. Each analysis need only account for those confounders that are expected to have an effect on the outcome and that have not already been accounted for by the inclusion of other closely correlated variables. Thus, the assessment of the potential for bias from confounding requires context-specific understanding of the relationship between treatment and outcomes and study-specific evaluation.[1, 6, 7]

Selection Bias

Selection bias refers to the selection of the subset of those eligible, when that selection is conditioned upon variables that are the common effect of causes of the exposures and outcomes.[1, 2] The specific risks of selection bias vary by type of observational study design. Differential loss to follow-up, for instance, is a concern for cohort studies but not for case-series or cross-sectional studies. On the other hand, a choice of control group that is intrinsically either more or less exposed than the source population of the cases, is a specific form of selection bias in case-control studies.

The concerns regarding attrition bias, however, are similar for RCTs and observational studies in theory: both types of studies may be weakened by high rates of overall or differential attrition. In practice, observational studies may have higher risks of attrition bias: they may have longer time horizons and fewer resources to follow up with participants

Lastly, RCTs often suffer less from inadvertent selection bias caused by the analysis. Usually, in RCTs, investigators will not adjust for other consequences of the treatment, nor will they adjust for consequences of the outcome. However, in large data-base analysis, that is sometimes done unwittingly, as with restriction of the analysis on folic acid supplementation on congenital malformation to live births.[2]

Performance Bias

The risk of performance bias may differ in practice between RCTs and observational studies. In a RCT the intervention is typically standardized so that study participants within groups are exposed (for the most part) to similar interventions or control conditions. Further, co-interventions can be standardized and/or monitored and fidelity to the intervention can be assessed and reported. For observational studies, this standardization may not be possible; therefore, researchers may not have control over how the intervention was administered or the level of exposure. As a result, observational studies may not be able to clearly define intervention states. Hernan and Taubman note that when the principle of consistency (a causal contrast between two or more well defined interventions) is not met (as with studies exploring whether obesity leads to increased mortality), other requirements of caual inference such as exchangeability and positivity cannot be met.[9] This will also vary by specific observational study design features, particularly whether the intervention/exposure occurred prospectively vs. retrospectively with respect to the conduct of the study. Blinding of the providers and

participants may also be of variable concern across different designs. For instance, in a retrospective study, the intervention or exposure would likely have been administered outside the context of a research study; therefore, blinding of the intervention/exposure may not be applicable. In a prospective study, by contrast, blinding of providers and participants is critical to limiting bias that arises as a result of knowing the study hypothesis and what the study participants are receiving. Blinding, or other measures that make assessments objective, is possible in assessing the exposure status of individuals in case-control studies.

Detection Bias

As with performance bias, detection bias may be more problematic in observational studies because outcome assessment may not be standardized as it is more typically with RCTs. For example, in RCTs the same outcome assessment tools are used often with protocols for their implementation and assessment of results. As above, this source of bias can also vary across observational studies, particularly for prospective vs. retrospective designs. In retrospective designs where the measurement of outcomes has occurred prior to the start of the study, the researchers have no control over how those assessments were made, including choice of measurement tools, whether tools were valid and reliable (or a process to ensure their validity/reliability was employed), and how results were interpreted. Blinding of outcome assessors serves to limit detection bias, but not all designs can employ a blinded approach to assessing outcomes. In some studies, researchers serve as outcome assessors and can be blinded; in other studies, participants provide self-reported outcome data and cannot be blinded. As above, blinding of outcome assessors may be of variable concern across different observational study designs.

Information bias is related to detection bias and arises from how measurement and assessment of exposure and outcomes are made. In theory, all designs run the risk of bias from the use of poorly validated measures.

Reporting Bias

Reporting bias is a concern across all research regardless of design: authors are more likely to publish studies and selectively report outcomes that show statistical significance. However, with the recent emphasis on standardized reporting of RCTs and prospective trial registration, reporting bias may become less problematic with RCTs or at least more easily detected as a problem in specific RCTs. While no empirical evidence exists that reporting bias is a greater concern for observational studies, fewer standards exist for observational studies, making reporting bias a concern of equal or greater magnitude than for RCTs.

References

1. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott, Williams, & Wilkins; 2008.

2. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep; 15(5):615-25. PMID: 15308962.

3. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol. 2012 Feb; 65(2):163-78. PMID: 21959223.

4. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.0.2. London: The Cochrane Collaboration; 2009. Available at: www.cochrane-handbook.org. Accessed June 9, 2011.

5. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001 Jul 7; 323(7303):42-6. PMID: 11440947.

6. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001; 22:189-212. PMID: 11274518.

7. Shrier I, Boivin JF, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol. 2007 Nov 15; 166(10):1203-9. PMID: 17712019.

8. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008 Mar 11; 5(3):e67. PMID: 18336067.

9. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. Int J Obes (Lond). 2008 Aug; 32 Suppl 3:S8-14. PMID: 18695657.

Appendix B: Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures

This item bank is intended to evaluate the quality of studies examining the outcomes of interventions, treatments, or exposures. Eligible study designs include observational studies (cohort studies, case-control, case-series, and cross-sectional studies). It is not intended to rate the quality of studies concerning the accuracy of diagnostic tests. Abstractors can use the empty text box included with each question to document an explanation of their rating for later review. This may be particularly helpful in relation to a "cannot determine" response choice.

Study Definitions

Case series

Description: A study that describes a group of patients with a similar diagnosis and/or treatment. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of a condition through a collection of individual case reports.

Design features:

1. There is no comparison between groups to assess the effect/association of an intervention/exposure and an outcome.
2. There is no comparison with the same group over time.

Cross-sectional study

A study in which both the exposure and the outcome status in a target population are assessed concurrently that is, at the same point in time or during a brief period of time. The temporal sequence of cause and effect cannot necessarily be determined. They are most commonly used to assess prevalence. A common method for data collection is a survey.

Case control study

A study in which participants are selected based on the known outcome(s) of interest (e.g., disease, injury). Exposure status is then collected based on the participants' past experiences. Exposure status is compared between the two (or more) groups: those who have the outcome of interest and those who do not have the outcome of interest (controls). This is a retrospective study that collects data on events that have already occurred.

Cohort studies

A study in which individuals in the group without the outcome(s) of interest (e.g., disease) are classified according to exposure status (exposed or unexposed) and then are followed over time to determine if the development of the outcome of interest is different in the exposed and unexposed groups.

Questions to Assess the Risk of Bias (Q1-Q11)

Q1: Do the inclusion/exclusion criteria vary across the comparison groups of the study?
[PI: Drop question if not relevant to all included studies. To use this question for studies with one group, the focus of the question on comparison groups and related response categories would need to be changed to individuals.]

PI:

Yes, varies..... ☐

Partially: some, but not all criteria, applied to all groups or not clearly stated if some criteria are applied to all groups..... ☐

No, does not vary ..... ☐

Cannot determine: article does not specify..... ☐

Not applicable: study has only one group and so does not include comparison groups ..... ☐

Explanation for rating:

Q2: Does the strategy for recruiting participants into the study differ across groups? *[PIs: Drop question if not relevant to all included studies. To use this question for studies with one group, the focus of the question on comparison groups and related response categories would need to be changed to individuals.]*

PI:

Yes, differs ..... ☐

No, does not differ ..... ☐

Cannot determine ..... ☐

Not applicable: one study group ..... ☐

Explanation for rating:

Q3: Is the selection of the comparison group inappropriate, after taking into account feasibility and ethical considerations? *[PI: Provide instruction to the abstractor based on the type of study. Interventions with community components are likely to have contamination if all groups are drawn from the same community. Interventions without community components should select groups from the same source (e.g., community or hospital) to reduce baseline differences across groups. For case-control studies, controls should represent the population from which cases arose; that is, controls should have met the case definition if they had the outcome.]*

PI:

Yes, inappropriate ..... ☐

No, not inappropriate ..... ☐

Cannot determine or no description of the derivation of the comparison group ..... ☐

Not applicable: study does not include a comparison group (case series, one study group) ..... ☐

Explanation for rating:

Q4: Does the study fail to account for important variations in the execution of the study from the proposed protocol? *[PI: Consider intensity, duration, frequency, route, setting, and timing of intervention/exposures. Drop if not relevant for body of literature.]*

PI:

Yes, fails to account	☐	***Explanation for rating:***
Partially, fails to account	☐	
No, does not fail to account.....	☐	
Cannot determine.....	☐	
Not applicable: not an intervention study or no variations.....	☐	

Q5: Was the outcome assessor not blinded to the intervention or exposure status of participants? *[PI: There may be circumstances where clinical evaluators cannot be blinded to exposure status. Drop if not relevant to the body of literature.]*

PI:

Yes, not blinded.....	☐	**Explanation for rating:**
No, blinded	☐	
Not applicable: assessor cannot be blinded.....	☐	

Q6: Were valid and reliable measures, implemented consistently across all study participants used to assess inclusion/exclusion criteria, intervention/exposure outcomes, participant health benefits and harms, and confounding? *[PI: Important measures should be identified for abstractors and if there is more than one, they should be listed separately. PI may need to establish a threshold for what would constitute acceptable measures based on study topic. When subjective or objective measures could be collected, subjective measures based on self-report may be considered as being less reliable and valid than objective measures such as clinical reports and lab findings. Some characteristics may require that sources for establishing their validity and/or reliability be described or referenced. If so, provide instruction to abstractors.]*

PI:

Yes, valid and reliable measure used..... ☐

No, valid and reliable measure not used..... ☐

Cannot determine or measurement approach not reported..... ☐

Explanation for rating:

Q7: Was the length of follow-up different across study groups? *[Abstractor: When follow-up was the same for all study participants, the answer is no. If different lengths of follow-up were adjusted by statistical techniques, (e.g., survival analysis), the answer is no. Studies in which differences in follow-up were ignored should be answered yes.]*

PI:

Yes, different or cannot determine ..... ☐

No, not different or remedied through analysis... ☐

Not applicable: cross-sectional or only one group followed over time..... ☐

Explanation for rating:

Q8: In cases of high loss to follow-up (or differential loss to follow-up), was the impact assessed (e.g., through sensitivity analysis or other adjustment method)? *[PI: Attrition is measured in relation to the time between baseline (allocation in some instances) and outcome measurement for both retrospective and prospective studies and could include data loss from switching. Attrition rates may vary by outcome and time of measurement. Specify the criterion to meet relevant standards for the topic. Specify measurement period of interest, if repeated measures. Cochrane standard for attrition is 20 percent for shorter term (<1 year) and 30 percent for longer term (≥ 1 year).]*

PI:

Yes, impact assessed ..... ☐

No, impact not assessed ..... ☐

Cannot determine..... ☐

Not applicable: no loss to follow-up or loss to follow-up was not considered to be high, cross-sectional study, or case-control study selected on outcome ..... ☐

Explanation for rating:

Q9: Are any important primary outcomes missing from the results? *[PI: Identify all primary outcomes that one would expect to be reported in the study, including timing of measurement.]*

PI:

Yes, important outcome(s) missing ..... ☐

No important outcome(s) missing..... ☐

Cannot determine..... ☐

Explanation for rating:

Q10: Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results? *[PI: Identify all important harms that one would expect be reported in the study, including timing of measurement. Drop if not relevant to body of literature.]*

PI:

Yes, important outcomes missing..... ☐

No important outcomes missing..... ☐

Assessment of harms not applicable to this study..... ☐

Explanation for rating:

Q11: Are results believable taking study limitations into consideration? *[Abstractor: This question is intended to capture the overall quality of the study. Consider issues that may limit your ability to interpret the results of the study. Review responses to earlier questions for specific criteria.]*

PI:

Yes, believable..... ☐

No, not believable..... ☐

Explanation for rating:

Questions to Assess Confounding (Q6, Q12-13)

Q12: Any attempt to balance the allocation between the groups or match groups (e.g., through stratification, matching, propensity scores). *[PI: Drop if not relevant to the body of evidence.]*

PI:

Yes or study accounts for imbalance between groups through a post hoc approach such as multivariate analysis..... ☐

No or cannot determine..... ☐

Not applicable: study does not include a comparison group (case series or one study group)..... ☐

Explanation for rating:

Q13: Were important confounding variables not taken into account in the design and/or analysis (e.g., through matching, stratification, interaction terms, multivariate analysis, or other statistical adjustment such as instrumental variables)? *[PI: Provide instruction to abstractors on known confounding variables and inadequate adjustment for confounding for each outcome.]*

PI:

Yes, not accounted for or not identified ..... ☐

Partially: some variables taken into account or adjustment achieved to some extent..... ☐

No: taken into account… ..... ☐

Cannot determine..... ☐

Explanation for rating:

Questions to Assess Precision (Q14-16)

The following questions assess precision. We do not consider precision as a component of risk of bias, so these questions are not designed to assess or evaluate the risk of bias of individual studies. They may be used, however, for assessment of quality of individual studies when systematic reviewers are unable to pool studies quantitatively.

Q14: Is the sample size insufficiently large to detect a clinically significant difference of 5% or more between groups in at least one primary outcome measure? *[PI: Specify a different percent, if clinically relevant for each outcome of interest. PI may want to consider optimal information size (OIS) standards established by the GRADE Working Group in Guidelines #6 on imprecision.(Guyatt et al., 2011) Question relates to precision; reviewers whose evaluation of quality is limited to considerations of systematic error or risk of bias (not random error/precision) need not include this question. Reviewers who include both precision and systematic error in their evaluation of quality but rely on meta-analysis for pooled estimates need not include this question. PIs who choose to include considerations of precision in their assessment may include the question, but should be aware of the need for collaboration between clinical and statistical expertise in determining the threshold for a clinically adequate sample size.]*

PI:

Yes, imprecise..... ☐

No, sufficiently precise..... ☐

Explanation for rating:

Q15: Are the statistical methods used to assess the primary benefit outcomes inadequate?
[Abstractor: Question relates to precision and may not be relevant for systematic reviews that are able to pool data. The statistical techniques used must be appropriate to the data for each outcome and take into account issues such as controlling for dose-response, small sample size, clustering, rare outcomes, and multiple comparisons. In normally distributed data the standard error, standard deviation, or confidence intervals should be reported. In non-normally distributed data, inter-quartile range should be reported. For cohort studies, if the outcome has a greater than 10 percent prevalence, consider if the risk ratio and relative risk need to be calculated]

PI:

Yes, inadequate..... ☐

No, adequate..... ☐

Cannot determine..... ☐

Explanation for rating:

Q16: Are the statistical methods used to assess the main harm or adverse event outcomes inadequate? *[Abstractor: Question relates to precision and may not be relevant for systematic reviews that are able to pool data. The statistical techniques used must be appropriate to the data for each outcome and take into account issues such as controlling for dose-response, small sample size, clustering, rare outcomes, and multiple comparisons. In normally distributed data, the standard error, standard deviation, or confidence intervals should be reported. In non-normally distributed data, inter-quartile range should be reported.]*

PI:

Yes, inadequate..... ☐

No, adequate..... ☐

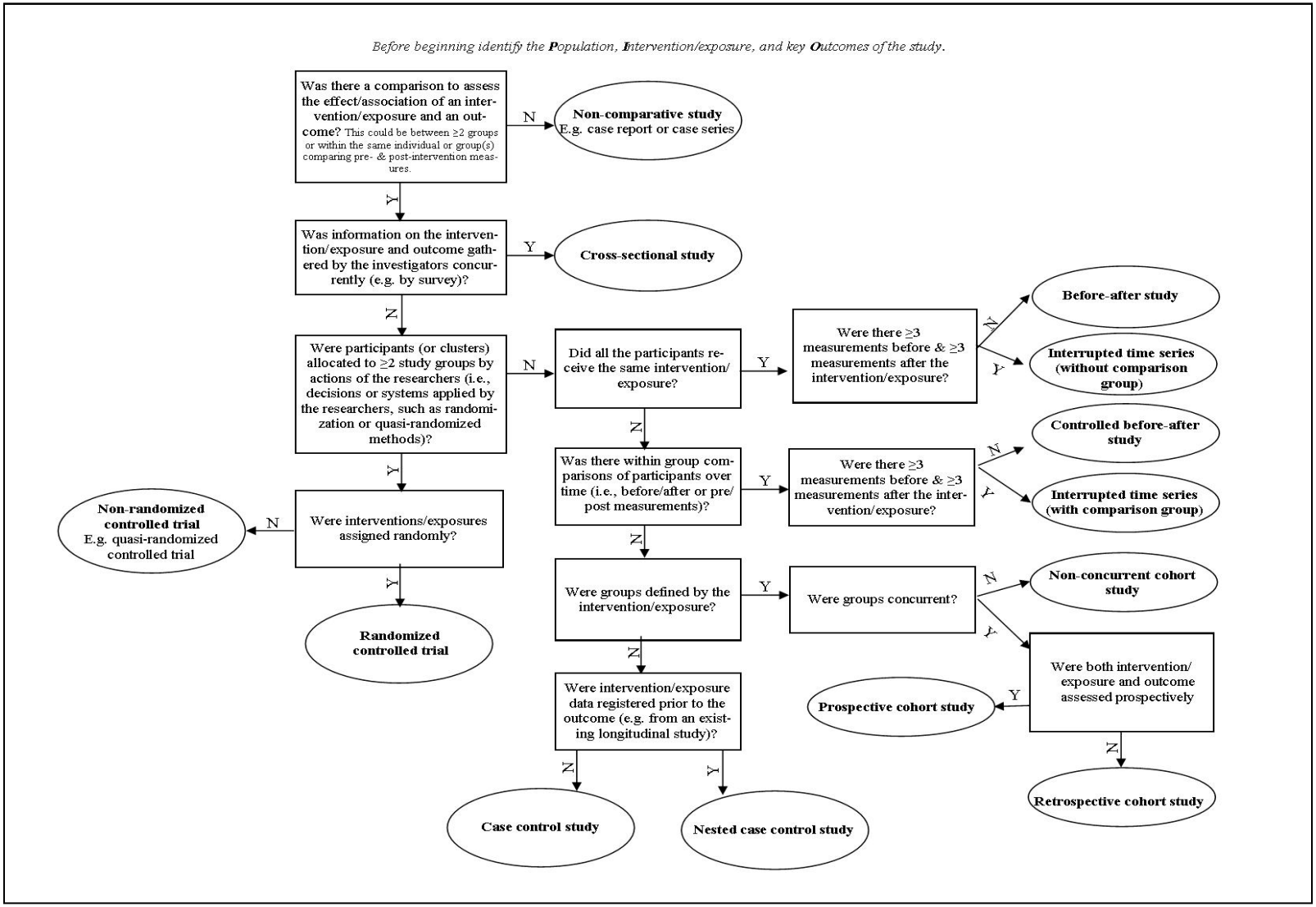
Not applicable: harms not reported..... ☐

Explanation for rating:

Modified from: Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol. 2012 Feb; 65(2):163-78. PMID: 21959223.

Appendix C: Taxonomy on Study Design

Users: To determine observational study design types, use the taxonomy and definitions below



DESIGN ALGORITHM FOR STUDIES OF INTERVENTIONS AND EXPOSURES

When using the algorithm, it is recommended that you do not rely on the design labels assigned by the authors of the report, but rather work through the questions in the algorithm based on the methods presented in the report and the definitions provided below.

Study Design Key

Below is a list of definitions that correlate with study designs assigned by the accompanying taxonomy. At the end of this list are some additional concepts that may be useful during study design classification.

Non-randomized trial

A study in which individuals or groups of individuals (e.g., community, classroom) are assigned to the intervention or control by a method that is not random (e.g., date of birth, date of admission, judgement of the investigator). Individuals or groups are followed prospectively to assess differences in the outcome(s) of interest. The unit of analysis is the individual or the group, as appropriate.

Randomized trial

A study designed to test the efficacy of an intervention on an individual, a group of individuals, or clusters (e.g., classrooms, communities). Individuals or clusters are randomly allocated to receive an intervention or control/comparison (e.g., placebo or another intervention) and are followed prospectively to assess differences in outcomes. The unit of analysis is the individual, group of individuals, or the cluster, as appropriate. Variations in treatment assignment and measurement produce different types of studies including factorial, cross-over, parallel, stepped wedge and Solomon four-group.

Prospective cohort study

A study in which individuals in the group without the outcome(s) of interest (e.g., disease) are classified according to exposure status at baseline (exposed or unexposed) and then are followed over time to determine if the development of the outcome of interest is different in the exposed and unexposed groups.

Retrospective cohort study

A study in which a group of individuals is identified on the basis of common features that were determined in the past. The group is usually assembled using available data sources (e.g., administrative data). Individuals are classified according to exposure status (exposed or unexposed) at the time the group existed and are followed up to a prespecified endpoint to determine if the development of the outcome of interest is different in the exposed or unexposed groups.

Interrupted time series with comparison group

A study in which multiple observations over time are “interrupted” by an intervention or exposure and in which two series are examined (one is a comparison group). There must be at least 3 observations before and at least 3 observations after the intervention or exposure for each group. The investigator(s) does not assign or have control over the intervention/exposure, which may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation, educational program, service delivery model) but does have control over the timing of the measurement and the variables being measured.

Controlled before-after study

A study in which the outcome(s) of interest is measured both before and after the intervention or exposure in two or more groups of individuals. In this study design the study group receives the intervention or exposure and the comparison group(s) does not. This type of study includes interventions that may be in the control of the investigator (e.g., a surgical procedure) as well as interventions that may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation). In all cases the investigator(s) has control over the timing of the measurement and the variables being measured.

Non-concurrent cohort study

A study in which 2 or more groups of individuals are identified on the basis of common features at different time points. Individuals in each group are classified according to exposure status (exposed or unexposed) at the time the groups existed or were created. They are followed to determine if the development of the outcome of interest is different in the exposed or unexposed groups.

Nested case control study

A study where exposed and control subjects are drawn from the population of a prospective cohort study. Baseline data are obtained at the time the population is identified; the population is then followed over a period of time. The study is then carried out using persons in whom the disease or outcome has developed and a sample of those who have not developed the outcome of interest (controls).

Case control study

A study in which participants are selected based on the known outcome(s) of interest (e.g., disease, injury). Exposure status is then collected based on the participants' past experiences. Exposure status is compared between the two (or more) groups: those who have the outcome of interest and those who do not have the outcome of interest (controls). This is a retrospective study that collects data on events that have already occurred.

Interrupted time series (without a comparison group)

A study in which multiple observations over time are "interrupted" by an intervention or exposure. There must be at least 3 observations before the intervention and at least 3 observations after the intervention; otherwise, the study is considered a before-after study. The investigator(s) does not assign or have control over the intervention/exposure, which may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation, educational program, service delivery model) but does have control over the timing of the measurement and the variables being measured.

Before-after study

A study of an intervention or exposure in which the investigator(s) compares the outcome(s) of interest both before and after the intervention in the same group of individuals. This includes interventions that may be in the control of the investigator (e.g., a surgical procedure) as well as interventions that may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation). In all cases the investigator(s) has control over the timing of the measurement and the variables being measured.

Cross-sectional study

A study in which both the exposure and the outcome status in a target population are assessed concurrently, that is, at the same point in time or during a brief period of time. The temporal sequence of cause and effect cannot necessarily be determined. They are most commonly used to assess prevalence. A common method for data collection is a survey.

Non-comparative study (case series)

Examples of this design include:

- A study that presents a description of a single patient or participant. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of the individual.
- A study that describes the experience of a group of patients with a similar diagnosis and/or treatment. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of a condition.
- A study in which data are collected at a series of points in time on the same population to observe trends in the outcome(s) of interest.

Additional Concepts

Cluster

The term 'cluster' refers to a unit of allocation or analysis in a clinical trial. Examples of clusters include hospitals, schools, neighborhoods, or entire communities.

Cluster randomized controlled trial

Synonym: *community trial; group randomized trial*

A randomized controlled trial in which the units of randomization and analysis are groups of people or communities (e.g., classroom, hospital, town). Typically, several communities receive the intervention and several different communities serve as controls.

Cohort

The term 'cohort' refers to a group of individuals (or other organizational units) who have a common feature when they are assembled (e.g., birth year, place of employment, medical condition, place or time period of medical treatment) and are followed over time. They can be followed prospectively or examined retrospectively.

Experimental study

A type of study in which investigators have direct control over the timing, course, and assignment of the intervention. Experimental studies investigate an intervention to determine its effect on the outcome(s) of interest. In an experimental study a population is selected to receive a specific intervention the effects of which are measured by comparing the outcomes in the experimental group with the outcomes of a control group that has received another intervention or placebo. Examples include randomized controlled trial, cluster randomized controlled trial, nonrandomized trial, n-of-one trial. See also **observational study**.

Observational study

A study in which the investigator(s) does not control the exposure/ intervention status of study participants (i.e., the assignment of the intervention or exposure of interest is not under the control of the investigator(s)). The simplest form of observational study is the case report or case series, which describes the clinical course of individuals with a particular condition or diagnosis.

Observational studies include descriptive and analytic studies. See also **experimental study**.

Quasi-experimental study

A type of study in which the investigator(s) evaluates the effect of an intervention but does not have full control over the timing, course, or allocation of the intervention. They are often used when it is not possible to conduct a true experimental study.

FROM: Hartling L, Bond K, Harvey K, Santaguida PL, Viswananthan M, Dryden DM. Developing and Testing a Tool for the Classification of Study Designs in Systematic Reviews of Interventions and Exposures. Agency for Healthcare Research and Quality; December 2010. Methods Research Report. AHRQ Publication No. 11-EHC-007. Available at http://effectivehealthcare.ahrq.gov/.