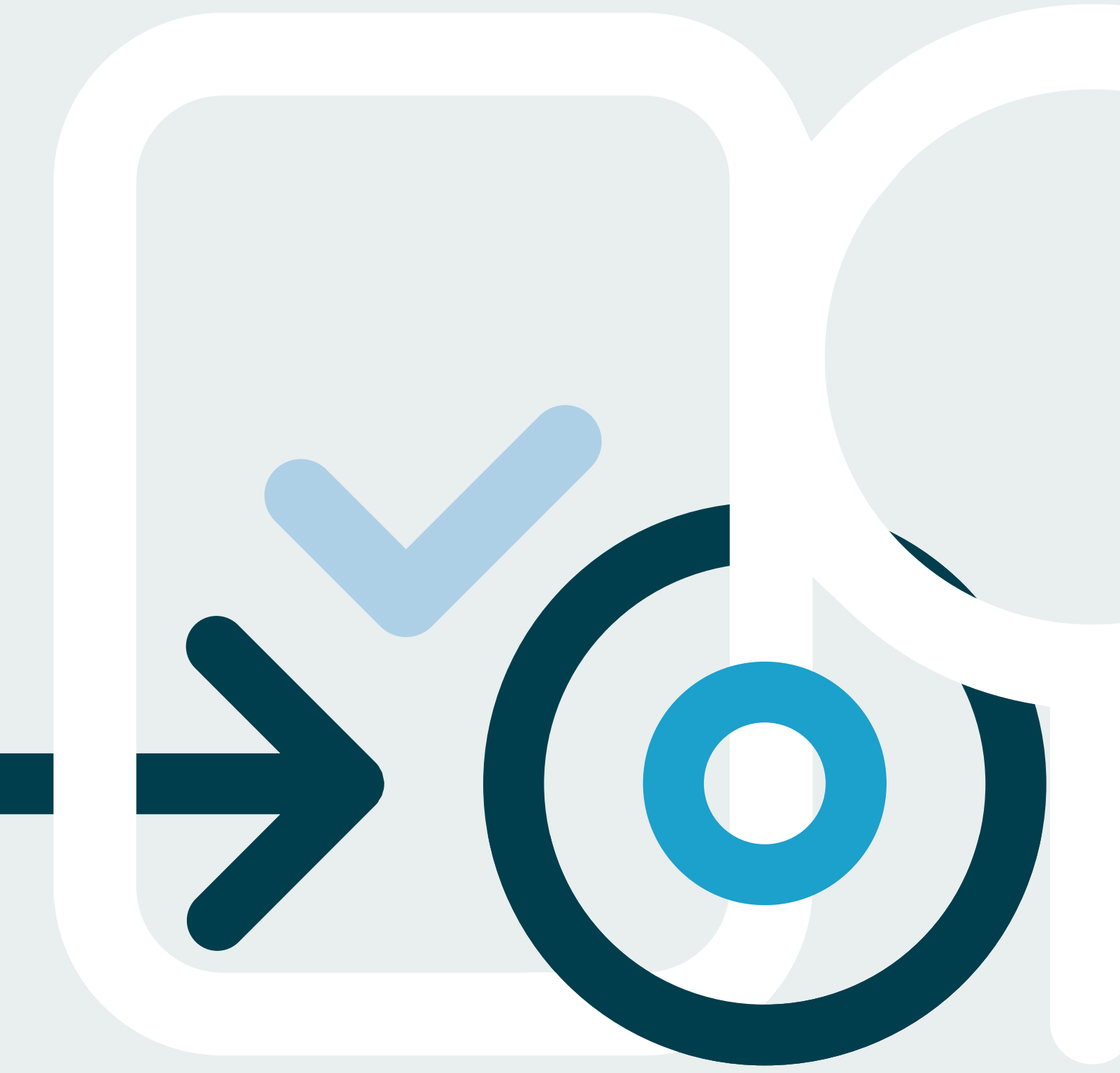
THE ANROWS EVIDENCE PORTAL

ANROWS Instrument for assessing Risk of bias in quantitative Impact Studies (ANROWS-IRIS):

RISK OF BIAS TOOL GUIDANCE DOCUMENT



ANROWS

Prepared by

Angela Higginson, Associate Professor, Queensland University of Technology
Elizabeth Eggins, Research Fellow, Griffith University
Lorelei Hine, Research Manager (Acting), ANROWS
Elizabeth Watt, Research Manager, ANROWS
Charlotte Bell, Senior Research Officer, ANROWS

Author acknowledgement

The authors would like to thank the Evidence Portal Policy Reference Group and Evidence Portal Advisory Group for their comments and input on the development of the ANROWS-IRIS. We would like to also thank Professor David B. Wilson for his invaluable comments and advice. Thanks to the following ANROWS staff: Dr Jane Lloyd for her research expertise and Dr Jennifer Sijnja for her knowledge translation expertise and support. Lastly, we would like to thank the following ANROWS staff members who were integral to the work on the tool: Dr Brigitte Gilbert, Liuissa Zhen and Sara Roetman.

Suggested citation

Higginson, A., Eggins, E., Hine, L., Watt, E., & Bell, C. (2023). *ANROWS Instrument for assessing Risk of bias in quantitative Impact Studies (ANROWS-IRIS): Risk of bias tool guidance document* (Evidence Portal Guidance document, EP.23.03/1). ANROWS.

ANROWS acknowledgement

This material was produced with funding from the Australian Government Department of Social Services. Australia's National Research Organisation for Women's Safety (ANROWS) gratefully acknowledges the financial and other support it has received from the government, without which this work would not have been possible. The findings and views reported in this paper are those of the authors and cannot be attributed to the Australian Government Department of Social Services.

Acknowledgement of Country

ANROWS acknowledges the Traditional Owners of the land across Australia on which we live and work. We pay our respects to Aboriginal and Torres Strait Islander Elders past and present. We value Aboriginal and Torres Strait Islander histories, cultures and knowledge. We are committed to standing and working with First Nations peoples, honouring the truths set out in the Warawarni-gu Guma Statement.

Published by

Australia's National Research Organisation for Women's Safety Limited (ANROWS)
PO Box Q389, Queen Victoria Building, NSW 1230 | www.anrows.org.au
ABN 67 162 349 171

Contents

Introduction

The ANROWS Instrument for assessing Risk of bias in quantitative Impact Studies (ANROWS-IRIS) is a bespoke risk of bias tool developed as part of the ANROWS Evidence Portal. It has been designed for use with the quantitative impact evaluations included in the ANROWS Evidence Portal as well as for systematic reviews in the social and psychological sciences more broadly. The tool is designed to be applied to quantitative impact evaluations of interventions to critically appraise them across six domains that collectively examine whether the design, reporting and implementation of an evaluation study can support the conclusion that the intervention caused a change in the measured outcomes, or if study flaws are likely to lead to over- or underestimates of the effect of the intervention.

This tool has been informed by Effective Public Health Practice Project (EPHPP) Quality Assessment tool, Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) and A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2), for the purposes of rating risk of bias in the estimates of intervention effectiveness from quantitative impact evalua-tion studies.

This document contains the ANROWS-IRIS guidance information for how to apply the tool to studies. It should be read alongside the *ANROWS Instrument for assessing Risk of bias in quantitative Impact Studies: Rating Tool* and the *Development of the ANROWS Instrument for assessing Risk of bias in quantitative Impact Studies (ANROWS-IRIS): Technical report.*

Domain 1: Study design

This domain provides an initial assessment of the study's risk of bias based on the study design. The subsequent domains are used to upgrade or downgrade the study's overall risk of bias.

Q1 Select the study design

a. Randomised controlled trial (RCT)

The key characteristic of an RCT is that researchers **randomly** allocate participants to a treatment or intervention group(s) and a control or comparison group(s) **before** the intervention is conducted. RCTs are therefore prospective experimental designs.

Examples of a random allocation process include:

- computer-generated random numbers
- reference to a random number table
- coin tossing, shuffling cards or envelopes, throwing dice and drawing lots.

If the authors report that the evaluation was a randomised trial or that participants were randomised, the study should be classified as an RCT unless the authors have described a non-random allocation sequence.

If the allocation of participants to groups is not random but might be predictable in advance or externally influenced, then the study is **not** an RCT *even if the author calls it a randomised controlled trial.*

Examples of a non-random allocation process include:

- alternate participants
- methods based on dates (e.g. dates of birth or admission)
- patient record numbers
- allocation decisions made by clinicians or participants
- allocation based on the availability of the intervention
- allocation based on a cut-off score on a pre-intervention measure (i.e. regression discontinuity design).

b. Quasi-experimental impact evaluation with comparison group(s)

The key characteristic of this category of design is that researchers compare outcomes for participants in a treatment or intervention group(s) to outcomes for participants in a control or comparison group(s), but the allocation of participants to these groups is **not random**. Quasi-experiments can be either prospective or retrospective.

There are many different ways to describe quasi-experimental impact evaluations, but broadly the following research designs fall under this category:

Controlled clinical trial or quasi-randomised trial
A prospective experimental design where researchers allocate participants to a treatment or intervention group(s) and a control or comparison group(s) **before** the intervention is conducted, but the allocation method is **not random**.

Examples of a non-random allocation process include:
- alternate participants
- methods based on dates (e.g. dates of birth or admission)
- patient record numbers
- allocation decisions made by clinicians or participants
- allocation based on the availability of the intervention.

Regression discontinuity design
A prospective experimental design where researchers allocate participants to a treatment or intervention group(s) and a control or comparison group(s) **before** the intervention is conducted, and the allocation is based on a cut-off score on a pre-intervention measure.

Cohort analytic with or without baseline measures of the outcome(s)
In this design the researchers do not control whether the participant receives the intervention. Instead, groups are formed retrospectively, based on whether the participants have already received the intervention. Participants are members of the treatment or intervention group(s) if they have received the intervention and are considered members of the comparison group(s) if they have not received the intervention. In this design, all participants must have been measured on outcomes after the intervention but may also have been measured on the outcome measure(s) before the intervention (baseline) and possibly at multiple time points after the intervention. The groups may or may not also be statistically matched on key variables.

Case-control design
Case-control studies are typically conducted to examine rare outcomes. In this design the researchers do not control whether the participant receives the intervention. Instead, groups are formed retrospectively based on whether the participants have **already displayed the outcome of interest.** The groups are then examined to determine if they differ based on their prior exposure to the intervention.

Multiple regression analyses
These are analytic techniques that can be applied to cross-sectional or longitudinal data to control for the potential impact of other key variables in the analysis. In multiple regression analysis the intervention is treated as one of a set of predictor or independent variables and the outcome of interest is treated as the dependent variable in the model. Multiple regression models attempt to statistically control for the influence of potential confounders by controlling for the effect of multiple predictor or independent variables.

Bivariate correlational design
In this design the researchers typically do not control whether the participant receives the intervention. The design uses cross-sectional data to calculate the bivariate correlation between the level of the intervention (e.g. intervention presence or absence, or intervention dose) and the level of the outcome among

participants. Although this design compares outcomes in participants who received the intervention (or who received more of the intervention) to outcomes in participants who did not receive the intervention (or who received less of the intervention), a bivariate correlational design does not attempt to control for alternate influences on the outcome. A conceptual equivalent is the analysis of cross-sectional data using a bivariate or simple regression model with only the intervention variable as an explanatory variable to predict the presence, absence or level of the outcome of interest.

Interrupted time-series design with comparison group(s)
In this design an aggregate measure of an outcome is **observed over multiple time points** (e.g., daily, weekly, monthly) before and after the intervention. The outcome is measured for **both** an intervention group and a comparison group. (Note that synthetic controls are considered a comparison group). There are several ways that an interrupted time-series design with a comparison group can be analysed, but a key characteristic is that the comparison group controls for the impact of alternative influences on the outcome over time.

c. Long interrupted time-series design without comparison group

In this design an aggregate measure of an outcome is **observed over multiple time points** (e.g. daily, weekly, monthly) for a single group that receives the intervention. A long interrupted time-series has 25 or more observations before the intervention and 25 or more observations after the intervention. There are several ways that a long interrupted time-series design can be analysed, but a key characteristic is that the pre-treatment observations function as the comparison group for the post-treatment observations.

If the study design has these characteristics and has **fewer than** 25 pre-intervention and post-intervention observations, it is a short interrupted time-series design and should be categorised as a single group pre–post design. This is because the data series is not considered long enough to appropriately control for the impact of alternate influences on the outcome.

d. Single group pre–post design

In this design there is only one group and all members of that group receive the intervention. There is no comparison group. Participants are measured on the outcome before and after the intervention. These designs can be prospective or retrospective.

This category also includes short interrupted time-series designs (fewer than 25 pre-intervention and fewer than 25 post-intervention observations) without comparison groups.

Q2 Is the comparison condition or group comprised of treatment refusers or drop-outs?

Note: This item is not scored for long interrupted time-series without comparison group(s) or single group pre–post designs.

a. **Yes**

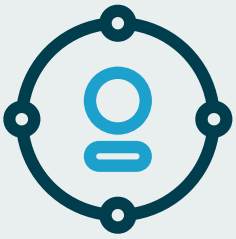
   The comparison group is made up of participants who refused or declined to participate in the intervention, or who failed to complete the intervention.

   If participants were reallocated from the intervention group to the comparison group because they dropped out of the intervention, select this option.

b. **No**

   The comparison group is no treatment, a waitlist control, placebo, treatment or business as usual, an alternative treatment, or an alternative dosage of the treatment.

Domain 2: Selection bias



This domain focuses on the **external validity** of the study. It assesses whether the methods used to select participants may affect the generalisability of the study results to the intervention's target population. This assessment is an interplay between the aims of the intervention, the possible population of eligible participants and the sampling approach of the study.

Note: The questions in this domain focus on participants' initial selection into the study, and *not* on participants' allocation to intervention or comparison group(s).

Q3 Do the authors clearly describe the target population?

The target population is the people or places for which the intervention is intended. The target population may not be explicitly stated in the study using this terminology, so critical judgement may be needed to identify the target population. The target population is the set of people or places to which the results of the study can be generalised.

It is important to distinguish between the sample that was analysed in the study, and the target population. For example, a study may be conducted with perpetrators in one specific city, but this does not mean that the perpetrators living in that city are the target population of the intervention – the target population is typically much broader than the sample in the study.

a. Yes

The authors might state that the intervention is designed for use in particular circumstances, or for particular populations (e.g. an intimate partner violence [IPV] primary prevention program for use in universities, a perpetrator program to be used as a diversion from arrest, a nurse-practitioner model of care for victims and survivors in refuges, a violence against women [VAW] bystander program in men's rugby league clubs).

In these circumstances, there is likely to be sufficient information to know who the intervention is aimed at in the broader population (e.g. students in higher education, perpetrators who have come to the attention of police, victim and survivor residents in women's shelters, male rugby league players).

b. No

The study authors name or describe the intervention but do not identify the particular populations, circumstances, or locations where the intervention is intended to be applied (e.g. an IPV primary prevention program, an IPV perpetrator program, a VAW bystander program).

In these cases, the target population should be assumed to be equally broad (e.g. people who have not perpetrated IPV, perpetrators of IPV or bystanders to VAW).

Q4 Do the authors clearly describe the sampling frame?

The sampling frame is the source that the study uses to access potential participants. Ideally the sampling frame will be appropriate for the target population, as even with random sampling, bias can be introduced to the study if the sampling frame does not match the target population.

Sampling frames can be broad or narrow, and can be made up of individuals, organisations or spaces, and they may facilitate direct or indirect recruitment. Examples of sampling frames include:

- a list of individual people and their contact details (e.g. the national electoral roll; the electoral roll in one particular city, a paid survey panel, the phone book, specific subreddits, the entirety of Twitter)
- a set of organisations that could facilitate the recruitment of individual participants to the study (e.g. a list of all women's refuges in the state, one women's refuge, the psychology department at one university, all schools in one school district, one police station, all police stations in one police district)
- a list of areas where recruitment of participants might occur (e.g. a large shopping centre or high street, three university campuses, high-crime street segments in one city)
- a pre-existing data set of aggregate data for a location (e.g. DFV arrest data for a city, aggregate data on maxillofacial injuries from one emergency department)
- an organisation that has implemented a new policy or practice either in a pilot location or across the entire organisation (e.g. one police district, one policing jurisdiction).

a. Yes

The authors clearly describe the broad parameters of the sampling frame. It is not necessary for authors to identify the exact location of the sampling frame (e.g. authors do not need to name a particular police department or school district), but there must be sufficient information to evaluate the conceptual appropriateness of the sampling frame. Typically, a clearly described sampling frame will give information on the type and number of source locations included, as in the list above.

b. No

The study authors do not describe the broad parameters of the sampling frame well enough to allow an evaluation of its conceptual appropriateness, or they give no indication of where the sample has been drawn from.

Q5 Is the sampling frame likely to be appropriate for the target population?

Critical judgement and subject matter knowledge are needed to assess the appropriateness of the sample frame. Think about which section(s) of the target population is being represented, and who might be underrepresented. Sampling frames may be more likely to be appropriate for the target population when the target population is clearly defined, but only if there is no other obvious source of bias.

a. Yes

The sampling frame appears to be generally appropriate for the target population and does not appear to present an obvious source of bias.

For example, if the target population is male rugby league players, a sampling frame of three rugby league clubs in a city would be likely to be appropriate, as the sampling frame matches the target population and there is no obvious source of bias.

b. Somewhat

The sampling frame presents an obvious source of bias, but the sampling frame still represents a reasonable subset of the target population.

For example, if the target population is IPV victims and survivors, a sampling frame of a women's shelter is somewhat likely to be appropriate, as it will overrepresent women who have left violent relationships and underrepresent those who remain in the relationship. The sampling frame is a subset of the target population.

For another example, if the target population has not been clearly described, but can be assumed to be IPV perpetrators, a sampling frame of IPV arrest data from a police jurisdiction is somewhat likely to be appropriate. The sampling frame would not allow the study access to those IPV perpetrators who had not been arrested, and there are likely to be systematic differences between arrested and non-arrested perpetrators, but "arrested perpetrators" is a reasonable subset of the target population.

c. No

The sampling frame presents an obvious source of bias and the sampling frame is not conceptually appropriate for the target population.

For example, if the target population is VAW perpetrators in a custodial setting, but the sampling frame is made up of first-year students from the psychology department at one university, the sampling frame is not likely to be appropriate to the target population. Although some first-year psychology students may have perpetrated VAW, they are also likely to have considerably different demographic characteristics to imprisoned VAW perpetrators, not least of which are age and socio-economic status. The more important issue is that the custodial setting is very different to a psychology lab. This choice of sampling frame is likely to introduce significant bias to the results.

d. Cannot tell

Only select this option if Q3 = b and the authors do not clearly describe the target population well enough to assess the appropriateness of the sampling frame.

Q6 Do the authors clearly describe the sampling approach?

The sampling method is the approach used to recruit participants to the study from the sampling frame.

Note: Although there can be overlaps of the methods used, the sampling method is *not* the same as the allocation process of assigning participants to the intervention or comparison group(s).

Sampling methods include:

- simple random sampling, stratified random sampling, cluster random sampling
- systematic sampling (i.e. selecting every nth member of the sampling frame)
- clinician/practitioner referral, participant self-referral
- snowball sampling
- purposive sampling, quota sampling
- convenience sampling, haphazard sampling.

In some cases the sample may be the total population of the sampling frame, particularly where an organisation has chosen to implement and evaluate a new practice, or where aggregate administrative data is being analysed.

a. **Yes**

   The authors clearly describe the general approach to sampling.

b. **No**

   The study authors do not describe the general approach to sampling well enough to allow an evaluation of its conceptual appropriateness, or they give no indication of how the sample was selected.

Q7 Are the study participants likely to be representative of the sampling frame?

Sampling can use either probability or non-probability methods. Probability sampling uses methods that result in a known probability of each member of the sampling frame being selected for recruitment. In non-probability sampling the probability of each member of the sampling frame being selected for recruitment cannot be calculated.

Probability sampling methods include:

- simple random sampling, stratified random sampling, cluster random sampling
- systematic sampling (i.e. selecting every nth member of the sampling frame).

Although not technically a probability sampling method, total population sampling also has a known probability of each member of the sampling frame being selected for recruitment, with the probability being 1. Total population sampling therefore guarantees that the sample is representative of the sampling frame.

Non-probability sampling methods include:

- clinician/practitioner referral, participant self-referral
- snowball sampling
- purposive sampling, quota sampling
- convenience sampling, haphazard sampling.

Statistical methods can be used to demonstrate that the sample is representative, or to control for known reasons for selection bias. These methods include comparison of the study sample to the sampling frame to demonstrate equivalence, statistical comparison of the sample versus those who declined to participate to demonstrate equivalence, and statistically controlling for potential confounders related to selection into the sample.

a. Yes

The authors provide detail to demonstrate that the sample is likely to be representative of the **sampling frame**. This may be established in the following ways:

i. Total population sampling is used, meaning that the entire sampling frame can be utilised (e.g. place-based experiments, criminal justice interventions that are implemented as a pilot study or new policy and are applied to all cases that fulfil specific criteria).

ii. Statistical comparison is used that demonstrates the equivalence of the study sample and the sampling frame (e.g. based on the statistical comparison of key socio-demographic variables).

iii. Statistical comparison is used that demonstrates the equivalence of those who agree to participate with those who decline to participate (e.g. based on the statistical comparison of key sociodemographic variables).

iv. The authors statistically control for potential confounders that are argued to be related to selection into the sample, and that would result in the sample being unrepresentative of the sampling frame if not controlled.

In cases where the study has statistically compared the sample to the sampling frame across multiple variables, and some (but not all) variables are statistically significantly different, use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the sample is not representative.

b. Somewhat

The authors have not statistically compared the sample to the sampling frame, but there is no clear indication of selection bias.

c. No

The authors provide detail to demonstrate that the sample is NOT representative of the sampling frame. Examples include samples that statistically differ from the sampling frame on key demographic variables, or participants who have self-selected into the study and no statistics are provided to demonstrate representativeness.

In cases where the study has statistically compared the sample to the sampling frame across multiple variables, and some (but not all) variables are statistically significantly different, use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the sample is not representative.

d. Cannot tell

Only select this option if Q4 = b **OR** Q6=b and the authors do not clearly describe the sampling frame **OR** the sampling approach well enough to evaluate the likelihood of the sample being representative.

Q8 Do the authors demonstrate that the participants are likely to be representative of the target population?

Note: This item focuses on participants' initial selection into the study, and *not* on participants' allocation to intervention or comparison group(s). It is concerned with the equivalence between the study sample and the target population, and *not* with baseline equivalence between the intervention and comparison group(s).

a. **Yes**

The authors provide detail to demonstrate that the sample is likely to be representative of the target population. This may be established in the following ways:

i. Total population sampling (census) is used, meaning that the entire target population can be utilised (e.g. place-based experiments, criminal justice interventions that are implemented across a jurisdiction as a pilot study or a natural experiment, or a new policy that is applied to all cases that fulfil specific criteria).

ii. Statistical comparison is used that demonstrates the equivalence of the study sample and the target population (e.g. based on the statistical comparison of key socio-demographic variables).

iii. Statistical comparison is used that demonstrates the equivalence of those who agree to participate with those who decline to participate (e.g. based on the statistical comparison of key socio-demographic variables).

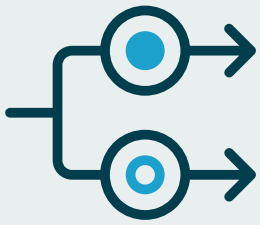
iv. The authors statistically control for potential confounders that are argued to be related to selection into the sample, and that would result in the sample being unrepresentative of the target population if not controlled.

In cases where the study has statistically compared the sample to the target population across multiple variables, and some (but not all) variables are statistically significantly different, use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the sample is not representative.

b. No

The study did not use any of the methods described above to demonstrate that the sample was likely to be representative of the **target population**.

Domain 3: Confounders



Note: This domain is not scored for single group pre–post designs.

The questions in this domain focus on participants' allocation to intervention or comparison group(s) and not to their initial selection into the study.

Q9 Do the authors state or demonstrate if the comparison group was equivalent to the treatment group prior to the intervention?

a. **Yes**

The authors discuss whether the groups are equivalent or balanced at baseline before the intervention, or they provide data or statistical analysis to evaluate group equivalence across key demographic variables.

In studies without a baseline measure of the outcome(s), authors can evaluate equivalence based on socio-demographic variables. For long interrupted time-series without comparison group(s), authors can evaluate the stationarity of the pre-intervention data series.

b. **No (score this domain and go to Q12)**

The authors have not discussed group equivalence and have not provided any data to evaluate group equivalence, or stationarity.

Q10 Are there any meaningful differences between the groups?

A meaningful difference is one that could reasonably be expected to bias the results of the analysis.

a. **Yes**

The authors state that the groups are not equivalent or not balanced at baseline, or statistically significant differences are reported on key variables.

If only raw data is presented without interpretation, answer "Yes" if there appear to be meaningful differences between groups on key variables. Use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the groups differ.

For long interrupted time-series without comparison group(s), authors have demonstrated that the pre-intervention data series is not stationary and any trends or seasonality in the pre-intervention data series has not been controlled for in the analysis.

In cases where the study has statistically compared the groups across multiple variables, and some (but not all) variables are statistically significantly different, use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the groups differ in a meaningful way.

b. No (score this domain and go to Q12)

The authors state that the groups are equivalent or balanced at baseline, or there are no statistically significant differences reported on key variables. If the authors state that the groups are equivalent or balanced at baseline, but the data demonstrates a meaningful difference, answer "Yes" to this item.

If only raw data is presented without interpretation, answer "No" if there do not appear to be meaningful differences between groups on key variables.

For long interrupted time-series without comparison group(s), authors have demonstrated that the pre-intervention data series is stationary or any trends or seasonality in the pre-intervention data series has been controlled for in the analysis.

In cases where the study has statistically compared the groups across multiple variables, and some (but not all) variables are statistically significantly different, use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate that the groups differ in a meaningful way.

Q11 Do authors attempt to control for confounding factors in their analysis?

Confounding factors are those that influence both the likelihood of receiving the intervention and the participants' score on the outcome measure.

Examples of possible confounders include age, ethnicity, gender, sexuality, education, marital status, family structure, socio-economic status (e.g. income or class), health status, prior contact with the criminal justice system, prior victimisation, pre-intervention score on outcome measure(s), and/or seasonality or autocorrelation in time-series designs.

This is not an exhaustive list, as confounders tend to be contextually dependent.

a. Yes

The study design includes statistical controls, such as including covariates in regression analysis or statistically matching the intervention and control group on confounding factors, or statistically controlling for seasonality or autocorrelation in time-series designs.

b. No

The analysis does not incorporate statistical controls for confounding factors.

Domain 4:
Data collection methods

Q12 Do the outcomes have face validity?

Face validity is the extent to which a test appears (at face value) to measure what it purports to measure. Face validity is a subjective judgement.

a. **Yes**
The primary outcomes appear (at face value) to be appropriate measures of the concepts that the study aims to evaluate.

b. **No**
The primary outcomes do not (at face value) appear to be appropriate measures of the concepts that the study aims to evaluate.

Q13 Do the authors describe how they measured each outcome?

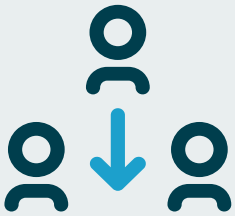
a. **Yes**
All outcomes are clearly named and described in replicable detail or the authors provide a citation to an existing standardised measure.
Replicable detail is the level of detail that would allow a different researcher to perform the same analysis. This level of detail may sometimes appear in text and sometimes in appendices or supplementary materials. Examples include text of survey questions and response categories, detail of how composite indexes were created including component-item text and response categories, and names of data fields or variables from administrative datasets. Alternatively, authors may report a citation to a standardised or validated item, or to a paper that provides replicable detail for the outcomes in question. Answer "Yes" if **all** outcomes are reported in this level of detail.

b. **Somewhat**
The study provides a brief description of the outcomes only and no citation to an existing measure.
The outcome measures are described well enough to evaluate face validity, but not well enough to allow replication.

c. **Mixed**
The primary outcomes meet the standard for a) but at least some secondary outcomes only meet the standard for b).
In cases where the study reports on multiple outcomes, use critical judgement and subject matter expertise to assess which outcomes should be considered primary and which outcomes should be considered secondary.

d. **No**
Only outcome labels are provided.

Domain 5: Withdrawals and drop-outs



Note: This domain is not scored for single group pre–post designs.

This domain assesses the level of attrition from the study. Attrition may occur while the intervention is ongoing, or it may occur after the intervention has been completed but before all outcome measurements are taken. When answering Q14–16, consider all possible sources of attrition that are reported in the study.

Q14 Is there a meaningful difference in attrition or drop-out between the treatment and comparison group?

A meaningful difference is one that could reasonably be expected to bias the results of the analysis.

a. **Yes**

   The authors state that the groups experience different levels of attrition, or statistically significant differences are reported on attrition between groups.

   If only raw data is presented without interpretation, answer "Yes" if there appear to be meaningful differences in attrition between groups. Use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate differential attrition.

b. **No (score this domain and go to Q17)**

   The authors state that the groups experience equivalent levels of attrition, or no statistically significant differences are reported on attrition between groups.

   If only raw data is presented without interpretation, answer "No" if there does not appear to be meaningful differences in attrition between groups. Use critical judgement and subject matter expertise to assess whether the differences are sufficient to demonstrate differential attrition.

c. **There is no attrition in either group (score this domain and go to Q17)**

   Answer c) if there is no attrition in either the treatment or the intervention group(s). This is likely to occur when using administrative data or aggregate data.

d. **Unclear (score this domain and go to Q17)**

   The authors do not provide sufficient information to assess whether there has been differential attrition.

Q15 Is the attrition systematic or at random?

a. **Systematic**

The authors state or statistically demonstrate that at least some measures of attrition are associated with one or more confounding factors.

b. **Random (score this domain and go to Q17)**

The authors state or statistically demonstrate that none of the attrition is associated with confounding factors.

If the authors state that the attrition is at random, but the data demonstrates a meaningful pattern to the attrition, answer a) to this item.

c. **Unclear (score this domain and go to Q17)**

The authors do not provide sufficient information to assess the source of differential attrition.

Q16 If systematic, did the authors control for the impact of differential attrition?

a. **Yes**

The study design includes statistical controls, such as including covariates in regression analysis or statistically matching the intervention and control group on factors associated with differential attrition.

b. **No**

The study design does not include statistical controls for factors associated with differential attrition.

c. **Unclear**

The authors do not provide sufficient information to assess whether the study design includes statistical controls for factors associated with differential attrition.

Domain 6: Intervention integrity and fidelity

Q17 Was the intervention implemented as intended (as per protocol)?

Intervention integrity refers to the delivery of the intervention as it was planned or intended. This domain focuses on whether there were changes to the way in which the intervention was delivered. Changes may be either planned or unplanned.

Please note: Issues with participant withdrawal are only considered in this domain if they result in a change to the way in which the intervention is delivered to the remaining participants.

- If not all of the participants in a treatment group received the full intervention (full dosage) because of withdrawal or drop-out, but the intervention was still delivered to the remaining participants as intended, do not report this as an issue with intervention integrity. Although some participants received a lower dose than intended, the way that the intervention was implemented did not change. This issue should instead be considered when rating Domain 5.
- If the intervention that was delivered to the remaining participants was adapted or changed because of withdrawals or drop-outs, this is a change to how the intervention was implemented. This may indicate an issue with intervention integrity and should be considered here.

a. **Yes**

   The authors state that the intervention was delivered as intended **OR** the authors do not report anything to suggest that the intervention was not implemented as intended.

b. **Somewhat**

   There were variations but the authors described the variations clearly **OR** the authors discuss where the intervention varied from its intended implementation and provide a clear description of the intervention as received by participants.

c. **No**

   The authors report that the intervention varied from its intended implementation, but do not provide a clear description of the intervention as received by participants.

Q18 Did the authors report that co-intervention or contamination occurred?

Co-intervention is where participants in the intervention group receive an additional intervention beyond what was intended. This may range from an additional intervention component through to an entirely separate additional intervention.

Contamination is when the comparison group receives some or all of the intervention.

a. **Yes**

Either co-intervention or contamination (or both) is reported to have occurred.

The authors may discuss that co-intervention or contamination may have occurred, but unless it is confirmed by the authors to have *actually* occurred, select "No".

b. **No**

Neither co-intervention nor contamination were reported to have occurred.

The authors may discuss that co-intervention or contamination *may* have occurred, but unless it is confirmed by the authors to have *actually* occurred, select "No".

Q19 If contamination or co-intervention was reported, did the authors report the results of relevant sensitivity analyses?

a. **Yes**

To determine the impact of contamination or co-intervention on the study, the authors may have conducted both an as-treated analysis and an intention-to-treat analysis, and then conducted a sensitivity analysis of the results.

b. **No**

No sensitivity analyses of the impact of co-intervention or contamination were reported.

ANROWS

AUSTRALIA'S NATIONAL RESEARCH
ORGANISATION FOR WOMEN'S SAFETY
to Reduce Violence against Women & their Children