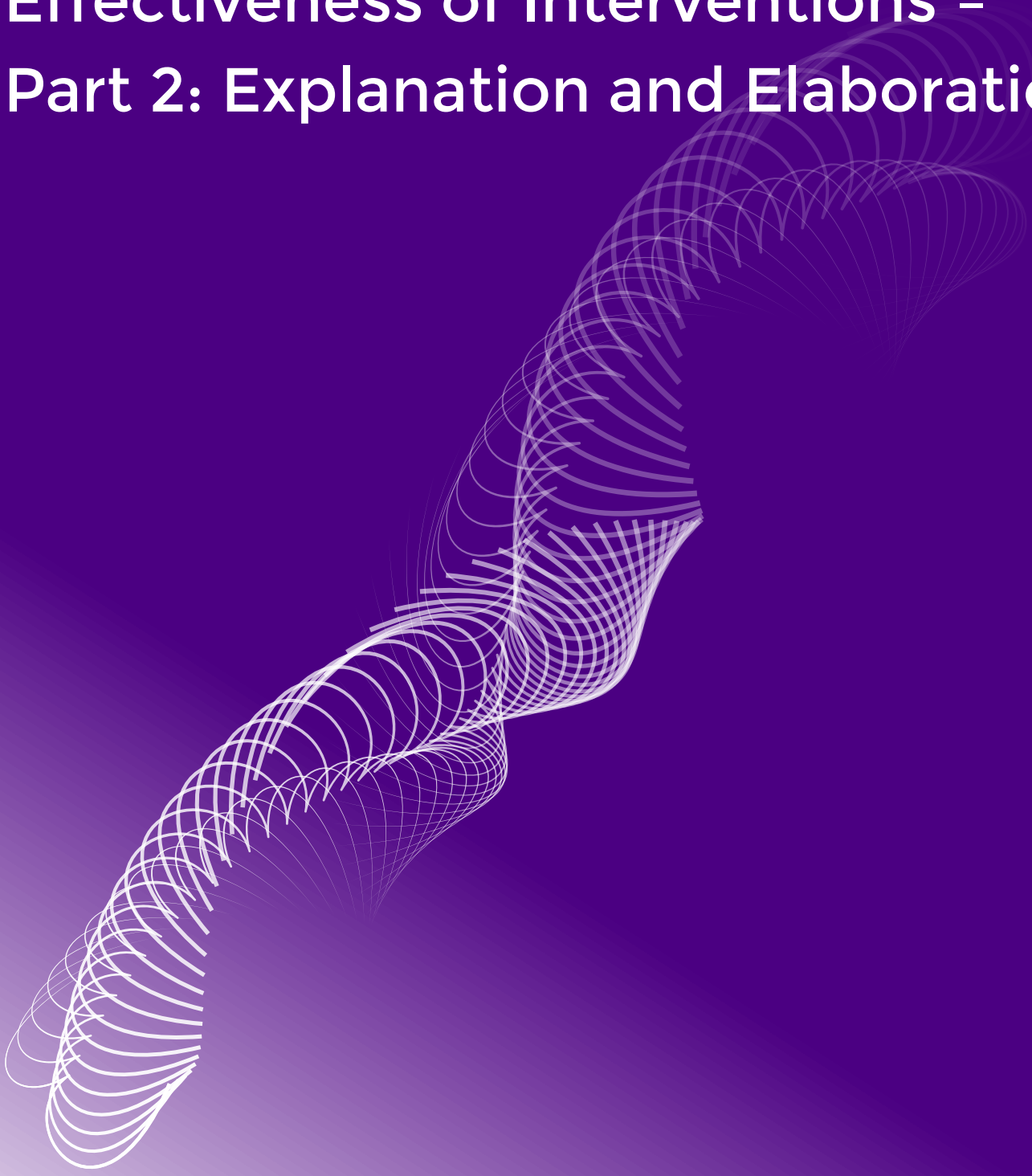
Standards of Evidence

Standard for Producing Evidence – Effectiveness of Interventions – Part 2: Explanation and Elaboration



Standard of Evidence 2 Part 2 (StEv 2-2:2016)

Standard for Producing Evidence – Effectiveness of Interventions – Part 2: Explanation and Elaboration

Standard of Evidence 2 Part 2 (StEv 2-2:2016)

Supported by:

Publishing and copyright information

Standard for Producing Evidence – Effectiveness of Interventions – Part 2: Explanation and Elaboration

Standard of Evidence 2 Part 2 (StEv 2-2:2016)

Published February 2016

ISBN 978-1-911056-02-7

This standard may be indexed under the following ICS classification:

03.100.40 (Research and development)

www.hact.org.uk

@HACThousing

info@hact.org.uk

HACT is registered as the Housing Associations' Charitable Trust, charity number 1096829, company number 04560091.

This document may be cited as:

"Vine, Jim (2016). Standard for Producing Evidence – Effectiveness of Interventions – Part 2: Explanation and Elaboration. HACT. London, UK."

Publication history

First published February 2016

Amendments issued since publication

Date	Text affected

Technical foreword

Information about this document

This standard is published by HACT and drafted by Jim Vine. It came into effect on 2 February 2016.

National and international standards bodies have established conventions for the drafting of standards. Whilst an Explanation and Elaboration document is not a common format in other standards, where appropriate these conventions have been adopted in the drafting of this standard. For the avoidance of doubt, the adoption of these conventions does not constitute a claim that any such body has overseen the creation of this standard.

> *[REFERENCES: ISO/IEC Directives, Part 2, 2011. http://www.iec.ch/members_experts/refdocs/iec/isoiec-dir2%7Bed6.0%7Den.pdf. Rules for the structure and drafting of UK standards, 2012. http://www.bsigroup.com/Documents/standards/guide-to-standards/BSI-Guide-to-standards-2-standard-structure-UK-EN.pdf]*

This part of the standard (part 2) is primarily informative in nature. As such, it provides supporting information explaining the 'normative' part (part 1), which sets out the provisions and requirements of the standard.

Where possible, the sections and numbers of this part of the standard match their respective sections in part 1.

Acknowledgements

The work to develop this standard would not have been possible without the kind support of Bromford, Look Ahead Care and Support, Metropolitan, Sanctuary Supported Living, Trafford Housing Trust and Public Health England.

Great thanks are also due to all of those who have acted as a Correspondence Group for the project of developing the standard, contributing their thoughts in ways that have very much improved the final

output, and to Peter Molyneux, who chaired meetings of the working group that led to the standard. Any errors that remain are, of course, the responsibility of the author.

The Correspondence Group included representatives of the following organisations: Academy of Medical Royal Colleges; Alliance for Useful Evidence; Care and Repair England; Centre for Mental Health; Chartered Institute of Housing; Children and Young People's Mental Health Coalition; Department for Communities and Local Government (DCLG); EDF; Economic and Social Research Council (ESRC); Homeless Link; Housing LIN; Joseph Rowntree Foundation (JRF); London School of Hygiene and Tropical Medicine; MDRC; National Housing Federation; National Institute for Health and Care Excellence (NICE); Place2Be; Royal College of Psychiatrists; University College London; University of Chicago; University of Durham; University of Glasgow; University of Stirling; University of Warwick; University of York; and Youth Access.

Contractual and legal considerations

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a standard cannot confer immunity from legal obligations.

Contents

1 Scope

1.1 Scope of this document

This document provides explanation and elaboration of the requirements and recommendations specified in StEv2-1 (Standard for Producing Evidence – Effectiveness of Interventions – Part 1: Specification). *[REFERENCE: Vine, 2016a.]* It provides information on why certain features are included in the process specified by StEv2-1.

This part of the standard (part 2) only provides explanations of the rationale behind the elements of the process and does not attempt to repeat in detail all of the elements of that process. The detailed specification of the process is provided separately, in part 1 of the standard.

1.2 Explanation and elaboration on the scope of StEv2-1

The scope of the process specified in StEv2-1 is intended to be equally applicable for any type of intervention, because there are many arears of activity where it is possible to produce evidence of the effectiveness of the intervention, and where such evidence could beneficially be used to inform decisions. Similarly, the process is applicable whether the intervention being studied is existing or new, since many existing practices will not be supported by robust evidence of their effectiveness (and may be ineffective).

Consequently, the process has been designed to make as few assumptions as possible about the nature of the intervention, outcomes or context in which the study is being conducted. It is hoped that this broad and general applicability will facilitate wide adoption of the standard. If different parts of organisations produce evidence of the effectiveness of interventions in the same way, hopefully the organisation's confidence in producing and using robust evidence of effectiveness with grow in general.

StEv2-1 was developed in the context of the housing sector, and specifically with input from a group of housing associations that were interested in increasing their engagement with the health sector. The housing sector has

historically had limited evidence of the effectiveness of its interventions, which is a particular constraint on engagement with sectors where evidence is routinely available to inform decisions. In specifying a process that the housing sector can use across its range of activities, there is more potential to build familiarity with evidence than there would be with a narrowly focused process, suited only to deployment in particular divisions of some organisations. The resulting, relatively generic, standard is still highly relevant to measuring health and wellbeing outcomes, and assessing the effectiveness of interventions in delivering outcomes that are wider determinants of health, but is also applicable to any other activity where an organisation can specify and measure an outcome that an intervention is intended to achieve. In becoming applicable to the breadth of activity of the housing sector, it is also likely to be relevant to many organisations outside of the sector.

In specifying a process for producing evidence, StEv2-1 also aims to increase the confidence that evidence users have in the evidence produced, in order to increase the chances that they will make use of the evidence. This document also plays a role in building this confidence, by explaining how the various elements of the process contribute to robust evidence creation.

The ultimate intention is that the adoption of the process should result in the increased use of robust evidence of the effectiveness of interventions to inform decision making. If organisations are able to select between different interventions based upon evidence of which is likely to be most effective, they will tend to achieve more of their intended outcomes, and to use their resources to greatest effect. Adoption of the process would support the increased use and re-use of evidence of effectiveness by:

- increasing the supply of evidence of effectiveness;
- building confidence in the evidence that is created; and
- ensuring that the evidence is as widely accessible as possible.

Whilst the focus of StEv2-1 is on creating evidence of the effectiveness of interventions, this is not intended as an indication that ‘what works’ is the only question that matters, nor the only thing that evidence can be generated on. StEv2-1 also specifies high-level approaches (rather than detailed processes) to guide economic evaluation and process evaluation, which will help to answer several closely related questions (such as ‘what works for whom, in what circumstances, at what costs, how and why?’).

StEv2-1 also provides only a high-level approach in relation to conducting systematic reviews. Although these are typically concerned with producing evidence of effectiveness, detailed processes are out of scope because they do so through a significantly different method than the primary studies that are the core focus of the scope.

The scope of StEv2-1 notes that as well as being used by those producing evidence, the standard can also be used in commissioning studies. The existence of the standard should simplify the process of commissioning as commissioners may specify to contractors that they require evidence to be generated in conformity with the standard (and at a specified level, as appropriate).

2 Terms and definitions

2.1 Terms and definitions used in this document

For the purposes of this document, the terms and definitions given in StEv1-1 (General Requirements for Evidence – Part 1: Vocabulary) apply. *[REFERENCE: Vine, 2016.]*

2.2 Explanation and elaboration on the terms and definitions adopted

The terms adopted have been selected with an intention that they should be relatively widely known and unlikely to contribute to misunderstandings. As far as possible, a single term has been selected for each concept and used throughout; the use of synonyms has been avoided. This results in a slightly more repetitious style of writing, but removes the potential for readers to be left with questions over whether terms are being used as synonyms or to convey slightly different meanings: "when it says 'study' there does it mean the same thing as when it says 'research' over here?".

Some of the terms were selected in the interests of generality and neutrality of interpretation. "Intervention" was chosen rather than "treatment" (which might imply a particular medical context) or any terms that might limit the applicability (such as "project" or "service"). "Study" was selected to describe projects to produce evidence, rather than "research" or "evaluation" as in some organisations those terms may be associated with particular teams or budget headings, and their use may have risked pigeonholing the use of the standard into either of those categories; the term also avoids the connotations in some circles for "evaluation" being defined as occurring *ex post*, i.e., during or after implementation, with "appraisal" being used for studies conducted *ex ante* (i.e., prior to implementation).

The term "outcome" is used to refer to all variables that it is believed (or hoped or feared) an intervention might have an effect on, where that effect would be of some interest. This is intended to include all effects, without

distinguishing between those that happen soon after the intervention and those that can only be detected some time later, and encompasses both intermediate outcomes and the ultimate outcomes of interest. In some other contexts, the ultimate outcomes of interest, or effects that take longer to become apparent, are sometimes separated out as “impacts”. Similarly, in the academic discipline of evaluation science, what is referred to here as “evidence of effectiveness” might be referred to as “outcomes evaluation” or “impact evaluation” depending on the proximity of the effect.

3 Process

The process flowchart is provided as a visual representation of the process specified in the standard.

Two points of iteration are specified in the process. Those iteration loops are specified to ensure that each of the relevant stages are reviewed in light of other information that is generated in other parts of the process.

After an intervention has been specified it is important to revisit the literature to check whether there is evidence specifically related to the intervention as it has been designed. It is also helpful to revisit the causal chain map to see whether the expected causal links are still likely to pertain given the specific design.

A further set of iterated stages is specified within the study planning part of the process. The study design, participant recruitment approach, ethical requirements, outcome measures and study team are interrelated. Consequently it is important to review them in light of each other.

For simplicity, the process flowchart illustrates the process ending once the evidence has been produced and put into use. In some instances, having produced one piece of evidence the process may start again. If a level 1 study identifies that an intervention is practicable and associated with an outcome of interest it may be appropriate to commence a level 2 study into the intervention. If a level 2 study finds an intervention is ineffective it may be necessary to commence the process again to attempt to specify another intervention that is more effective.

4 Issue description

Preparation of an issue description helps to ensure that it is clear what questions the evidence is intended to answer and why they are important. This stage specifies a requirement for there to be a conscious process to describe the issue that the intervention is intended to address. Without actively undertaking this work, it would be easy for the nature of the issue to remain an implicit and unstated assumption. The issue might be a specific problem or negative situation to be resolved or moderated, or it might be an improvement that it is hoped can be achieved.

Part of this process is also about checking that the identified issue is important; for problems this means establishing that the situation is actually bad. Sometimes the problem will be so clear and the intended outcomes so obvious that this step will just require a few sentences to document them; in other cases, the process will help to uncover assumptions that may not be shared by everyone involved in the intervention or study, or that need closer examination before they should be acted upon.

Whilst many issues will be self-evidently problematic, in some cases it will not be so obvious whether an issue is actually perceived or experienced as a problem by those it affects. In these cases it would be worth undertaking a small piece of research with the affected population to check your intuition. This will typically require a piece of qualitative research, perhaps using interviews or focus groups.

The same principals apply where the study is examining prospective improvements: sometimes the benefits will be immediately apparent; in other cases they should be examined to test that they would be experienced by all as improvements.

Another potential role for qualitative research in problem identification might arise where you have a general idea about the existence of an issue but need to investigate it in more depth to fully understand exactly what the aspects of it are that you want to address or to inform the creation of potential interventions. In research terminology the process of identifying an intervention that you intend to study for effectiveness can be expressed as a hypothesis: 'if we implement intervention A we will see improvements in outcome X'. Qualitative research is well suited to hypothesis generation *[SOURCE: Lewin et al., 2009.]*.

This step is important whether you are looking to create a new intervention or test an existing one. Even activities that organisations have carried out for many years can be susceptible to being conducted without a clear shared understanding of what they are intended to achieve; in fact, it may be the case that existing activities have more unstated assumptions sitting behind them, and clarifying exactly what they are intended to achieve and for whom can be particularly valuable.

At this stage the description of the issue should capture the breadth of experience of the issue. The intervention that is adopted for the study may subsequently be designed to address the issue specifically for a subset of those experiencing it. The target population for the study is recorded later in the process (see section 7.5). If a particular focus is already established it should be noted in the section of the description relating to the "relevance of the study".

The issue description is also used to ensure that the study is grounded in practice. The section on the relevance of the study encourages consideration to be given to the questions of what decision(s) you are trying to inform and what you need to know to make that decision.

Preparing an issue description helps to establish the rationale for conducting a study. Ensuring that a study will be able to inform decisions around practice (depending on its findings, of course), rather than merely satisfying curiosity, is important for a variety of reasons:

- **Resource implications**. Is this an issue that matters? If not, resources should be deployed on more important matters, which might include conducting studies on other subjects. Will the evidence inform responses to the issue? If the relevant decisions have already been made and cannot be swayed by evidence of some intervention being effective, may similarly not want to waste resources.
- **Ethical grounds**. Many studies will involve human participants. Where there is the potential of depriving some participants of an effective intervention it is particularly important that the knowledge gained from the study will advance practice to ensure better outcomes for others in the future. Even where the risk is low, involvement in a study may be taking up participants' time, which should not be wasted on studies that have no prospect of impact on practice.
- **Funding**. Whether funding is being provided by the organisation itself or sought from an external funder, describing how the study will deliver

evidences that has a realistic prospect of influencing practice will justify expenditure on the study.

- **Motivation**. Establishing that a study has the potential to improve practice in relation to some outcome of importance will help to provide motivation to those conducting the study, those delivering the intervention, and potential participants alike.

Understanding the link to practice and decisions will help to ensure that the evidence produced will be well-received, attended to, and acted upon by its potential users. Consequently, it will normally be necessary to involve evidence users in the preparation of the issue description, or at least to consult them on its content. There may be deadlines by which decisions need to be made, or it may be that evidence users can plan timetables such that they are able to schedule the decisions for when the findings of the study will be available; engaging with them at this stage will help to establish issues of timing, amongst other things.

5 Intervention design

5.1 General

There are several steps that need to be completed before studying starts if the aim is to create clear evidence that will be easy to use. Together they form a considered approach to intervention design, ensuring it is thought through to have a good chance of meeting a well-defined set of needs.

Not only are these steps essential for successful evidence production and use, they are also likely to have broader benefits, helping to design an intervention that is more likely to be effective and building a shared understanding of plans between the members of the team delivering the intervention.

The process specifies iteration around these steps. This is important because something established in one step may cause you to reflect back on conclusions you had provisionally reached in earlier steps.

These steps are equally relevant in relation to producing evidence of the effectiveness of an existing intervention or designing a new intervention.

5.2 Evidence review

5.2.1 General

In order to establish that it is appropriate to conduct a study, it is important to establish what is already known on the subject. Identifying that there is genuine uncertainty about the comparative benefits of alternative courses of action helps to establish the ethical case for conducting a study, and will also support the business case for using resources to study the potential interventions. Uncertainty about which intervention is most appropriate (either for the population as a whole or for a sub-population of interest) can be established by conducting an evidence review, and finding an absence of reliable evidence or inconsistency in the existing evidence.

Checking the existing state of evidence of the effectiveness of interventions related to the issue helps to establish what is already known and what remains unknown, forming the context in which further evidence can usefully be produced. Doing so will also help to support a focus on producing evidence that will add value and ensure that the process is rigorous.

The process of checking the existing evidence of the effectiveness of interventions relating to the issue will comprise three broad activities:

- Collecting evidence of the effectiveness of interventions that are relevant to the issue;
- Reviewing evidence of the effectiveness of interventions from individual studies;
- Considering the overall picture that emerges about an intervention if there are multiple studies relating to it.

As well as the practical motivations for undertaking an evidence review, it has also been argued that it is ethically important to conduct a review of the existing evidence before starting (particularly when producing evidence involving people). *[SOURCE: Chalmers and Nylenna, 2014]*

For a new intervention the existing evidence might help shape the intervention specification. For an existing intervention it is likely that the evidence review will instead be important for establishing whether the intervention is already well-studied, and hence informing the decision about whether to proceed to a study or not (see section 6.1).

As with many aspects of creating evidence, those checking the existing state of evidence in relation to a particular intervention or issue can contribute to the wider sector if they are able to share their evidence check, where appropriate. StEv2-1 contains a recommendation to consider publishing elements of an evidence review. If conducted on a particularly systematic and rigorous basis, the review of evidence becomes one of the most powerful and compelling types of evidence in its own right (see Annex A), but even when conducted on a lighter-touch basis, a completed evidence review might allow other people considering investigating the same topic to more quickly access some of the most relevant evidence.

5.2.2 Search for evidence

A search for evidence existing evidence may seek to identify what ideas have previously been tried to address the issue. Depending on the scope of the study, it might be appropriate to place some constraints around the search, to focus on a particular class of interventions.

A search for evidence for an intervention that has been specified should seek to identify whether there are previous studies that have been conducted examining its effectiveness. This should extend to searching for evidence of similar interventions. A thorough search would include an examination of whether the intervention has been studied regarding its effects in relation to other issues as well as the one that is the subject of this study.

Given the potential to expand the search in multiple directions, and the wide range of places that could be searched, this activity should be proportionate to the scope of the study that it is intended to inform; a large study might merit a very detailed and systematic approach to reviewing a wide range of relevant previous evidence, whilst a small study might only need to comprise searches for a few key terms to ensure that the most prominent evidence is taken into account.

5.2.3 Review of evidence from individual studies

The review of evidence from individual studies provides a consistent format to ensure that relevant information is captured in relation to each study identified. Completion of the information will support the evidence review to consider the important aspects of the study being reviewed. A proportionate approach will be required; if the information is not readily available from the report(s) of studies, the person conducting the evidence review will need to assess whether it is important enough to the process to merit deeper investigation.

This stage is intended to form a review of the evidence from a particular study. If more than one report emerged from a particular study, these should normally all be considered together within a single review of that study. If more than one study has independently examined the same (or similar) interventions, these should be considered separately.

For each item recorded about the study, the sheet notes that a record should be kept of where in the report the information was found. This is intended to

support the person reviewing the evidence, or anyone who uses the information gathered in the future, to refer back to it.

Background information is collected in order to uniquely identify the study and report(s):

- Name of study
- Brief description of study
- Title
- Author(s) and affiliation(s)
- Date of publication
- Source (web address or other)

Information on the conduct of the study is collected to support the assessment of the study, including the assessment of its relevance in the review's context:

- Study location
- Dates
- Roles
- Study population
- Intervention

The information on the study location, dates and study population will support consideration of whether its findings are likely to be valid 'here', 'now' and 'with your target population'. Information on the intervention, and the roles of those involved in its delivery, will help to identify whether it is likely to be feasible to deliver in the context of the available resource capacities. Section 5.4 of the standard provides a format for fully describing an intervention; collecting a full intervention specification as part of an evidence review will often be disproportionate (or impossible if not contained in the report) but the headings may serve as useful prompts for key considerations. As this stage forms part of an iterative process, those reviewing evidence may return to complete additional information about an intervention if they find that it is particularly important (for example if they are considering replicating it).

The reported results, including details of the outcomes measured, will support the assessment of whether the study appeared to find that the intervention was effective in achieving outcomes of interest in your context:

- Outcomes measured

- Reported results

Identifying the outcomes measured in a study also provides options for outcomes that could be measured in your study (see sections 7.3.1 and 7.3.2). For an outcome measure that you are considering using in your study, it may be appropriate to collect full details of the measure as specified in section 7.3.2. Study reports may also provide information on other determinants of the outcome of interest (i.e., determinants other than the intervention), for example identifying certain demographic factors that were associated with the outcome. Recording these will provide options for other measurements that could be made in your study (see section 7.3.3).

Assessment information is recorded to address several key aspects that will inform how heavily the evidence from the study should be weighted in considering the course of action:

- Evidence level
- Assessment of quality
- Pragmatic attitude
- Assessment of applicability

These cover both an assessment of the extent to which the evidence being reviewed is robust in its own right, and the extent to which it will be applicable in the context of interest.

5.2.4 Collating evidence on interventions

Piecing together the various assessments of one or more studies of an intervention is intended to support a judgement about the validity of a statement along the lines: "this intervention is effective in contexts like the one that I am interested in". 'Validity' is an overarching term for a judgement about the extent to which relevant evidence supports a proposition. The principal component elements of a validity judgement are:

- Do the studies reveal a reliable correlation between the presumed cause and effect? (**Statistical conclusion validity**)
- Is there reason to believe that any correlation observed was a result of a causal relationship? (**Internal validity**)
- Do findings from the intervention as delivered in the study generalise to the intervention of interest? Do the changes in the outcomes that the

study measured convincingly demonstrate an impact on outcomes that are of interest? (**Construct validity**)

- Will the causal relationship remain when the intervention is delivered for different people in different settings? (**External validity**)

[ADAPTED FROM: Shadish, Cook and Campbell (2002), p34 (definition of validity) and pp37-38 (validity typology).]

For the purposes of assessing the applicability of evidence to your context, the relevant question of external validity is more specific: do you expect the causal relationship to generalise to the people that you are interested in delivering it to, and the setting in which you will be delivering it?

Assessing the quality and applicability of evidence is conducted at the 'intervention' level not at the study level (although these become effectively the same where only one study is found in relation to a particular intervention). Where multiple studies have been conducted on very similar interventions, it makes sense to consider the overall picture that is built up from the different studies, not just each of them independently.

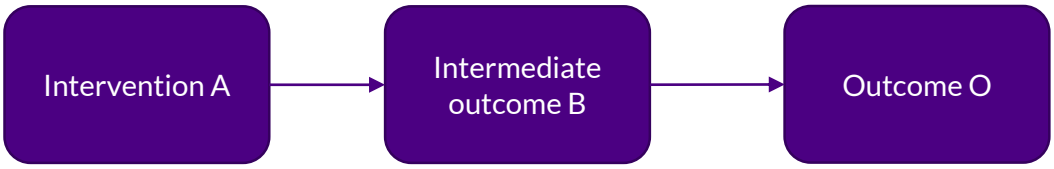
Where there are multiple studies on the same intervention, this stage will help to building an overall impression of the evidence of effectiveness of the intervention, including identifying any inconsistencies between the findings of different studies. It is not always appropriate to assume that several studies of weaker design add up to something more robust; there might be something systematic in the conduct of the studies that means that weaker designs will tend to over- or under-estimate the impact of an intervention. However, if there have been multiple relatively strong studies that each show an intervention to be successful in differing contexts, that would typically support a conclusion that the intervention is likely to be successful in a range of settings.

As noted in the process diagram (see section 3), the activity of reviewing evidence is one of a set that should be iterated around. The evidence that is located and assessed can inform causal chain mapping and intervention specification and those stages can in turn may inform a need for (further) evidence reviewing to be conducted.

5.3 Causal chain mapping

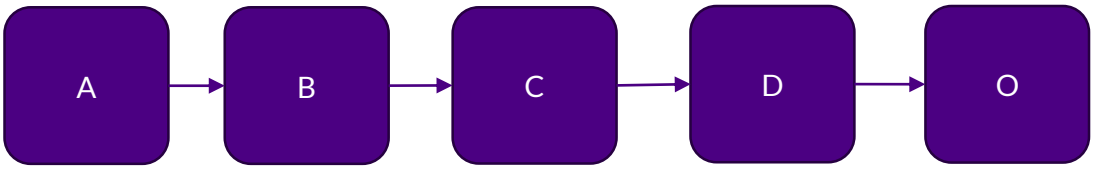
There exists a range of exercises that are designed to support those designing an intervention to think through how the inputs will feed through a set of interactions to create the ultimate outcome of interest. At their core these techniques encourage those designing interventions to build chains that start from the intervention and end with the ultimate outcome of interest. The chains are commonly presented visually, as a flow chart with arrows indicating the steps that flow from each other.

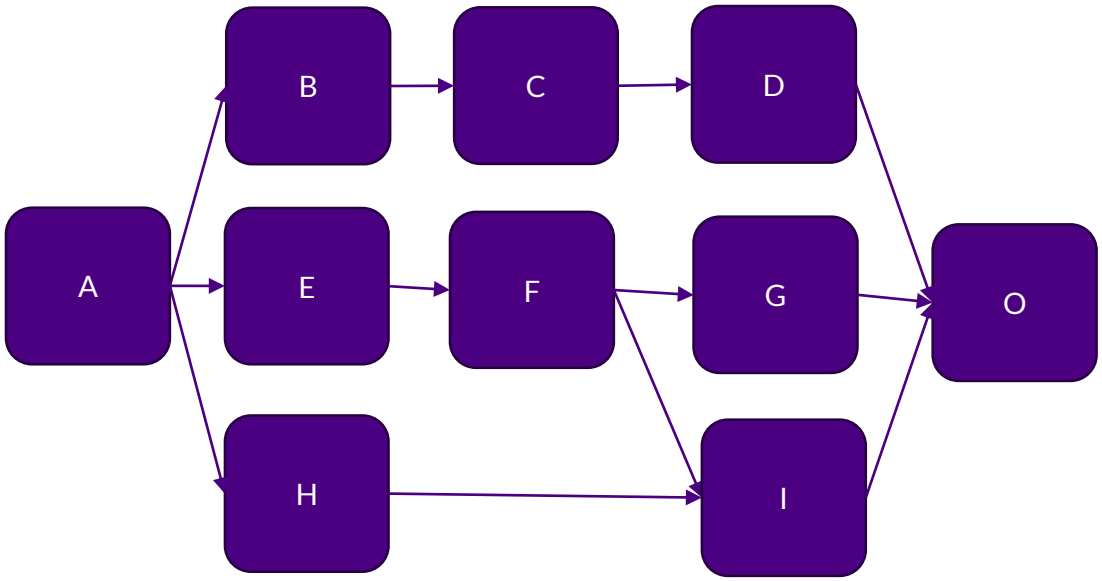
A very simple causal chain looks like this:



Each of the links in the chain represents an assumption, which may be interpreted as "intervention A will cause intermediate outcome B; intermediate outcome B will cause the ultimate outcome of interest, O".

These chains can get more complex either through a longer chain of intermediate outcomes or through identifying a more complex range of mechanisms by which it is thought an intervention might lead to the same ultimate outcome:





Considering the path by which it is hoped that the proposed intervention will contribute to improved outcomes can help to increase the clarity with which all elements of the study are planned. This includes ensuring that the intervention is clearly articulated and that the intended outcomes are well specified.

Causal chain mapping can be useful to prioritise what outcome measures to select (see section 7.3.1). Mapping the expected causal pathways can particularly help to inform the selection of intermediate outcomes that could be measured alongside the main outcome measures of interest.

Alternatively, where the ultimate outcomes of interest are very long term, causal chain mapping may support the identification of intermediate outcomes that are more practical to measure on a shorter timescale. Where the scope of the study is such that long term measurement is impractical, these intermediate outcomes may allow a study to be designed that still generates some useful insights. Even where a study is able to run for long enough to measure the ultimate outcomes, shorter term intermediate outcomes may still be able to provide early indicative findings.

Not all interventions are unambiguously beneficial, and causal chain mapping can also help to identify pathways by which negative outcomes might be caused by an intervention.

As noted in the process diagram (see section 3), causal chain mapping is one of a set that should be iterated around. It can inform the initial design of an intervention and can also be revisited as the detailed intervention specification is developed. It can be informed by an evidence review that finds existing evidence of certain causal chains and can also provide insight into where it would be helpful to search to see if there is more evidence that could be reviewed.

5.4 Intervention specification

If the evidence produced is to be usable subsequently it is vital to have a clear definition of the intervention. Providing sufficient information to fully describe the intervention includes not just a detailed description of what will be delivered, but also requires relevant information on how it is delivered, by whom, where, when, and why (the rationale behind elements of the intervention). The TIDieR checklist has been developed to ensure that those reporting on interventions have gathered sufficient information to adequately describe an intervention. StEv2-1 specifies an intervention specification that closely follows the elements of the TIDieR checklist.

> *[NOTE: Detailed explanation of the elements of the TIDieR checklist are not duplicated here. For information on those points, refer to the TIDieR checklist and guide (Hoffmann et al 2014).]*

A description is required for all arms of the study, which will typically include a comparison group as well as the proposed intervention(s) to be tested. This is important in order to ensure there is clarity over what the intervention was compared with. The "why" section of the specification for the comparison group should include a rationale for why this is a suitable intervention to be used as a comparison in the study. This may include it being the 'business-as-usual' approach, the current best available intervention (based on previous evidence of effectiveness) or, for a no-intervention comparison, it could be justified on the grounds of there being no proven effective intervention.

Preparing an intervention specification is as important for existing interventions as it is for new ones. Existing interventions will have often developed over time from their initial state or never been fully described. This can also be a useful opportunity to investigate variation in delivery: if there are multiple people delivering the intervention, do they all deliver it in the same way, or is there variation? Note that in a pragmatic study the

variation in delivery may be allowed to remain, unless there is reason to believe that upon wider deployment of the intervention it would be possible to ensure implementation in a specified fashion.

The actual design of the intervention should draw upon the evidence review and the causal chain mapping. Having identified and reviewed evidence, it can then be used to strengthen the efforts of a project to create new evidence. The form of this activity will be dependent on the state of the existing evidence. It may include modifying interventions to incorporate promising features from elsewhere, identifying a novel intervention where previous ones have been found to be consistently ineffective, or replicating the study of an intervention that has been found to be effective in other contexts.

As noted in the process diagram (see section 3), the activity of specifying the intervention is one of a set that should be iterated around. The intervention specification may be informed by evidence found in the evidence review, and as it is developed may inform a need to search for more evidence, particularly to see if there is anything relevant to the specific design. The intervention specification may also be informed by causal chain mapping and may require revising a causal chain map to ensure the assumptions still appear appropriate as the intervention is specified in more detail.

6 Decision to proceed

6.1 Proceeding to study

Whilst the standard has been designed to support the production of evidence, and is based on an assumption that such evidence is valuable, this stage is included in the process in recognition of the fact that in some instances it is not appropriate to produce further evidence. A study should only normally be conducted:

- where there is a knowledge gap that a study would address; and
- where gaining the knowledge that the study will provide is likely to be able to inform practice in some way.

If a study does not have a reasonable prospect of meeting these criteria, it is likely to be a poor use of constrained resources, and (particularly if it is being conducted with human participants) may present ethical concerns.

The evidence review (section 5.2) will inform the assessment of whether there is a knowledge gap. Referring back to (and refining if necessary) the issue description (section 4) will support the assessment of whether there is a match between the knowledge that would be produced by a study and the knowledge that would inform practice decisions.

For many types of interventions, an assessment of whether there is a significant knowledge gap will involve the use of judgement. Specifically, the effectiveness of many interventions is likely to context specific, and even if tested robustly in other times and places the effect of deploying it in a new context will be subject to some uncertainty. The weight to place on that uncertainty in making a decision to proceed to study will be dependent on an assessment of the similarity of the context of interest to the contexts of any earlier studies, but also linked to practical considerations regarding the importance of the issue, the resources required to deploy an intervention, the resources required to conduct a study, and other potential pieces of evidence that could be produced using the resources of the study.

For an issue of central importance, where significant resources are going to be expended on the intervention selected, it may make sense to resolve even small degrees of ambiguity in the evidence base.

For a more peripheral issue, which is going to account for relatively low resource usage, it may be appropriate to use an intervention that has accrued a moderate amount of evidence in other contexts rather than trying to absolutely optimise outcomes through further study.

Where an organisation has several potential issues where evidence of the effectiveness of interventions would be useful, resource requirements may necessitate prioritising the production of new evidence. All else being equal, it may be appropriate, for example, to study first those issues where there is no existing evidence of interventions being effective; interventions that have some evidence of effectiveness from other contexts could be deployed in the first instance and scheduled for later study in the context of interest.

Finding evidence for an intervention being effective does not automatically create a requirement that it should be deployed without further study. In some cases, as above, it may be appropriate to conduct further study to establish whether it is effective in a different context. Alternatively (or in addition) it may be appropriate to study other interventions aimed at achieving the same outcomes, to see if they are even more effective (or as effective at lower cost).

Similarly, finding that the existing evidence base demonstrates that a particular intervention has previously been found to be ineffective does not create an absolute assumption that it should not be studied again. To justify further study, however, it would normally be appropriate to ensure there is a plausible rationale for believing that the intervention might be more effective the context of interest than it was in the ones where it was previously studied, or that modifications being made to the intervention would plausibly boost its effectiveness.

6.2 Study levels

The specification of three levels in the standard has been informed by the many hierarchies of evidence that have been drawn up across various sectors. The levels draw on the well-established understanding of the relative merits of different study designs, and their abilities to produce robust evidence of the effectiveness of interventions, with each suited to different purposes, and relevant in different stages in the evolution of the evidence base for a particular intervention.

This hierarchy is typical of hierarchies of evidence found in the medical literature:

1. Systematic reviews and meta-analyses;
2. Randomised controlled trials with definitive results (confidence intervals that do not overlap the threshold clinically significant effect);
3. Randomised controlled trials with non-definitive results (a point estimate that suggests a clinically significant effect but with confidence intervals overlapping the threshold for this effect);
4. Cohort studies;
5. Case-control studies;
6. Cross sectional surveys; and
7. Case reports.

[SOURCE: Greenhalgh, 1997.]

Whilst different hierarchies sometimes extend to different points at the top or bottom or group certain study designs together, there is little dispute between experts on the relative abilities of different study designs to generate evidence of effectiveness (subject to studies being designed and conducted well).

The standard provides a framework for considering which level of evidence is most appropriate to produce. Consideration of the purpose of the study, the acceptability of various limitations, and the usage that the evidence will be put to should inform the decision about the appropriate study level.

The different sets of requirements and recommendations of the standard for studies at different levels are summarised in a table. Whilst there are many similarities in overall form between level 3 studies and those at level 1 or 2, the details are fairly different. Consequently, the requirements and recommendations for level 3 studies are presented separately in an annex.

7 Study planning

7.1 General

The process specifies iteration around several steps within the study planning process. This is important because something established in one step may inform thinking about how best to specify other elements of the study.

7.2 Assemble study team

Depending on the type of study being planned, a range of skills and expertise will be needed to give the study the best chance of being completed to a satisfactory conclusion. The skills to provide the research skills for the study might include:

- Statistician
- Trials methodologist
- Qualitative researcher
- Economist
- Someone with knowledge of relevant data sets
- Fieldworkers who can deliver surveys

In addition, there will need to be suitable skills available within the team to deliver the intervention.

A further consideration in relation to the study team will relate to independence (and perceived independence). If all of those responsible for a study have vested interests in the intervention being effective, there is a risk of the study's findings being viewed with suspicion, even if it was scrupulously conducted.

7.3 Measurements

7.3.1 Selection of outcome measures

Describing the issue and selecting the main outcomes that an intervention has been designed to address will often be two sides of the same coin: the issue is that people are in a certain bad situation, and the desired outcome might be that we would like fewer of them to be in that situation, or for the extent to which they experience that situation to be reduced. Reference to the issue description (see section 4) will be necessary to ensure that outcome measures are selected that will produce evidence that will answer the questions of evidence users.

Primary and secondary outcomes

The primary outcome measure will provide the principal assessment of whether the intervention is effective or not. It should be selected on the basis of being the best measure of the main outcome of interest in the study. It will also have special status in the study, for example in reporting the findings and in setting the sample sizes.

As well as a primary outcome measure, secondary outcome measures can be selected for a number of reasons:

- Establishing the effectiveness of intervention at delivering other benefits besides the primary target;
- Assessing the extent of any adverse outcomes;
- Monitoring intermediate outcomes to assess the propagation of the impact of the intervention along the causal chain;
- Assessing the effectiveness of the intervention at timepoints other than the primary timepoint of interest;
- Collecting an alternative measure for compatibility with other studies.

Often, there will be **multiple ultimate outcomes** that could relate to the same issue. In this case it is important to identify the primary (main) outcome that is felt to be the most relevant: which one thing is most important to achieve. Secondary outcome measures can be selected to assess the effectiveness of the intervention in achieving other benefits that it might deliver but gaining clarity about which is the primary outcome will help to

make it easier to make decisions if trade-offs are required, and to allow the intervention to be designed with a clear target in mind.

Secondary outcome measures can also be selected to assess the extent of potential **adverse outcomes** (side effects or unintended consequences). Because adverse *events* can happen in any population, irrespective of the presence of an intervention, simply monitoring the number of adverse events in the intervention arm of a study would not give a robust assessment of the level of adverse *outcomes* (i.e., adverse events that are causally linked to the intervention). Establishing a causal link between an intervention and adverse events requires the same methods as establishing a causal link to a beneficial outcome. The use of a causal chain map (see section 5.3) may help to identify potential adverse outcomes that might be associated with the intervention.

Measuring **intermediate outcomes** can help to understand in more detail how an intervention works, or why it does not, by supporting the identification of where any drop-off in performance occurs. Reference to the causal chain map (see section 5.3) may support consideration of all of the points at which the link between the intervention and the primary outcome might break down. If it is possible to specify secondary outcome measures for some of the links in the chain, it might be possible to identify 'why not', if an intervention fails to achieve its desired impact.

Intermediate outcomes explaining why an intervention did not work

"The [intervention] combines social mobilisation with government subsidy for toilet construction [in the state of Odisha in India]. The study collected data on several indicators of latrine use such as the smell of faeces, stain from faeces or urine, the presence of soap, the presence of a broom or brush for cleaning, and the presence of slippers. The researchers also tested for faecal indicator bacteria in water sources and in household drinking water, as well as on children's and mothers' hands and on children's toys. They tested for hand contamination of household members using hand rinse samples. And they set fly traps to measure the density of flies. … And so when there was no improvement in child diarrhoea

despite a substantial increase in the latrine coverage the researchers could identify the likely reason: not all family members were using the latrine."

[SOURCE: White, 2015.]

For outcomes that are measured as a snapshot (as opposed to things like 'time to event' outcome measures), it may be possible to collect data at **multiple timepoints**. Data on the primary outcome, but collected at timepoints other than the time of primary interest, should be treated much the same as secondary outcomes. The same considerations regarding the importance of selecting one primary outcome also apply to selecting a primary timepoint of interest.

There are reasons for adopting, where appropriate, the **same outcome measures as other studies** (see below). However, it may sometimes happen that an outcome measure that has been used in other studies appears to represent too much of a compromise from another possible outcome measure (for example because it is not a good enough fit for the needs of evidence users). In these instances it may make sense to collect data on it as a secondary outcome measure, in order to facilitate comparisons with other interventions or merging of data in future meta-analysis of studies of the same intervention.

Burden of outcome measures

Selecting a large number of outcome measures can present at least two potential problems. Firstly, the increased of data collection (and analysis) is likely to require additional resources and may place additional burdens on those implementing the intervention and/or study participants. The extent to which this will happen and its impact will depend on the nature and context of the study, and whether any of the outcome measures are based upon routinely-collected data.

The second issue relates to the potential for multiple comparisons to generate spurious findings of associations: as an increasing number of outcome measures is selected, it becomes more and more likely that one of them will vary between the intervention and comparison arms due to pure chance. The impact of this problem can be minimised by treating all findings from secondary outcomes as tentative and through statistical techniques to control for multiple comparisons.

Outcomes relevant to evidence users should be favoured

Producing evidence that directly relates to the outcome that evidence users are specifically interested in will increase the likelihood of the evidence being acted on.

Considering the accuracy and reliability of representation of the underlying outcome

Where it is impossible or impracticable to directly measure the underlying outcome of interest (for example, mental health states), some measureable alternative will need to be selected. In these instances it will be necessary to assess the construct validity of potential measures, i.e., to consider whether changes in them would convincingly demonstrate an impact on the ultimate outcomes of interest. Where measures have previously been used in other studies there may have accrued evidence of construct validity.

It is also particularly important to seek a measure that is reliable. There are three major forms of reliability that should be sought: test–retest, internal consistency, and inter-rater. *[SOURCE: Lilienfeld et al, 2015].*

Direct measures of outcomes are favoured over surrogates

Surrogate outcomes are used when the outcome of interest is hard to measure – perhaps that it will take a long time to become visible. In addition to surrogate outcomes being less directly relevant to evidence users, a further potential problem occurs if these markers turn out not to be such good indicators of the actual intended outcome as was previously believed.

When surrogate outcome measures mislead

In a health study, if the outcome of interest is lowered incidence of heart attacks, but the target population is expected to be at an elevated risk of heart attacks that would be detectable over a period of many years rather than months, the study might choose instead to look at some other marker that is thought to correlate with heart attacks.

In one notable example, anti-arrhythmic drugs (drugs designed to stabilise an irregular heartbeat) were found to be effective at the surrogate outcome of stabilising the heartbeat; it was thought that because an irregular heartbeat is normally associated with a higher risk of heart attack and death that this stabilisation would lower death rates. However, when longer-term studies looked at actual measures of interest (heart

attacks or deaths) they found that patients treated with the anti-arrhythmics actually had a significantly elevated risk of death.

[SOURCE: Epstein et al 1993]

Ease of data collection should be a factor in selecting outcome measures

Data collection can comprise a significant component of the resource requirements of a study. Consequently it is desirable, if possible, to specify outcome measures where data are already collected as part of normal operations, or where normal practice could be easily amended to collect the data for little extra effort. As well as reducing the resource requirement for those conducting the study, using routinely-collected data may reduce the burden upon study participants and those delivering the interventions, lowering the likelihood of drop-outs, dissatisfaction and lower data quality through fatigue.

Favouring the re-use of outcome measures from other studies

As well as the possibility that outcome measures used in other studies will have accrued evidence of construct validity (see above) there are further benefits from learning from the efforts of previous studies. Those who have previously specified an outcome measure will have already done the work of establishing exactly how it should be defined, which may be important where various possibilities are available. Furthermore, using the same outcome measure as other study is particularly helpful in comparing and merging the findings from different studies. If two different interventions have been assessed for effectiveness using the same outcome measure it will be easier to make a comparison between them. If the same intervention has been the subject of multiple studies using the same outcome measure it will facilitate the combining of the findings of the studies in subsequent systematic reviews and meta-analyses.

7.3.2 Specification of outcome measures

The outcome measurement sheet should be completed for the primary and secondary outcomes selected as being of interest in the study.

For each outcome, the minimum practically important difference should be established. This difference is the smallest amount of difference that would

matter for comparing the intervention to the alternative. Where a new intervention is being studied, this might be the amount by which it would have to outperform the existing standard approach to be worth considering deploying. Establishing the minimum practically important difference may require a small qualitative piece of research with decision-makers in the implementing organisation to identify. Having an accurate impression of the minimum practically important difference is particularly important for the primary outcome as this will be the main factor affecting how large the sample sizes need to be for the study.

7.3.3 Specification of other measurements

The collection of data other than outcomes measures, such as baseline demographic information on study participants, can be useful in various ways depending on the study design.

Where there are important known determinants of the outcome of interest, recording data on these at the study outset can help to establish whether the members of the intervention arm and any comparison arm are similar on terms of those characteristics; this can include establishing whether randomisation has been 'successful' in a randomised controlled trial.

Alternatively, baseline demographic data can be used in sorting participants into the arms of the study. In a randomised controlled trial this can be achieved by stratified randomisation, whereby participants are split into groups based on an important determinant of the outcome (e.g. gender) and then randomised within those groups (e.g., men randomised between the arms and women randomised between the two arms). This avoids the potential for 'chance bias', whereby the arms end up unbalanced in relation to the determinant by pure chance (e.g. men over-represented in the intervention arm and women over-represented in the control arm).

In some study designs this data can be used in statistical analyses in an attempt to control for the effect of any differences between the members of the different arms of the study.

7.4 Study design specification

7.4.1 Non-causal designs

Non-causal designs are those that can give an indication of whether an outcome appears to be associated with the intervention but cannot give a robust indication that the intervention actually causes the outcome. They are helpful in gaining an early impression of what the scale of any impact might be, and are particularly suited to being used alongside process evaluation techniques to test the feasibility of delivering the intervention.

Non-causal designs may also be suitable where the intervention being examined is inherently small in scale, to the extent that its scale prohibits the use of robust causal designs.

As a minimum, non-causal designs require the measurement of the outcome(s) of interest after the intervention has taken place. On its own, however, such a measurement will generally provide little information about whether the intervention is associated with that value for the outcome, because there would be nothing to compare it against. Consequently, one of the following approaches should normally be used:

- **Pre/post**: measurement of the outcome measures before and after the intervention.
- **Post-test with comparison group**: measurement of the outcome measures for both the group receiving the intervention and some other group that does not receive it.
- **Pre/post with comparison group**: measurement of the measures outcome both before and after the intervention for the group receiving the intervention and some other group that does not receive it.

In the variants where measurements are taken before and after the intervention it is possible to assess whether there is any change in the outcome. In the variants where there is a comparison group it is possible to assess how the intervention group compares to a group that did not receive the intervention. It should be stressed, however, that even where both techniques are combined in these methods there will remain the possibility that an association (for example the outcome of interest improving for the intervention group whilst it remains steady for the comparison group) could be due to factors other than the intervention. Consequently, finding an

association should only be treated as indicative of a **possible** impact, not evidence of the intervention being effective.

Where possible, the comparison group should be selected to be similar to the intervention group in terms of characteristics that are likely to be relevant to the outcome of interest. This will increase the likelihood that any association observed is a consequence of the intervention rather than a result of differences between the groups. However, this will still not account for all possible causes of difference between the groups, so the findings will still not support robust causal inference. These sorts of comparison groups are called 'nonequivalent groups' in the technical descriptions of these designs *[SOURCE: For example, Shadish, Cook and Campbell, 2002, p136.]*, in contrast to randomly assigned control groups, which are statistically equivalent on all observed and unobserved characteristics.

Where designs using non-equivalent groups are combined with appropriate statistical analyses, they may constitute quasi-experimental methods and be able to support robust causal inference (see section 7.4.2).

7.4.2 Designs that support robust causal inference

Study designs that support robust causal inference are those that have the potential to generate evidence of the effectiveness of interventions, addressing the issue of cause and effect, rather than just correlation. These designs rely upon their ability to establish a robust 'counterfactual', i.e., an indication of what would have happened to those receiving the intervention if they had not received it. The creation of evidence at level 2 will typically require organisations to draw on external research expertise, both due to the skills requirements of the methods and because independence in the study team can increase the perceived credibility of findings.

Where it is possible and acceptable to do so, a well-designed and well-delivered randomised controlled trial is the strongest way (in a single study) of establishing the effectiveness of interventions. The key advantages of randomised controlled trials in establishing the effectiveness of interventions arise from their ability to avoid the potential for selection bias to be present:

> *"[W]hen properly implemented, [randomisation] eliminates selection bias, balancing both known and unknown prognostic factors, in the assignment of treatments. Without randomisation, treatment comparisons may be*

> *prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment."*

[SOURCE: Moher et al, 2010.]

By reducing or removing the potential for selection bias, a randomised controlled trial removes a major alternative potential explanation of any differences observed between those receiving the intervention and those in the comparison arm.

However, there are also few who would deny that there are circumstances in which randomised controlled trials are not possible, feasible or appropriate; in these instances other designs seek to control for the range of potential biases, in order to create a robust assessment of whether any observed difference in outcomes between the arms of the study are causally due to the intervention. The circumstances in which RCTs may be unsuitable are discussed in Annex B; some alternative study designs are discussed in Annex C.

Several variant RCT designs exist, increasing the situations in which they can be deployed. Cluster randomisation designs, for example, involve randomisation at the level of a social unit rather than an individual (for example, randomising different neighbourhoods into the assignment groups), which can address some concerns of contamination. Waiting list based designs exist in which all of the participants will eventually get the intervention, but the order is randomised so some get it immediately and other have to wait, which may overcome ethical concerns in situations where roll-out of a known-effective intervention would have to be phased anyway. Some challenges caused by studies being unrepresentative may be overcome by designing them to be more 'pragmatic' (see below).

Whether specifying an RCT or an alternative design, the requirements of a level 2 study are likely to include the need for particular expertise. Those with the necessary expertise to implement an intervention will typically not be the same people with the expertise to conduct a study with a design that supports robust causal inference. The use of independent study partners would facilitate access to necessary expertise and also reduce the potential for perceived conflicts of interest, such as might occur where those responsible for implementing an intervention are also responsible for producing evidence of its effectiveness.

Studies should be designed to be **pragmatic** (as opposed to 'explanatory') in attitude. The intention with a pragmatic study is to establish whether the intervention is likely to work in real practice, whereas explanatory studies seek to identify whether an intervention can work in perfect conditions. Consequently, pragmatic studies are designed to resemble the situation in normal practice as closely as possible, rather than being delivered under tight controls. *[SOURCE: Schwartz and Lellouch, 2009.]* They would typically also be characterised by not having extremes of resources, training, or specialist staff conducting the intervention. (The study team will still have expertise that will not generally be present in ordinary practice, but that reflects the resources to assess the effectiveness of the intervention, not to implement it.) The following table outlines the key differences between exploratory and pragmatic trials:

	Exploratory	Pragmatic
Question	Efficacy—can the intervention work?	Effectiveness—does the intervention work when used in normal practice?
Setting	Well resourced, "ideal" setting	Normal practice
Participants	Highly selected. Poorly adherent participants and those with conditions which might dilute the effect are often excluded	Little or no selection beyond the clinical indication of interest
Intervention	Strictly enforced and adherence is monitored closely	Applied flexibly as it would be in normal practice
Outcomes	Often short term surrogates or process measures	Directly relevant to participants, funders, communities, and healthcare practitioners

Relevance to practice	Indirect—little effort made to match design of trial to decision making needs of those in usual setting in which intervention will be implemented	Direct—trial is designed to meet needs of those making decisions about treatment options in setting in which intervention will be implemented

[SOURCE: Table 1 in Zwarenstein et al 2008.]

Although the goal in studies is often to identify an intervention that is an improvement on the existing practice (known as 'superiority'), this is not the only potential use. One alternative framework is to design studies that test 'non-inferiority', to establish whether a new intervention is at least as good as the status quo in terms of outcomes (which can be useful if the new intervention has other desirable features, such as being cheaper to deliver or more acceptable to a target population). Another framework is 'equivalence', where the intention is to find out whether two options deliver the same outcomes.

It may be possible to design a study that is able to address equity issues by identifying disadvantaged groups for subgroup analysis. One factor affecting the feasibility of this will be the available sample sizes, and whether it is possible to have enough participants from the subgroup identified to come to statistically significant conclusions. The PROGRESS-Plus mnemonic can be used to consider which disadvantaged groups could be focused on in the study:

> *"Disadvantage can be measured across categories of social differentiation, using the mnemonic PROGRESS-Plus. PROGRESS is an acronym for Place of Residence, Race/Ethnicity, Occupation, Gender, Religion, Education, Socioeconomic Status, and Social Capital, and Plus represents additional categories such as Age, Disability, and Sexual Orientation."*

[SOURCE: Ueffing et al 2012, citing Evans, 2003 and Oliver, 2008.]

7.5 Specification of participant recruitment approach

Establishing an approach to participant recruitment should normally start by identifying the target population for the study, and is distinct from the approach to assigning participants into intervention and comparison arms. The study population may be the entirety of people that the organisation works with who are affected by the issue the study is considering, or it could be some subset of them, such as those living in particular areas.

Eligibility criteria

Within the overall population, the eligibility criteria establish any particular characteristics that make potential participants eligible or ineligible for participation in the study. In general the more inclusive these criteria can be, the more likely it will be that the resulting evidence will be generalisable (i.e., applicable in a range of contexts).

The standard requires the eligibility criteria to be accompanied by details of the extent to which they include the typical population the intervention would be delivered to. This supports a pragmatic approach to the study design. If the eligibility criteria closely mirror the criteria that would be used to allocate the intervention if it were demonstrated to be effective, the study will have a more pragmatic attitude.

Criteria excluding participants should be used primarily for reasons of safety / harm reduction, if there are groups where there would be a substantially elevated risk of negative outcomes if they were included in the study.

> *[NOTE: Historically there was a practice of separately stating inclusion criteria and exclusion criteria. This practice is now viewed to be unnecessary as criteria can be reframed in the positive or the negative, either to include or exclude people.]*

Sample size

The minimum target sample size will be driven (especially for designs that support robust causal inference) by the minimum practically important difference that has been identified for the primary outcome measure (see section 7.3.2). Relevant calculations should be undertaken to ensure that a

difference of this size would be detectable by the study. (The minimum detectable effect size of a study is determined in large part by the sample size; consequently, the sample size should be set such that the minimum detectable effect size is at least as large as the minimum practically important difference.)

Conducting these calculations – known as power calculations – will require someone with statistical expertise.

Further power calculations could also be conducted in relation for any particularly important secondary outcomes that it would be important to detect (such as important adverse outcomes). They could also be conducted for any important subgroup analyses that are planned (for example analyses of the effectiveness of the intervention for a particular disadvantaged group).

Transparency in sample size calculations supports those reviewing the evidence from the study to satisfy themselves that they have been correctly conducted. It will also allow future studies to refer to the assumptions made in the calculations and to re-use them, if appropriate.

7.6 Ethical considerations

A thorough guide to the topic of research ethics is beyond the scope of this document, but the information below contains some pointers to some of the most relevant issues to consider. For a more detailed grounding in the issue, see the Research Ethics Guidebook. *[SEE: Boddy et al., n.d.]*

The ethical implications of research can appear intimidating, especially to those who are new to the field. A general principle is that research should aim to do good and avoid doing harm. In this light, one of the most important ethical considerations is the amount of activity that is undertaken by organisations with no significant evidence of effectiveness. This inevitably means that at least some of these scarce resources are being wasted on interventions that are ineffective (or possibly even harmful), and that consequently the best possible outcomes are not being achieved.

Ethical requirements should not be viewed as a brake on producing evidence. Adherence to ethical approaches will contribute to ensuring the continued trust of study participants. Ethical considerations will often point in the direction of **more** robust evidence production: it is ethically

problematic to experiment on people (i.e., try out new interventions that might or might not be beneficial) in a way that fails to produce evidence of the effectiveness of those interventions.

A proportionate approach to study ethics should be adopted. Even in healthcare and the testing of medicines there are currently proposals being discussed to allow for much lighter touch consents where the investigation is looking at two interventions that could reasonably be given anyway (such as two drugs that are licensed for the same condition and there is currently not evidence of which works better) *[REFERENCE: Collett, 2014.]*.

The Research Ethics Guidebook notes that there is not, generally, statute law requiring those producing evidence to observe ethical guidelines. *[REFERENCE: Boddy et al, n.d., page: Legal requirements.]* However there may be a legal obligation to seek **ethical review** if the study could be classified as health research. *[REFERENCE: National Research Ethics Service, n.d.]* Even in the absence of a legal obligation, it may in some cases be a requirement of a research funder that ethical approval be sought, if external funds are being drawn upon to conduct the research. In any event, everyone generating evidence still needs to act in an ethical way, even where formal ethical approval is not required from any particular body.

One of the main justifications for it being ethical to undertake a study is that there exists a state of '**equipoise**' – i.e. that there is a degree of uncertainty about which of the options being compared is superior. If an intervention has already conclusively been shown to be effective then it would normally be ethically problematic to compare it to a control group receiving nothing. It would, however, still be acceptable to test it against another intervention of unknown effectiveness, or to compare it against 'no treatment' in a new context where it is not yet known whether the intervention would be effective. One exception to the requirement for equipoise can be a situation where there is a **natural delay**: where the intervention would be rolled out in a phased way to different beneficiaries anyway, then it can be ethical to conduct a study comparing those who get the intervention early against those who have not yet received it because they are due to get it later.

Even aside from ethical concerns, there will typically be a legal requirement under the Data Protection Act to maintain the **confidentiality** of individuals. This is normally complied with by reporting findings in aggregate and/or by ensuring that any individual-level information is reported in such a way that it is not possible to identify the person concerned. In sensitive situations, plans should be made to identify the conditions under which it is ethical to

break confidence; this will normally be when there is a safeguarding issue (i.e., when someone involved in conducting the study identifies a situation in which someone is likely to cause harm to self or others).

Participation in studies should normally be voluntary, and plans should be made to ensure that research participants give **informed consent**. This means that people agree to participate, and that they have had things explained to them such that they understand what they are agreeing to. There are some circumstances where this might not be feasible or appropriate, such as an intervention to test the effectiveness of receiving different wording in letters to encourage specific actions. Another example might be informational interventions, where information might be delivered at a neighbourhood level; in that situation it could be unfeasible to seek everyone's permission and might undermine the findings anyway if people were warned in advance that there could be a boost in the information being delivered in their area. Where consent is not sought, extra care should be taken to ensure that the study has a minimal risk of harm.

There are also ethical issues related to ensuring research evidence is accessible and used. **Publication bias** can be viewed as an ethical issue: failing to publish null findings means that others might waste resources by continuing to deliver the intervention, and that service users continue to receive ineffective interventions. Some have argued that **not searching the existing evidence base** before considering trying something new is ethically problematic, *[SOURCE: Chalmers and Nylenna, 2014]* especially in situations where the effectiveness of the proposed new intervention might already have been conclusively established.

7.7 Process evaluation design specification

Process evaluation comprises work to understand **how** an intervention is implemented, **why** it seems to work or not, and **what** contextual factors are affecting it; this is distinct from, and complementary to, the task of assessing **whether** an outcome was achieved (the effectiveness of the intervention). Process evaluation is also particularly suited to generating the evidence that will help to improve interventions that are not initially effective or identifying potential improvements to those that are effective to maximise their impact.

When conducted as part of a level 1 study, a key purpose for this step is to assess both the feasibility of delivering the intervention and the feasibility of conducting further studies. If the intervention is progressed to the stage of undertaking a level 2 study, it will be helpful to know whether it is likely to be possible to successfully run a study (evaluation feasibility), as well as knowing whether the intervention can be delivered by the organisation (implementation feasibility). This protects against the respective threats of evaluation failure and implementation failure. Consequently, specifying a process evaluation design is a requirement of level 1 studies.

The relative importance of these aspects will vary depending on the situation for the study:

- For a **level 1 study of a novel intervention** (or an existing intervention in a new context with a different group of service users), the process evaluation might be designed to examine both implementation feasibility and evaluation feasibility.
- For a **level 1 study of an existing intervention**, where the feasibility of implementation is already well understood, the process evaluation might be focused particularly on evaluation feasibility.
- For a **level 2 study**, the process evaluation is most likely to be used to study implementation feasibility, as there is less likely to be a future study for the intervention to progress to (so assessing evaluation feasibility would be of less importance).

Attempts to evaluate feasibility may not simply result in 'yes' or 'no' assessments. They may instead seek to establish what additional resources, training, etc. might be needed to make the intervention or evaluation feasible.

Many (but not all) of the methods associated with process evaluation will be qualitative in nature. When using qualitative methods, the four central principles laid out in the HM Treasury Magenta Book supplement on quality of qualitative evidence should be followed, with research being:

- **Contributory** in advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field;
- **Defensible in design** by providing a research strategy that can address the evaluative questions posed;
- **Rigorous in conduct** through the systematic and transparent collection, analysis and interpretation of qualitative data; and

- **Credible in claim** through offering well-founded and plausible arguments about the significance of the evidence generated.

[SOURCE: Spencer et al. 2012, p11.]

Whichever methods are being used, clarity and transparency over the intended approach will help to ensure credibility of the study.

7.8 Economic evaluation design specification

The overall resources available to deliver outcomes are always finite, so decisions need to be made about how to allocate them in order to achieve the maximum overall benefit. Consequently, whilst establishing which interventions are effective at achieving beneficial outcomes is a necessary condition for allocating scarce resources optimally, it is not on its own sufficient; it is also necessary to understand the costs and impacts of competing potential interventions, and to be able to compare them.

Economic evaluation can be conducted alongside or separate from other studies. There are pros and cons of each approach. Conducting after a study of the effectiveness of the intervention has been completed, for example, ensures that an economic evaluation is carried out on something that has already been shown to work, but does comprise another round, with consequent cost implications and time delays. Conversely, conducting an economic evaluation alongside a study assessing the effectiveness of an intervention will tend to mean a reduced overall data collection burden and alignment into a single study, but does create the risk that the study will be seeking to assess the cost effectiveness of something that is subsequently found to not even be effective (so inherently cannot be cost effective).

On balance, where an intervention has a reasonable likelihood of being found to be effective, the advantages of conducting economic evaluation alongside a level 2 study will probably outweigh the risks of wasted effort.

Some of the main types of economic evaluation are:

- **Cost minimisation**: Where two interventions are known to deliver the same outcome, a cost minimisation approach can be adopted to examine which delivers the outcome for lowest cost.

- **Cost-effectiveness analysis**: By comparing the costs of delivering an intervention to the amount of its outcome it achieves, a cost-effectiveness ratio can be derived, expressed in terms of the cost for each unit of the outcome.
- **Cost-benefit analysis**: By placing monetary values on the benefits as well as the costs, cost-benefit analyses report a ratio that is expressed in the same terms (e.g. £2 of benefits for every £1 of expenditure). This also enables different outcomes to be compared on the same scale, as they are all converted into the same units (money).
- **Cost-utility analysis**: Commonly used in health sectors, cost-utility analysis converts various outcomes to a measure of 'utility'. (In health sectors the common measure of utility is the Quality Adjusted Life Year, QALY.) As with cost-benefit analysis, this allows the comparison of different outcomes on a common scale, though in this case it is not a monetary scale.

8 Study protocol

8.1 Contents

Having undertaken the steps detailed above, producing a study protocol should largely be a process of drawing together the information previously documented. Once assembled, the protocol contains the information needed to conduct the study, which will be equally useful for the members of the study team and those in the future who want to replicate it or adopt elements from it. In completing all elements contained within the standard protocol items, those conducting studies will also be able to be confident that they have addressed the most important issues when designing their study.

8.2 Registration

Registration of a study is the act of lodging its protocol in an online registry where other users and producers of evidence can access it. Registration makes details of the intervention and study design available for others to learn from. Prospective registration – lodging the details as soon as plans for the study have been finalised and before the study has commenced – can deliver additional benefits, both for those registering their studies and the wider community.

For those registering their studies, prospective registration can increase the credibility of findings through up-front transparency. One issue that is known to affect studies, for example, is the potential to conduct subgroup analyses or trawl multiple measured outcomes after the results have been gathered, and produce reports that focus on the factors that show most impact. Where a study has not been prospectively registered it may not clear to users of the evidence whether the outcomes and the subgroups were pre-defined or investigated after the fact, which can reduce their confidence in the evidence.

For the wider community of evidence users, one of the main intended benefits of prospective registration is its ability to reduce both the incidence and severity of publication bias. Publication bias is the tendency for null findings (i.e., studies that indicate that an intervention was not effective) to

be less likely to be published; it results in an impoverished, and potentially misleading, evidence base.

Not only does prospective registration make publication bias less likely to occur, in some circumstance it might also decrease the impact of it when it does still happen. Although there is still the possibility that a study might not be reported, those conducting future systematic reviews will be more likely to identify that it was conducted, and therefore be more able to make contact in order to seek access to the findings.

9 Study conduct

9.1 Adherence to protocol

Although there is a high priority on adhering to the pre-defined protocol when conducting a study, there may in practice be reasons why exact adherence becomes impossible or inappropriate. In these instances it is important to document deviations (and the reasons behind them) to inform eventual reporting of the findings. Reasons requiring changes might include external information becoming available from other studies, internal financial difficulties, or a disappointing recruitment rate *[SOURCE: Moher et al 2010.]*.

Changing the primary outcome measure during the study should only be done in exceptional circumstances, and for reasons like it becoming impossible to collect the data. The primary outcome should **not** be changed just because the intervention seem to be ineffective for that outcome but appears more effective at doing something else.

9.2 Flow of participants

It is important to record the flow of participants through the study in order to be able to report on these details. Knowing the flow of participants is important for those reviewing the evidence in order to understand whether there are any potential sources of bias, such as might be introduced by differential drop-outs between an intervention and comparison arm.

The information collected should be sufficient to inform readers about the reasons for participants:

- not receiving the intervention they were allocated to,
- being lost to follow-up, or
- being excluded from the analysis.

Information about participants who were excluded after allocation is important because they are unlikely to be representative of all participants in the study. For example, participants may not be available for follow-up evaluation because they experienced a significant change in their

circumstances (improvement or worsening), which could be due to chance or be a result of the intervention they were receiving.

It is important to be able to distinguish between the various reasons for attrition, such as loss to follow up (which may be unavoidable), investigator-determined exclusion for such reasons as ineligibility, the participant choosing to withdraw from the intervention, and the participant not adhering well to the intervention as designed.

Information about whether all participants who were allocated to an arm of the study were included in the analysis, in the arms to which they were originally allocated (intention-to-treat analysis), is of particular importance. Erroneous conclusions can be reached if participants are excluded from analysis, and imbalances in such omissions between arms may be especially indicative of bias.

Knowing the number of participants who did not receive the intervention as allocated or did not complete the intervention permits the reader to assess the extent to which the estimated effectiveness of the intervention might be underestimated in comparison with ideal circumstances.

[ADAPTED FROM: Moher et al 2010, Item 13a Explanation.]

9.3 Adverse events

Adverse events are not necessarily adverse outcomes: adverse events can occur in comparison arms as well as interventions arms, even where there is a 'no treatment' comparison. Sometimes bad things just happen by chance. The reason for recording adverse events it to support analysis of whether there are more harms happening in an intervention arm than chance would explain, or if certain types of harm are more common.

Establishing if there are adverse events associated with an intervention allows a more holistic assessment of the intervention; evidence users will be able to take into consideration it harms as well as its benefits.

10 Findings and other study outputs

10.1 General

Recording, reporting and disseminating the details of any evidence being produced is an important part of contributing to the wider knowledge of the field.

Where there is commercial sensitivity relating to the success of interventions, organisations should still seek to adhere to all aspects of the documentation of studies, except for steps that result in making the documentation public. This will ensure that internal users of the evidence receive it in the same comprehensive format of other evidence being produced in conformity with the standard, and will also make it easy to publish findings later should they be 'declassified' if the information is felt to no longer be commercially sensitive.

The standard specifies that at least two reports of findings are to be prepared: a main report and a summary report. This is intended as a minimum, and other reports or means of publishing the findings may be used where these are helpful in communicating to wider audiences. These further reports are expected to link back to the main report, to facilitate those readers who wish to access more of the detail. Whilst they might contain less detail or be framed in more casual language than the main report, care should still be taken to ensure that any claims or indications they make are compatible with the evidence generated by the study. Particular care is required to ensure that a level 1 study avoids the use of language that implies a causal finding, irrespective of tone or audience.

10.2 Structured reports of findings

10.2.1 Main report

The intention of specifying a standardised structure for reporting findings is that it should assist in the preparation of an output that is thorough

(providing all the information evidence users and other potential readers could need to understand the study), unambiguous, and in a consistent well-organised format. In doing so, it will support the evidence produced to be useful and used in the future. Not only will a thorough and structured report make it easier for others to make use of the evidence, it is also likely to be useful to evidence producers themselves, when they have to refer back to it after a period of time, or to make use of it when making a case to future partners or stakeholders.

Capturing the standard information specified in the report format will benefit both evidence producers and evidence users. Evidence users will benefit from a clear understanding of how a study was conducted. Evidence creators will benefit from the support of checking that they have addressed all relevant issues.

Indication of conformity

By conforming to the standard in producing evidence, certain important requirements will have been met. Including an indication of conformity to the standard on the report will help to support the credibility of the findings. The standard specifies that such indications of conformity shall include the level of the study conducted; prominently displaying this provides readers with a simple summary of the approach that they can expect to find taken in the study when they read the report.

Intervention specification

All of the details on the design and conduct of the study are important for evidence users to gain a full understanding of the study and to build confidence in its findings. But even amongst these, the intervention specification stands out as being of particular importance: even if evidence users are willing to take the results and conclusions of a study at face value, without fully reviewing the details of the study, they will find it impossible to implement an intervention that is reported to be effective if they do not know what that intervention was.

Results: primacy of primary outcome results

It is important to stick to the primary outcome as specified before the study commenced, and to treat it as of primary importance throughout the process, including in reporting.

Given that the study will have been designed with a focusing on finding evidence in relation to the primary outcomes, any findings on secondary outcomes should be treated somewhat more tentatively. A major reason for this is the problem of multiple comparisons: if 20 other possible benefits were measured as secondary outcomes alongside the primary outcome, then even if the intervention does nothing you might find that there are statistically significant differences between the intervention and comparison arms on one or two of the secondary outcomes, just by chance.

Statistical techniques to control for problems of multiple comparisons, for example Bonferroni correction, which effectively sets a higher bar for a finding being deemed statistically significant. These can reduce the chances of a statistical artefact being incorrectly reported as a robust finding, but are not a substitute for retaining primary outcome as specified in advance.

Alternative timepoints for the primary outcome measure should be treated in the reporting the same as secondary outcomes.

Adverse events

The main report should report on harms and adverse events observed during the study. These might include unexpected negative consequences or side effects that were predicted during the planning stages of the study. Where the risk of potential side effects was foreseen, these should normally have been measured thoroughly to facilitate a holistic consideration, balancing harms against benefits.

Exploratory findings

Once data have been collected, it can be tempting to want to analyse them extensively to identify any associations that can be discerned. These sorts of exploratory findings, where various potential associations can be examined until a correlation is discovered, are particular susceptible to the problems of multiple comparison; any findings are fairly likely to be due to chance. Consequently, such findings should be presented with a prominent warning about their limitations. The findings may merit further study, if they identify an association that would be of practical importance if it were robustly demonstrated to persist.

Conclusions

Reporting should be designed such that it facilitates the answering of the 'so what?' question, i.e., to allow potential users of the evidence to easily identify what the implications are for practice or commissioning.

The conclusions presented in the report must be reasonable and supported by findings. The most robust findings will relate to the primary outcome, and this should normally be the focus of the conclusions. Conclusions around effectiveness of the intervention at achieving secondary outcomes should normally be much more tentative.

The conclusions should also be suited to the level of study they relate to. A level 1 study, for example, might conclude that an intervention is promising if an association is found and the process evaluation identifies good prospects for evaluation feasibility and implementation feasibility.

Checklists

Formats have been devised (principally in the medical health research sector) that ensure that all relevant information is captured. These relevant guidelines should be followed when compiling the main report to address each aspect of the evidence.

These guidelines are in general relatively broadly drawn, and most elements will be directly applicable in any relevant evidence reporting context. However, as they have mostly been developed for a health evidence context they do occasionally include some health-specific references, for example:

> *"Describe the health or health service problem that the intervention is intended to address…"*
>
> *[SOURCE: Zwarenstein et al. 2008.]*

In these cases they should be treated as if a more generic reference were included, such that the above quote might be read as simply:

> *"Describe the problem that the intervention is intended to address…"*

Completing checklists fully normally requires noting where in the structured report the information can be found, not just a tick to confirm that it is in there.

10.2.2 Summary report

As a summary of the main report, the relevant explanations behind those elements' purpose are equally applicable in relation to the summary report.

The summary report is specified to be written in 'plain English'. This is an important requirement for making the findings of studies accessible more widely.

10.3 Lodging reports in repository

Another important feature of the standard is to outline processes that will support the communication of evidence. This is designed to ensure that evidence is not only produced, but that it is accessible and usable, to maximise the impact it has on future practice.

Lodging reports of studies in a repository allows those who might want to directly use the results to do so. It is also necessary for future use of the evidence in a systematic review or meta-analysis. Lodging evidence further increases the chance that the evidence will be used, by increasing the likelihood that it will be found and included in these types of studies.

Failing to make evidence available and accessible always results in a smaller evidence base for potential evidence users. However, the biggest concerns occur when there is bias in relation to which evidence is made available and which is not, known as publication bias; in these circumstances, the evidence base is not just reduced but actually distorted. The most common concern is that publication bias tends to result in positive findings being more likely to be published than null findings. In the extreme case, this could result in the same ineffective intervention being studied many times; by chance, one of these might reach an incorrect result that the intervention was effective, and it might only be this study that gets published. Even in the absence of publication bias, it means that some evidence users are forced to make decisions based on less evidence than has been produced. It can also result in interventions being re-studied when ample evidence has already been produced establishing their effectiveness or ineffectiveness.

Consequently, it is important that wherever possible evidence is made available and accessible, whether the findings show an intervention to be effective, ineffective or are inconclusive.

For evidence producers, one further motivation for making their evidence widely available is the potential for reciprocity: others within their sector will see that it is established practice to lodge reports and will be more likely to do so with their evidence in return, making more evidence available for all potential users.

10.4 Publication of data and analysis

For most users of a piece of evidence the main thing they will be interested in will be the results – was this intervention effective – possibly combined with any other lessons for practice. Others, however, will be interested in digging deeper into information that was produced as part of the process. One particularly valuable use of data is the ability to use statistical techniques to combine information from multiple studies to strengthen the findings, known as meta-analysis. It is recommended that data should be published in a publicly accessible repository; where the data relate to people they must be suitably anonymised before publication. This will also make it possible for others to check your work, improving the credibility of the findings.

The credibility of findings can be further enhanced through the publication of the analysis alongside the data. This might take the form of a spreadsheet file in which the calculations were performed or programming code if the analysis was undertaken using an advanced statistical package such as R, Stata or SPSS.

10.5 Open access publishing

The publication of findings in an academic journal is not considered to be the primary output of a study following the process in the standard. It may, however, be a secondary output, and one that researchers undertaking evidence production might be interested in pursuing, in order to communicate findings to an academic audience.

Requirements for open access publishing are becoming more common to be specified anyway, so academics becoming more aware of it. In UK, there are requirements attached both to REF (the process by which universities are ranked on their research) and Research Council funding. Consequently, many academics will be happy to comply.

Annex A (normative) Systematic identification, reviewing and analysis of multiple causal studies

The principal focus of the scope of the standard is producing new primary studies. It would be remiss to not address systematic reviews at all, as they form an important part of a hierarchy of evidence of the effectiveness of interventions. Consequently, Annex A to StEv2-1 provides an overview of requirements and recommendations to undertake a systematic review.

The purpose of evidence synthesis is to bring together the best available evidence on a topic, draw conclusions from multiple studies and identify and remaining uncertainties or gaps in knowledge. A systematic review should address a clearly formulated question, use systematic and explicit methods to identify, select, and critically appraise relevant research, and collect and analyse data from the studies that are included in the review *[REFERENCE: Moher et al, 2009].* Thorough reviews will make particular efforts to find information on studies that have not been published, to lower the risks associated with positive reports being disproportionately published. Depending on the quantity of evidence relating to the particular question, a systematic review can be a very substantial undertaking; it does, however, have the potential to create an evidence base that is more compelling than the sum of its parts.

A systematic approach to reviewing the evidence on a topic requires planning and documenting the approach in advance. This is called the protocol for the review.

Where appropriate, meta-analysis (i.e., statistical analysis of data from multiple studies) should be conducted as part of the systematic review.

A realist synthesis approach to systematic reviews could also be considered where relevant. Realist synthesis aims to take a context-sensitive approach to systematic reviewing, building up a theory of what works, for whom and in what circumstances, through the analysis of previous research.

Annex B (normative) Circumstances where randomised controlled trials may be unsuitable

Whilst RCTs produce some of the most robust evidence available on the effectiveness of interventions, there are circumstances where they should not be used. The list provided in Annex B to StEv2-1 attempts to outline those circumstances.

One of the most common concerns about RCTs amongst those less familiar with the design is that randomising interventions is unethical. In general, this is not the case: where there is equipoise (i.e., uncertainty over which course of action is most effective) it can be argued that random allocation is one of the fairest ways of distributing the benefits, which may accrue to those in the intervention arm, the control arm, or may not be present at all.

As noted in the list, however, there are specific ethical concerns where an intervention is known to be more effective than the alternative course of action. Even these concerns may not completely rule out the potential for an RCT to be ethically conducted. If there is a natural delay, meaning that it takes time to roll out the effective intervention so some people will receive it before others in any event, it may be possible to design an ethical study.

Annex C (informative) Quasi-experimental methods

Annex C to StEv2-1 outlines a selection of alternative designs that may be suitable for use in level 2 studies where a randomised controlled trial is unsuitable. The technical details of these designs are substantial. Interested readers are directed to Shadish, Cook and Campbell (2002) for further information.

Bibliography

Boddy, J., Neumann, T., Jennings, S., Morrow, V., Alderson, P., Rees, R. and Gibson, W. n.d. The Research Ethics Guidebook. Institute of Education, University of London. London. (WWW resource)
http://www.ethicsguidebook.ac.uk/
Legal requirements page: http://www.ethicsguidebook.ac.uk/Legal-requirements-76

Chalmers, I., Nylenna, M., 2014. A new network to promote evidence-based research. The Lancet 384, 1903–1904. doi:10.1016/S0140-6736(14)62252-2

Collett, C., 2014. Seeking informed consent for simple and efficient trials in the NHS. Draft guidance: For comment. Health Research Authority. http://www.hra.nhs.uk/documents/2014/10/seeking-informed-consent-simple-efficient-trials-nhs-draft-guidance-comment.pdf

Epstein A E, Hallstrom AP, Rogers W J, Liebson P R, Seals A A, Anderson J L, Cohen J D, Capone R J, Wyse D G. 1993. Mortality Following Ventricular Arrhythmia Suppression by Encainide, Flecainide, and Moricizine After Myocardial Infarction: The Original Design Concept of the Cardiac Arrhythmia Suppression Trial (CAST). JAMA. http://jama.jamanetwork.com/article.aspx?articleid=409358

Greenhalgh, T. (1997) How to read a paper. BMJ 1997;315:246. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127173

Lewin, S., Glenton, C., Oxman, A.D., 2009. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ 339, b3496. doi:10.1136/bmj.b3496

Lilienfeld, S.O., Sauvigné, K.C., Lynn, S.J., Cautin, R.L., Latzman, R.D., Waldman, I.D., 2015. Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. Front. Psychol. 1100. doi:10.3389/fpsyg.2015.01100

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6. doi:10.1371/journal.pmed.1000097

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G., 2010. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 340, c869. doi:10.1136/bmj.c869

National Research Ethics Service, n.d. Does my project require review by a Research Ethics Committee? http://www.hra.nhs.uk/documents/2013/09/does-my-project-require-rec-review.pdf

Rogers, P., n.d. Develop Programme Theory [WWW Document]. Better Evaluation. URL http://betterevaluation.org/plan/define/develop_logic_model (accessed 10.13.15).

Schwartz, D., Lellouch, J., 2009. Explanatory and Pragmatic Attitudes in Therapeutical Trials. Journal of Clinical Epidemiology 62, 499–505. doi:10.1016/j.jclinepi.2009.01.012

Shadish, W.R., Jr, Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs for Generalised Causal Inference. Wadsworth, Cengage Learning, Belmont, CA.

Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2012) Quality in qualitative evaluation: a framework for assessing research evidence (supplementary Magenta Book guidance). HM Treasury, London. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190986/Magenta_Book_quality_in_qualitative_evaluation__QQE_.pdf

Ueffing, E., Tugwell, P., Welch, V., Petticrew, M., Kristjansson, E., 2012. Equity Checklist for Systematic Review Authors - Version 2012-10-02. http://equity.cochrane.org/sites/equity.cochrane.org/files/uploads/EquityChecklist2012.pdf

Vine, J., 2016, General Requirements for Evidence – Part 1: Vocabulary. HACT. London, UK.

Vine, J., 2016a, StEv2-1 (Standard for Producing Evidence – Effectiveness of Interventions – Part 1: Specification. HACT. London, UK.

White, H., 2015. Understanding what's what: the importance of sector knowledge in causal chain analysis. Evidence Matters.

http://blogs.3ieimpact.org/understanding-whats-what-the-importance-of-sector-knowledge-in-causal-chain-analysis/

Zwarenstein, M., Treweek, S., Gagnier, J.J., Altman, D.G., Tunis, S., Haynes, B., Oxman, A.D., Moher, D., 2008. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 337, a2390. doi:10.1136/bmj.a2390

HACT

ISBN 978-1-911056-02-7

9 781911 056027 >

Non-fiction £20.00