LEARNING FROM RESEARCH: SYSTEMATIC REVIEWS FOR INFORMING POLICY DECISIONS

A QUICK GUIDE

David Gough, Sandy Oliver and James Thomas
EPPI-Centre, Social Science Research Unit, Institute of Education
University of London

December 2013

The Alliance champions the use of evidence in social policy and practice. We are an open–access network of 1,600 individuals from across government, universities, charities, business and local authorities in the UK and internationally. The Alliance provides a focal point for advancing the evidence agenda, developing a collective voice, whilst aiding collaboration and knowledge sharing, through debate and discussion. We are funded by the BIG Lottery Fund, the Economic and Social Research Council and Nesta. Membership is free. To sign up please visit: www.alliance4usefulevidence.org

Please cite as: Gough D, Oliver S, Thomas J (2013) *Learning from Research: Systematic Reviews for Informing Policy Decisions: A Quick Guide.* A paper for the Alliance for Useful Evidence. London: Nesta.

CONTENTS

PREFACE

"Instead of just mooching through the research literature, consciously or unconsciously picking out papers here and there that support [our] pre-existing beliefs, [we] take a scientific, systematic approach to the very process of looking for scientific evidence, ensuring that [our] evidence is as complete and representative as possible of all the research that has ever been done." (Ben Goldacre 2012)

Systematic reviews provide a vital tool for policy makers and practitioners to find what works, how it works – and what might do harm. Whether it is policing, teaching, social care or any other area of social policy, reviews can inform us about what is known, how it is known, how this varies across studies. It also tells us what is not known from previous research.

It now seems extraordinary that only a few decades ago it was acceptable practice to review research by ad hoc literature reviews or panels of experts. Summaries of research were not done in a systematic way and all sorts of biases crept in: reviewers did not attempt to identify all the relevant research, check that it was reliable, or write up their results in an accountable manner. But over the last forty years, we have a more rigorous systematic approach has been refined to review what is known and not known, as described so well by Ben Goldacre above.

But let's also dispel a myth about systematic reviews. These are not just technical and statistical exercises appropriate for medicine and health. They include qualitative research and are highly relevant across a wide range of social policy areas. This guide succinctly outlines all these different approaches and will support those thinking of commissioning a review of research to take the best course of action.

Even though the idea is simple – bringing together all the best available research – the practice is fraught with challenges. This guide will help anybody navigate those challenges. Systematic reviews have made a vital contribution to the history of science. They inform decisions that affect people's lives in many domains, from the way we teach our children to preventing crime. We thus have a responsibility to make sure reviews are done well. This guide will help us do that.

Dr. David Halpern
What Works National Adviser, Cabinet Office

INTRODUCTION

We would like to make the best decisions possible in all aspects of living, from our work to our private lives. Findings from research, considered alongside our values and resources, can help us. One might argue that in public policy and practice we have an obligation to use relevant research findings; and that it would be unethical to make decisions that affect other people without being informed. Research evidence is, of course, not the only thing that could or should inform decision making, but if it is available we should at least make good use of it.

Although simple in principle, finding relevant, reliable research can be difficult in practice. Research about the same issue can vary considerably in its assumptions and methods. It is not unusual to find different pieces of research with contradictory findings. This may be because of chance variation, differences in context, mistakes in some of the research, or because researchers have different world views and thus understand issues in different ways.

Systematic reviews can help address these problems. They provide a meticulous way of finding relevant, high quality studies; and integrating their findings to give a clearer and more comprehensive picture than any single study can produce. Moreover, they are based on a clearly stated set of assumptions. Readers can base their judgments on a coherent set of studies encapsulated within a single report

Systematic reviews enable us to establish not only what is known from research; but also what is not known. They can inform decisions about what further research might be best undertaken, thereby creating a virtuous cycle. They enable researchers, policymakers and practitioners to answer key questions: '*what do we know, how do we know it?*' and '*what more do we want to know and how can we know it?*'

This guide is a short introduction for decision-makers and researchers or anyone else considering whether a systematic review may be appropriate to fill a gap in knowledge or to use as a resource. It will help anybody planning on commissioning a review of what research is already out there. You may be an analyst, evaluator, policymaker or commissioner. The guide is aimed at anybody from central government, local authorities, public service providers, regulatory and advisory bodies, charities or the consultancy sector. We also hope it will inform staff engaged in developing a 'What Works' centre; synthesizing research is a core principle of the new independent 'What Works' institutions (Cabinet Office, 2013).

For those decision-makers considering commissioning a review of research, Parts One and Four will be most useful: Part One of this guide describes the logic of systematic reviews and how they fit within the broad range of research. Part Four highlights some of the key issues to consider when commissioning or using a review and some of the infrastructure organizations that have been developed to support the use of research. Parts Two and Three are more technical and will be useful for researchers and others engaged in making a review successful: Part Two describes the main stages of undertaking a systematic review. Part Three explains the wide variation that exists in approaches to reviewing and how to choose between them. More detailed information on review methods can be found in a range other publications (for example, Gough et al. 2012b, 2012c)[1].

---

[1] Please also see the EPPI-Centre website: http://eppi.ioe.ac.uk Other detailed guides for undertaking systematic reviews of the impact of health intervention studies have been produced by The Centre for Reviews and Dissemination, York University (http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf) and the Cochrane

1 THE NATURE AND LOGIC OF SYSTEMATIC REVIEWS?

Reviews of existing research can provide a valuable addition to the evidence base and enable policy makers and practitioners to access that evidence. Doing a review systematically and reporting it clearly allows readers to judge how relevant and reliable it is for their own purposes; just as they may judge individual studies. Systematic reviews can be undertaken for all types of research questions and study designs.

What are we trying to review? Research and how it varies

Research is necessarily varied. Study designs have to accommodate: different disciplines, from social policy, to botany and materials science; different ideological and theoretical assumptions; different questions; and different methods. Studies that address the same research questions may be underpinned by very different theoretical assumptions and may therefore require very different methods. Study methods are driven by the nature of the research question being posed:

Question	Method
What is this phenomenon?	Describe analytically (e.g. to develop a classification system or theory)
What is its meaning or value?	Describe and develop meaning (e.g. an ethnographic study of people's experiences)
How many are there?	Describe and measure (e.g. a quantitative survey)
How do these phenomena compare?	Describe and compare (e.g. interviews or focus groups)
How are these phenomena related?	Describe associations and causal links (e.g. a randomized controlled trial comparing the impact of two different interventions)

Collaboration (http://www.cochrane.org/training/cochrane-handbook). The Campbell Collaboration has prepared guidance on undertaking systematic reviews of studies of the impact in social policy (http://www.campbellcollaboration.org/resources/research/the_production.php)

To put it simply, the first two questions are 'qualitative'; the latter are more 'quantitative'. However, this distinction is more complex (see Box 1). We prefer to describe the first two questions as deserving a 'configurative' approach, the latter as 'aggregating'.

These are important distinctions for anybody commissioning a review and we go into more detail in Part Three. But essentially, configurative approaches invoke studies that seek to develop and explore theory by *configuring* ways of understanding phenomena and their meaning or value to people; while the latter questions invoke studies that seeks to test theory about the presence, distinctions or associations of phenomena by *aggregating* observations within that particular theoretical framework (Box 1). If, instead of collecting data from the field, data are collected from relevant literatures, the first two questions invoke literature reviews that *configure* ways of understanding from different studies to shape a body of literature; and the latter questions invoke reviews that *aggregate* the findings of similar studies to judge the strength of a conclusion *(Gough et al. 2012b, c, Sandelowski et al. 2012)*.

> ## Box 1: Qualitative and quantitative methods or configuring and aggregating studies
>
> The two dominant paradigms in social research have been portrayed as quantitative and qualitative. However, these terms are imprecise. Qualitative research is predominantly concerned with the exploration, **configuration** and development of meaning and theory, although some of the methods employed for these purposes may be quantitative! Quantitative research is predominantly concerned with **aggregation of** information and testing hypotheses, although some of the methods employed for these purposes may be qualitative! There are often tensions between the two paradigms (Oakley 2000), some of which relate back to nineteenth century debates about the role of science and the humanities. Some quantitative researchers are anxious about the seeming lack of rigour and potential bias of qualitative research; while some qualitative researchers are anxious about the lack of critical focus on theory and dangers of 'empty empiricism' in some quantitative research. Although the methods employed in reviewing literature may be qualitative or quantitative, different ways of approaching the task of reviewing are more accurately described as configuring and aggregating of data in studies. (Gough et al. 2012a, Sandelowski et al. 2012). See Part 3 for more discussion on these terms.

In addition to variation in research questions driving the use of different research methods there can be many further assumptions of how research is formulated and undertaken. Those with differing ideological and theoretical assumptions will understand and thus study issues differently. So in any one topic area there can be more than one systematic review representing these different perspectives. It is also possible to undertake reviews specifically to analyse and understand such variation in perspectives.

Integrating what we know: the logic of systematic reviews

Research can be used to inform decisions made by policymakers, practitioners or service users. Knowledge of previous research can also inform the planning of new research by academics and the funders of research through analysing what is and isn't known (a gap analysis).

However, finding relevant, reliable research for decision-making can be problematic due to:

- Quantity of research: we live in an age where there is an enormous amount of research undertaken. It is difficult even for full time academics to keep up with research in their field.
- Lack of research: even if there is considerable research on a topic, the focus and method and quality of this research may not be relevant or sufficient for your particular review question.
- Physical access: research is published in so many different forms, many of which require a subscription to access.
- Intellectual access: many research studies require technical knowledge to interpret and understand.
- Reviewing skills: understanding the totality of the research requires the skills (and means) to interpret primary studies, organize them and synthesize their findings.

For many years, literature reviews have provided decision-makers with information about what is known and not known. However, relatively little time and money has been invested in literature reviews compared to primary research. In other words, society has invested considerably in research but less in bringing together what is known from that research. Consequently, reviews have traditionally been idiosyncratic, simply summarising research that the reviewer had previous knowledge of or could easily access. As systematic reviews were not expected to use and report explicit, systematic methods to identify and synthesise the findings of primary research, it was not possible to know how well a review had been carried out, what were the parameters and boundaries of the evidence being presented, and whether the findings could be relied upon.

Existing research can also be reviewed by academic experts, either individually or in groups ('*expert panels*'). Experts have many specialist skills; but relying on expert opinion can be risky if that opinion is not supplemented by a systematic review of the research. The authority or reputation of experts may obscure their ideological and theoretical assumptions, the boundaries of their knowledge (and consistency of depth of knowledge within those boundaries), and possible flaws in their methods of synthesizing knowledge. However, a great deal of decision-making in government is still guided by experts rather than systematic reviews of evidence.

Research may also inform decisions due to chance, fashion and financial or other vested interests. Individual academics are likely to believe in their own work and actively promote their findings; while policymakers may seek out research that supports prior policy decisions (Weiss 1979).

Systematic reviews emphasize the need for the same degree of rigour and clarity of method that is expected of primary research. In the words of Ben Goldacre, when conducting systematic reviews, "instead of just mooching through the research literature, consciously or unconsciously picking out papers here and there that support [our] pre-existing beliefs, [we] take a scientific, systematic approach to the very process of looking for scientific evidence, ensuring that [our] evidence is as complete and representative as possible of all the research that has ever been done" (Goldacre 2012: Kindle Locations 174-176).

A brief history of systematic reviews

There is a long history of using research evidence to support public policy decisions. However, systematic methods for appraising and collating evidence have only developed over the last

forty years. Critical appraisal and synthesis of research findings in a systematic manner first emerged in 1975 under the term 'meta analysis'. The phrase was coined by Glass who conducted syntheses in the areas of psychotherapy and school class size (for example, Glass and Smith 1979).

Although these early syntheses were conducted in the fields of public and social policy, systematic research synthesis grew most rapidly in the fields of medicine and health.

Archie Cochrane's seminal text 'Effectiveness and efficiency' (1972) urged health practitioners to practice 'evidence based medicine', later defined by David Sackett as 'the conscientious, explicit, judicious use of current best evidence in making decisions about the care of individual patients.' (Sackett et al. 1996). In 1992 the Cochrane Collaboration[2] was formed to support effectiveness reviews in healthcare and manage the knowledge generated. A sister organisation, the Campbell Collaboration[3], was later formed to support effectiveness reviews in wider areas of public and social policy. Despite the success of these umbrella organizations they represent only a minority of all systematic reviews undertaken.

The development of systematic methods to answer questions other than the level of impact of interventions and thus to synthesize the findings from other forms of research has been slower. More recent developments have taken place in meta ethnography (Noblitt and Hare 1988), mixed methods reviews and reviews addressing other fields such as management, conservation, international development, software engineering and economics.

Despite increasing awareness of systematic reviews, there are still many myths. Some believe that reviews are only relevant to quantitative research and are thus of no interest to qualitative research or social policy. Some see systematic reviews as purely technical exercises. They fail to recognize the research skill and creativity required or the new scientific discoveries they make; and so they are afforded insufficient resources or academic credit.

There is also much confusion created by different uses of terminology. A systematic review is using systematic methods for reviewing literature. It can be considered a 'meta level' (about or beyond) process as it is research on research. The term 'meta analysis' could therefore be used to describe all systematic reviews but in practice the term is used only for statistical meta analysis where statistics are used to synthesize the numerical data from different studies. There is also the term 'meta evaluation' (Scriven 1969) which can include the synthesis of evaluation studies but also has broader meanings (Scriven 1969, Gough et. al. 2012b).

Questions and answers

Q 1.1: What is the difference between a systematic and non systematic review of the research literature?

A 1.1: The only difference is that a systematic review has an explicit rigorous methodology. This makes the results of the review accountable and open to criticism and debate.

Q 1.2: Is a systematic review a piece of research?

A 1.2: Yes it is a form of secondary research. It uses rigorous methods to bring together the results of individual primary studies. It is the first thing to do before undertaking any new primary research in order to: (i) check that new research is necessary and not simply duplicating prior research; and (ii) that the new research is informed by the lessons of previous research.

Q 1.2: Why might a non researcher want to read or commission a systematic review?

A 1 2: In order to know what is known from research about an issue. Other factors may also influence the decision that is made and so research findings are just one form of information to inform decision making.

Q 1.3: Are not all systematic reviews about randomized controlled trials (RCTs) to assess the impact of programmes and these are an outdated focus on positivism which has been long discredited?

A 1.3: Reviews can answer all types of research question and causal questions are not the only, or even necessarily the most important, questions that reviews can address. However, if the review question is on the extent of impact of a programme then you do need to take seriously possible counterfactuals (other explanations for causal attribution) and RCTs are particularly good for dealing with this type of problem.

2 THE MAIN STAGES OF A SYSTEMATIC REVIEW

There are many types of systematic review. But they all share the same basic logic and stages.

Stage	Description
Need	• Who is asking the question and what will they do with the answers.
Review question	• What will the review do and what are its underlying assumptions.
Scope	• Establishing the criteria used to select studies.
Search	• Designing and running a search strategy to find promising sources.
Screen	• Checking which studies are relevant and discarding those that are not.
Code	• Collecting information for mapping, quality assurance, and on study findings.
Map	• Describing the nature of the research field defined by the inclusion criteria.
Appraise	• Judging the relevance, utility and quality of the studies.
Synthesise	• Bringing together the findings of different studies to answer the review question.
Communicate	• Describing how the review was done, its findings and implications for future decisions.

The need for a review

An early step in any review is establishing whose skills and expertise are required. More specifically, reviewers need to ask:

- Whose questions and perspectives are important?
- Will they be involved as individuals or will they be representing a wider group? If so, what does such representation mean in practice?
- At what stage(s) of the review process should they be involved? Should they be involved only in establishing the review question; or in other stages of the review too?
- Are they to be consulted or offered a share in making decisions about the review?
- How will those involved be engaged?

Involving relevant experts, decision makers and service users in a systematic review can broaden the range of voices when setting the research agenda. As well as having democratic benefits, this may also make the research more useful; for example, research on health services that involves health service users may help inform the development of services better suited to the users. Guidance about user involvement in reviews is available from organizations that support systematic review teams (Green et al. 2011; Rees and Oliver 2007, 2012; Campbell Collaboration International Development Coordinating Group 2012). This issue is also discussed at the end of the next section on Review questions and conceptual frameworks

Review questions and conceptual frameworks

The review *question* gives each review its particular structure and drives key decisions about: what types of studies to include; where to look for them; how to assess their quality; and how to pool their findings. Although the question may appear to be simple, it will include many assumptions that need to be clarified. Each concept within the question has to be carefully defined, as this will affect which studies are included or excluded. Without clear definitions, decisions about inclusion or exclusion will not be consistent. Thus a review on the effects of homework on children would require clarity of what was meant by both '*children*' and '*homework*', and also what '*effects*' were to be considered. The importance of clear definitions is illustrated by a comparison of six reviews about the same general topic of older people and accidents. The reviews together considered several hundred studies; but only one study was common to all six reviews. This was because each review defined 'older people' and 'accidents' differently; and each judged the effects of different interventions in terms of different outcomes (Oliver et al. 1999). Each review provided different conclusions to reflect their focus[4].

Research questions should be developed with the review method and resources available in mind. Reviews can be relatively broad or narrow and relatively shallow or deep. If the review question is very broad but resources are limited, the review may be relatively shallow. This may provide only an initial impression of the research evidence.

The review must also explicitly state any particular theory or ideology underpinning the research question. This might include a logic model or theory of change to be tested by the review. Such approaches are essential when complicated (multi-component) and complex (containing intricate interactions) interventions are synthesized (Rogers 2008). In these situations, the *programme theory*, which details the various pathways through which an intervention is thought to have an effect, is documented and often represented in graphical form in a flow diagram (called a 'logic model') (Anderson et al. 2011). A review may examine one small part of a programme theory or may be more ambitious in attempting to examine the complexity of many components of a theory.

Since reviewers need to ensure that the review will meet the needs of those commissioning the review, it can be helpful (if not essential!), to include diverse opinions in the review design process, which can help clarify the priorities and conceptualization of the review. The reviewers may benefit from listening to others if they are not subject specialists. If they are subject specialists, listening to others may help them avoid overlooking some critical aspects because of the implicit assumptions they hold. Likely users of a review can be particularly helpful in offering relevant insights into the questions being addressed. Involving people living in disadvantaged circumstances can be particularly challenging.

---

[4] for another example, see Woodman et al. 2012

Such efforts can be complemented by analysing studies in a review in terms of equity (Konnerup and Sowden 2008; Oliver et al. 2012 pp69-70; Ueffing et al. 2009). This raises broader issues about the extent that a review question is concerned with different groups of people (sub-populations) and contexts and the extent that these issues are specified in the review question.

Inclusion criteria

The review question and conceptual framework help determine what type(s) of studies to include in the review. In simple cases, this can be all studies that involve all of the different concepts within the review question. A review on the effects of homework on children would include all studies on: 'children', 'homework' and 'effects', however these were defined. However, the inclusion criteria may also specify the type of research method(s) to include, or a more specific type of homework to examine. As this review is concerned with the *effects* of homework, it may only include studies that examine effects. In some reviews, the methods may be defined quite tightly; for example reviews of randomized controlled trials (RCTs) only. In others, a wider or different range of research designs may be included.

Inclusion criteria may also specify the countries where the study has taken place or the language in which the study is reported. The countries specified may be limited because of the focus of the review question; for example, a review of studies on the effects of homework on children in Australia or in nations with a particular level of economic development. Language restrictions may be more for pragmatic, i.e. resource, rather than methodologically justified reasons.

The date a study was undertaken may also be a criterion for inclusion. For example, a review of homework in Australia may only consider studies since a pertinent change in Australian schools policy. Including only recent studies may be a practical way of managing the workload; though this will provide a less comprehensive synthesis.

Whatever restrictions are imposed on the review, it is important that they are coherent and consistent. For example, it is difficult to justify a date inclusion criterion based on a change in policy in one country if studies from other countries are included in a review. Any restrictions by geography, language or time should be stated explicitly in the review, preferably in the title. Direct explicit restrictions on language may not be common in reviews but they often occur indirectly by the type of sources used to search for studies and are therefore hidden rather than transparent criteria.

Search strategy

The inclusion criteria for studies determine the strategy used to search for potentially relevant studies for a review.

Bibliographic databases of published journal articles are a common place to search for relevant studies, providing details such as the study title, publication source and abstract. These databases usually cover a particular academic discipline and/or topic area. As databases may not be comprehensive and research questions do not always precisely match existing academic disciplines or topics it may be necessary to search several databases (e.g. Brunton et al. 2012; Stansfield et al. 2012).

Databases can be *'free text'* searched for any words that occur in the database records for each publication. A review on the effects of homework on children, would involve a *'free text'* search for the co-occurrence of the terms '*children*', '*homework*', and *'effects'* (i.e. the terms: *children* AND *homework* AND *effects* would need to appear in the document text). To ensure the search is as comprehensive as possible, different versions of the terms should also be used, for example searching for '*child*' as well as '*children*', or '*impact*' as well as '*effects*'. Databases often have the facility to search for similar words (known as 'stemming' and 'synonym searching'), using functions such as *'child*'* to identify all words where *'child'* is the root.

The choice of bibliographic databases is critical as it determines what can be found in the search. Normally several databases would be searched to cover the concepts within the review question. It can be important to include UK databases to increase the likelihood of identifying UK studies and particularly those studies that are not formally published (grey literature) which are often less well represented in large international databases (Grayson and Gomersall 2003).

Bibliographic databases usually index records according to their topic using *'keywords'* or *'controlled terms'*. A comprehensive search strategy usually involves searching both free and controlled terms. The advantage of controlled terms is that some of the work has already been undertaken to organize the papers. The disadvantages are that the terms may not have been applied in the way best suited to the review, or they may not have been applied consistently to all the records in the database. However a database is searched, the review report should set out which databases were searched, the search terms used (*'search string'*) and the date(s) of searching.

Searches of the internet can be broader as they are not restricted by the limits of a particular database; however, they are likely to generate a lot of non-relevant material. Some search engines, such as Google Scholar, are restricted to academic work, but they could miss the large number of research papers that are not formally published (sometimes referred to as *'grey literature'*). The other disadvantage of internet searches is that it is more difficult to record the search method in a way that others can reproduce. Despite this, internet searches can provide a useful double check of a more specific search strategy and identify important grey literature that cannot be found in established databases. A more focused approach is to search specific websites that are likely to include relevant studies. These might include the websites of specific research centres, research associations or government departments.

A search strategy may also involve 'hand searching' journals or books, particularly those journals not included ('*indexed*') in the main bibliographic databases and for research topics where primary research is commonly reported in books. Examining the reference lists of relevant studies can also lead reviewers to other relevant studies – and citation indexes such as Google Scholar and Web of Science can tell reviewers which papers have cited documents they have selected as relevant. In turn their reference lists may then lead to even further studies. This 'snowballing' technique can be highly productive, but may restrict the search to a selective group of studies that reference each other. Finally, it is also worth asking experts in the field. One test of a good search is to find studies that were not known to you or to topic experts.

A practical constraint for reviews of effectiveness is '*publication bias'*. It has been shown that studies of impact that find no effect are less likely to be published. Similarly, some study results are withheld by those with vested interests in the outcome of those studies. This has been raised as an issue in the pharmaceutical industry, where studies showing no effect of a drug may affect sales (Chalmers 2006; Goldacre 2012; see also the BMJ Open Data Campaign[5] and the All Trials initiative[6]).

---

[5] http://www.bmj.com/tamiflu
[6] http://www.alltrials.net/

Such publication bias can in turn bias the synthesis of findings, which may show positive effects for a particular intervention that might not otherwise be the case.

Searching is often an inefficient process. Many of the papers elicited will not to be relevant. The more thorough the search, the greater the proportion of relevant studies that are likely to be identified (increasing the *'sensitivity'* of the search); however, the number of non-relevant studies is also likely to increase (lowering the '*specificity'* of the search).

The aim of some searches is to identify all relevant studies. Reviews that are predominantly aggregative would seek to include all relevant studies to avoid any bias in reporting of the findings. In reviews that predominantly arrange, or configure, study findings, the need to find every study is less critical. This is similar to primary research where aggregating research seeks complete or representative populations whilst configuring research seeks often smaller samples from specific or diverse contexts, rather than statistically representative samples. Detailed guidance for finding effectiveness studies is available from the Campbell and Cochrane Collaborations (Hammerstrøm et al. 2010; LeFebvre et al. 2011).

Guidance for finding a broader range of studies is available from the EPPI-Centre (Brunton et al. 2012).

Screening

Studies identified by the search are then checked ('*screened*') to ensure they meet the inclusion criteria. If part of the search has been by hand, then the full paper will be immediately available and a decision can be made as to whether it meets the inclusion criteria ('*one stage'* screening). In web and database searching, only the title and abstract may be available. In that case, initial screening is required to determine whether the study is likely to be relevant and then a full copy of the paper is required to complete the screening exercise ('*two stage'* screening). In many cases, it will be clear from the abstract whether or not the study is relevant. For example, a search for studies on the effects of homework on children might identify papers on work in a children's home or working at home for those with children. But in some cases it may be difficult to make a clear judgment from the abstract and many full papers have to be retrieved.

Screening can have large resource implications. A search might identify 20,000 abstracts, which can make the screening process lengthy. Even if 95 per cent of those abstracts are screened out as irrelevant, the reviewers will still then need to source and read 1,000 papers. Information technology solutions can now assist with this process. If a sample of abstracts is coded as being potentially relevant, text mining technology can rank the remaining abstracts in order of their likely relevance (Thomas et al. 2011).

Coding

Once relevant studies have been identified, the relevant information needs to be extracted from each study. This information is used to describe:

- The characteristics of included studies to provide a 'map' of the research on this issue (as defined by the review question and inclusion criteria);
- How the research was undertaken to allow assessment of the quality and relevance of the studies in addressing the review question; and
- The findings of each study so that these can be synthesised to answer the review question.

Mapping the research field

A description of each study identified in the review, once compiled, describes the field of interest. This field of interest is defined by the review question and inclusion criteria which could be anything and so may not necessarily fit with traditional topic areas. What is described in the map might simply be the names of the authors and the country in which the studies were undertaken. It is however possible to describe many more details of each study and the details coded and 'mapped' will depend upon the interests of the reviewers and potential users of the review. Maps do not typically contain research findings because maps occur before quality and relevance appraisal and including findings may lead readers to assume that these were quality assured findings.

There are three reasons for systematically describing, or '*mapping*' the research field.

i)   Maps as a research project

A map of the included studies may be useful in clarifying the nature of research that has been undertaken to date on an issue.

Maps can include any variable to describe a research field, for example: where the studies were undertaken, the conceptual assumptions involved, the sample and the research methods applied. A review of research on personal development planning (PDP) in higher education shows how a map can add useful context to a review. The included studies were coded on numerous variables including their research design and the country where the study was located. The resultant map showed that US studies were predominantly experimental and aggregative testing whether PDP worked, while the UK studies were predominantly interpretative configuring approaches to examining the processes by which PDP had its effect (Gough et al. 2003).

Systematic maps of research fields can also highlight gaps in research. It could be argued that there were gaps in both the US and UK evidence bases on personal development planning, given the focus on aggregative and configurative studies respectively. Similarly, maps can be used to compare policy and practice on the ground with what has been studied in research; they may reveal that only a specific sub-set of policy and/or practice has been studied.

ii)   Maps as a stage to synthesis

Mapping can also inform the selection of studies for the synthesis of study findings. The studies contained within a research field may be too numerous or heterogeneous for meaningful synthesis; it might be methodologically too difficult or just take too much time. The map provides an opportunity to select a sub-group of studies for synthesis. This can involve undertaking a single synthesis based on a narrowed review question and set of inclusion criteria; or undertaking a series of syntheses. For example a review of children and homework may involve syntheses on sub-topics such as the effects of homework on attainment or the process of parental involvement

Syntheses can also be restricted to studies employing specific research methods: the personal development planning review was asking a question of extent of impact and mapped all outcome studies of personal development planning but restricted the synthesis to experimental evaluations.

iii)   Maps to inform synthesis

Maps can also provide useful contextual information to inform the synthesis. In particular, they can inform the interpretation and integration of findings from the studies included in the synthesis. This might include contextual information about studies that may be relevant to the interpretation and application of the findings of reviews in local contexts.

Appraising the quality and relevance of studies

Once studies have been assessed as meeting the inclusion criteria and so be suitable for inclusion, the next step is to assess their quality and relevance to addressing the review question. This assessment determines how much weight is placed on the evidence of each study included in the final synthesis. There are three key components to the assessment: the study's relevance to the review question, the appropriateness of its methods in the context of this specific review, and the quality of the execution of these methods (Gough 2007).

The first component is the relevance of the study to the research question. Some studies may meet the inclusion criteria but be less relevant than others due to specific aspects of the study design. This might be due to characteristics of the sample, the nature of the intervention, or the way that data were collected. For example, a review of homework may include some studies of very atypical aspects of homework or studies that used a very unusual measure to assess the effects of homework. Issues of the ethics of the primary studies can also be included in this component.

The second component is the extent to which the research design and methods used in a study, however well executed, were appropriate for addressing the review question. This is not a critique of the study *per se*, but of its relevance to a particular review question. Depending on the review question, some well executed studies may be less relevant to the review than other studies that may not have been executed as well. This is more likely in reviews with broad inclusion criteria that consider studies with a variety of research designs. Some people use the term 'hierarchy of evidence' to describe the relative fitness for purpose of different study designs, but as different research questions will be best answered by different types of study, there is not a single such hierarchy of evidence.

The final component is the extent to which the study has been well executed. This is a generic judgment that could be applied to a study irrespective of the review. A study of processes would be evaluated against expected quality criteria for a process study; while a randomized controlled trial would be evaluated against quality criteria for such experimental studies.

All three components are relevant when considering the quality and relevance of a study. But there is considerable variation in how they are combined to produce an overall judgment on a study. Some reviews score studies on each component and then combine these scores. This has the advantage of clearly setting out the criteria by which quality and relevance has been assessed. However, the final judgment will be affected by how much emphasis is placed on each component. Other reviews specify the minimum criteria that studies have to meet for each component.

Many published scales are available for assessing study quality – each incorporate criteria relevant to the research design being evaluated. Quality scales for experimental studies tend to focus on the extent to which a study protects against experimental bias. Quality scales for appraising qualitative research tend to focus on the extent to which the study reflects the meaning of the data.

Experimental studies judged to be of low quality may be excluded from the review and/or final synthesis. Quality judgments of qualitative studies may simply be used to help the reader judge for themselves the utility of individual studies and the review as a whole.

There are also systems that both evaluate the research evidence and link this to the recommendations for decision making such as guidelines for professional decision making. The GRADE[7] system, for example, has been developed to assess the quality of evidence used in reviews on the effectiveness of interventions. It includes a quality appraisal of both the studies included in a review and the recommendations based on the synthesis of the study findings. The assessment comprises four components:

- Quality of evidence (based on the quality appraisal of the studies included in the synthesis).
- Uncertainty about the balance between desirable and undesirable effects (as interventions may have negative as well as positive impacts).
- Uncertainty or variability in values and preferences (as users of services may or may not value the service or its impacts).
- Uncertainty about whether the intervention represents a wise use of resources (as the most effective intervention may only be marginally more effective than a much cheaper intervention and so may be less cost effective).

Synthesis

Each stage of the review proceeds towards the end goal of synthesis. Synthesis is the process of integrating the findings from the included studies to answer the review question. This involves examining the available data, looking for patterns and interpreting them.

A synthesis is more than a list of findings from the included studies. It is an attempt to integrate information and produce a more definitive answer to the review question than the individual studies included the review can provide.

The methods of interpreting and integrating the data depend upon the research question and the methods used in the primary studies included. The differences between aggregative and configuring approaches to synthesis are discussed in Section 3.

Reporting of a review

The report sets out the background to the review, its aims (and potential users and how their perspectives have been engaged in the review), the review methods (the 'protocol'), and the results of the map and synthesis (the answer to the review question). The report should also include discussion of the strengths and limitations of the review and how this should effect interpretation of the findings (Moher et al. 2009). This may include information on contextual issues relevant to the interpretation and application of review findings.

As well as reporting the key findings, systematic reviews require clear reporting of method. This should include details of the decisions made at each stage of the review process.

---

[7] http://www.gradeworkinggroup.org/

There are also publication standards for reviews of evaluations of interventions (PRISMA[8]), meta narrative (Wong et al. 2013a) and realist syntheses (Wong et al. 2013b)[9]. Such technical information can take over a hundred pages to describe. The need to be explicit about the approach taken can conflict with the need for accessible findings. It is common therefore for reviews to be reported in a range of formats with different degrees of detail. A commonly advocated approach is for a series of three formats: a 1 page summary; a 3 page extended summary; and a 25 page full report. An alternative approach advocated by the authors is to produce a:

- 1 to 2 page summary
- 25 page report
- 100 + page technical report or appendix
- Access to all data codings

Questions and answers

Q 2.1: Should I expect to follow these stages for any review?

A 2.1: All reviews have each of these stages, but they may differ in how the work is done at each stage, the order of each stage, and the importance of each stage. For instance, for a review testing a hypothesis, all the searching is done before the studies are appraised and synthesised; moreover, seeking as many studies as possible is important to avoid being misled by only looking at studies that are easy to find. In contrast, a review building a theory may revisit the searching stage as the findings begin to emerge and suggest some key issues needing more information. Theory building reviews may also have a different approach to the quality and relevance appraisal of studies. In multi component reviews there are several sub-reviews each with their own stages and these may then be combined in an overall synthesis.

Q 2.2: I have read of other terms to describe reviews and their stages, so why are these not included here?

A 2.2: There are a confusing array of terminologies and it is not possible to list them all. Many of the labels do not accurately distinguish different ways of reviewing so we have tried to present some of the organising principles and descriptions of methods.

---

[8] http://www.prisma-statement.org/
[9] See also Gough 2013

3 HOW REVIEWS VARY

Reviews can address a diverse range of research questions. Consequently, as with primary research, there are many different approaches to the systematic review of research. The approach chosen will be dictated by the review questions, assumptions, methods, and answers. Attempts to classify review methods are not always comprehensive and consistent. However, there is one broad distinction that can be helpful for understanding how reviews differ. Reviews can either *configure* findings to create to develop ideas or theories; or *aggregate* findings to test a theory (see also Box 1 in part 1 of this booklet). Most reviews have elements of both.

Configurative approaches to synthesis

Configurative approaches to primary research and research reviews are concerned with meaning and interpretation. They are used to develop ideas and theories. They tend to use exploratory and iterative approaches that emerge through the process of the research – they are *'inductive'* rather than *'deductive'*. Research that is predominantly configuring often relies on a small number of detailed cases to develop an understanding of processes and mechanisms and meanings. Heterogeneity is more useful than homogeneity. A spread of different and unusual cases may provide greater insights than a representative sample that reveals more about typical cases.

This approach is reflected in configurative approaches in reviews. Ethnography, for example, is a primary research method where the researcher engages with the real world to explore social meanings. Meta ethnography (Noblitt and Hare 1988) uses similar methods to explore and integrate the findings of previous ethnographies in order to create higher level explanations of phenomena.

There are many other forms of configurative approaches in reviews (see for example, Barnett-Page and Thomas 2009), reflecting the variety of methods used in primary research. For example, meta theoretical (see Patterson et al. 2001) and meta narrative reviews (Greenhalgh 2004, Wong et al. 2013a) have been undertaken to review how heterogeneous topics have been studied in different ways.

Aggregative approaches to synthesis

Aggregative approaches to the analysis of data in both primary research and research reviews are concerned with measuring, then adding together and comparing the measures taken. Aggregative approaches specify their methods in advance (*a priori*) and then apply them – they are *'deductive'* rather than *'inductive'*. For example, RCTs and reviews of such trials test the hypothesis that an intervention has the impact that has been predicted. In some cases, a simple '*black box*' study will determine whether there is a difference between groups receiving and not receiving an intervention. In other cases, more complex logic models or theories of change are tested to determine what causes the observed change. Aggregative approaches to research do not develop theory directly; though it can lead to theory development through empirical testing of the assumptions of a theory.

As with primary aggregative approaches, aggregative approaches to reviews seek to avoid bias. To do this, they aim to include every study that meets the inclusion criteria. Studies can only be easily aggregated if they are similar to each other. Therefore, unlike configurative approaches, aggregative approaches seek homogeneity rather than heterogeneity between the studies.

The best known aggregative approaches to reviews are reviews of the effectiveness of interventions. Relevant statistics from each study (such *as 'effect size'*) are combined statistically ('*statistical meta analysis*') and represented graphically (in a '*forest plot*') to provide an overall measure of effect (Figure 1). The forest plot shows:

- the estimated effect size for each study and a measure of its size/weight in the meta analysis – shown by the position and size of square boxes for each study the confidence intervals for each of these effect sizes (the probability that the effect size is within certain ranges) – shown by the horizontal lines for each study;
- whether these confidence intervals cross the vertical line of effect (and thus limit the confidence that there was an effect) – whether they cross the vertical line at zero;
- the overall effect size from the meta analysis, and its confidence interval – the location of the diamond and its breadth; and
- whether this effect size and its confidence interval show a positive or no effect of the intervention - it shows a positive effect in the example in Figure 1.

Figure 1: Example of a forest plot
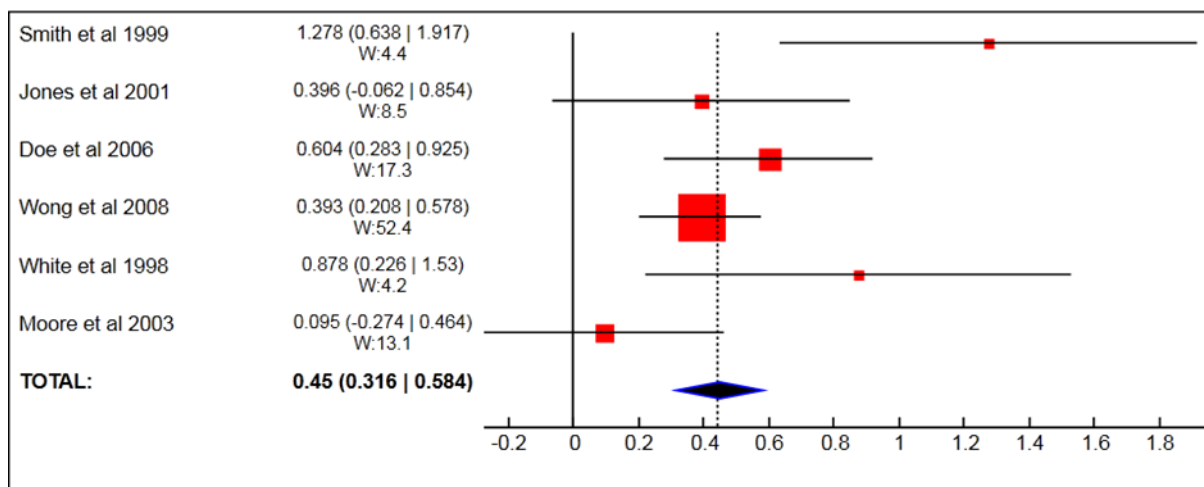
In some reviews, insufficient data are available or studies are not sufficiently similar to allow a statistical meta-analysis. This creates difficulties for a synthesis. It is not sufficient to simply count the number of primary studies showing a positive, negative or no impact, as the studies will vary in sample size and quality of execution. However, looking at patterns in the studies may be fruitful; for example, if the results of the high quality studies in a review all suggest an intervention is effective, it may lend greater weight to a conclusion about overall effectiveness. If findings are mixed across studies it may be very misleading to simply 'vote count'.

There are many other types of aggregative approach to reviews. Cost effect analyses collect data on the economic costs of interventions and then compare this with data on the effectiveness of interventions to enable cost effectiveness comparisons.

Reviews of diagnostic accuracy consider the ability of a test to make an accurate diagnosis (see for example, Harbord et al. 2007).

Approaches to reviewing that both aggregate and configure findings

The previous two sections have detailed two important ways in which study findings are combined in systematic reviews. Few reviews employ one approach to the exclusion of the other, however, and some policy-relevant review questions require that findings are both aggregated AND configured. For example, answering the question "what is the best school-based intervention to increase children's physical activity?" requires that a wide range of research is included, covering many different types of intervention. The reviewer will put the studies identified into different 'piles', according to the approach taken in the intervention evaluated. They will then combine (aggregate) the findings within each intervention category in order to see how well each type of intervention can increase physical activity. They will then explore and explain differences in findings (configuration) between the different approaches in order to identify which approach is most suitable for different children in different contexts. Research methods used in this sort of configurative analysis include (meta-) regression, structural equation modeling and multilevel modeling. However, another example of aggregation within a configurative analysis might be to count how many times a particular concept or theory appears within the studies and use this tally as a way of determining how important or significant that theory / concept is.
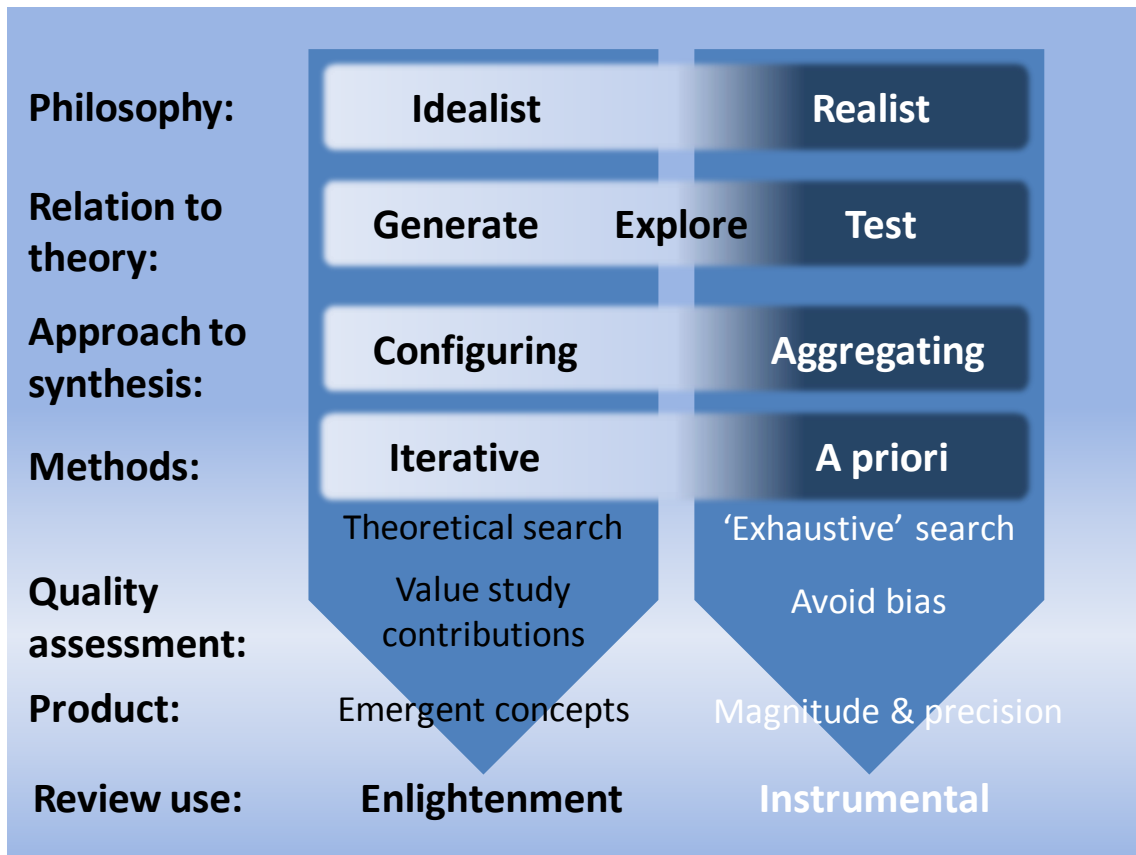
Comparing configurative and aggregative approaches

The different research paradigms underpinning configurative and aggregative approaches to reviews are reflected in the differences in the methods used – from initial screening, through to mapping, synthesis and reporting (Figure 2).

For configurative approaches, screening involves appraising studies against the needs of the review rather than against standard criteria; once a review team has found some examples of a particular element in their configuration (e.g. a specific context), they may not seek other examples as they have reached *'saturation'* for that part of their synthesis and additional studies would not add any new information. Aggregative approaches, however, wish to be as exhaustive as possible (within the constraints of the inclusion criteria) to avoid bias and obtain greater confidence in the precision of their findings. As configurative approaches are less likely than aggregative approaches to produce an exhaustive list of studies, they may not produce a comprehensive map of the research field. Nevertheless, arranging studies in a map is a configuring activity (and some configurative reviews have that as their main aim, as in meta narrative reviews describing research traditions (Wong 2013a)). Quality and relevance appraisal criteria for configurative approaches may reflect the need for insight and meaning rather than for the ability to aggregate data. Synthesis in configurative approaches is an interpretive process that involves examining patterns in data or integrating concepts to produce higher level explanations. Synthesis in aggregative approaches involves aggregating data to describe, compare or test hypotheses.

Figure 2: Summary of differences between configurative and aggregative approaches (from Gough et al. 2012c)

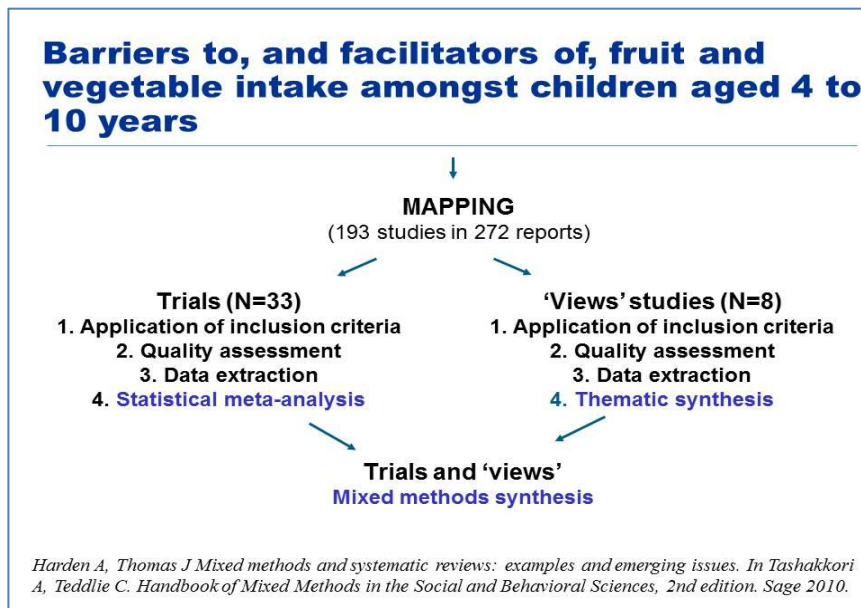|	Configurative		Aggregative
Philosophy:	Idealist		Realist
Relation to theory:	Generate	Explore	Test
Approach to synthesis:	Configuring		Aggregating
Methods:	Iterative		A priori
Quality assessment:	Theoretical search Value study contributions		'Exhaustive' search Avoid bias
Product:	Emergent concepts		Magnitude & precision
Review use:	Enlightenment		Instrumental

Aggregative and configurative approaches to reviews provide distinct products that influence decision makers in different ways: configurative approaches inform ways of thinking and have an enlightenment function; aggregative approaches have a more instrumental effect (Weiss 1979).

Mixed method, multi-component reviews

'*Mixed methods*' reviews are increasingly common, possibly because questions about policy often do not sit within neat academic boundaries, but require evidence located in different disciplines using a range of methods (Brannen 2005; Hammersley 2000). In order to develop policy, it is often necessary to bridge the divide between the micro (people's experiences, understandings and in-depth contextual knowledge) with the macro (the patterns apparent on a population or sub-population level).

Figure 3: Example of a mixed methods synthesis

Barriers to, and facilitators of, fruit and vegetable intake amongst children aged 4 to 10 years

↓

MAPPING
(193 studies in 272 reports)

Trials (N=33)
1. Application of inclusion criteria
2. Quality assessment
3. Data extraction
4. Statistical meta-analysis

'Views' studies (N=8)
1. Application of inclusion criteria
2. Quality assessment
3. Data extraction
4. Thematic synthesis

Trials and 'views'
Mixed methods synthesis

Harden A, Thomas J Mixed methods and systematic reviews: examples and emerging issues. In Tashakkori A, Teddlie C. Handbook of Mixed Methods in the Social and Behavioral Sciences, 2nd edition. Sage 2010.

As in primary research, review methods can be mixed in a variety of ways (Brannen 2005). Some reviews analyse data originating from different research traditions separately before bringing the analyses together. This involves dividing the review question into sub questions, as with the example given in Figure 3. In effect this review involved three syntheses: a statistical meta-analysis of the impact of health promotion interventions to increase fruit and vegetable intake by young people (aggregative strategy); a qualitative synthesis of young people's views about eating (configurative strategy); and an interrogation of the aggregative strategy with the results of the configurative strategy configurative strategy) (Thomas et al. 2004).

Another approach involves treating all research data as evidence, and pulling that evidence together to create, identify and test theories. 'Realist synthesis' (Pawson 2006, Wong et al. 2013b, Rycroft-Malone et al. 2010) appraises social policy interventions through a two stage process: firstly, unpacking the underlying theory; secondly, examining the empirical data to test the unpacked theory. This strategy may not seem different from combining a configurative analysis of theory with an aggregative strategy to test theory. However, realist synthesis differs in using an iterative approach to gather empirical data rather than the *a priori* approach used in traditional aggregative approaches to reviews (Gough 2013).

Another mixed method is framework synthesis (Oliver et al. 2008; Carroll et al. 2011). Evidence from the different studies is incorporated in a framework that evolves as understanding grows through the review process. Such an approach has two advantages for policy reviews. It enables pre-existing expertise and perspectives to shape the framework of the review. Policymakers, practitioners and service users can also help to shape the initial framework, and adapt it as findings emerge. This approach particularly suits exploring new policies where there has been little theoretical thinking.

Systematic reviews are just one tool in the research toolbox. A systematic review rarely exists in isolation. It may be preceded by different forms of primary research and other reviews based on different perspectives and assumptions.

The need for aggregative and configurative components in a new review will depend upon the question(s) being asked, the theoretical and empirical work already undertaken and the resources available. It may be decided to undertake a complex review with multiple components (as in a mixed methods review); or a much more limited review that helps to fill in just one small gap in the evidence base. It is also possible to undertake mini reviews or primary studies to fill in gaps from the previous reviews. This may provide a more informative and cost effective approach than building an overly narrow or shallow rapid review from scratch.

Other types of reviews

As well as full systematic reviews, a number of other types of reviews might be suitable, depending on: the policy imperative; the question(s) to be addressed; and the resources and time available.

Scoping reviews

A scoping review is a brief, non-systematic look at the research that has been undertaken in a particular field. Scoping reviews are often useful to inform the development of a systematic review. Some, however, use the term more broadly to refer to quick often unsystematic reviews. Others use it to refer to broad searches to describe the nature and extent of evidence (which we would call a map).

Rapid reviews

The need for quick responses to research evidence to inform policy decisions as well as a lack of resources to undertake reviews has led some to advocate rapid reviews (Thomas et al. 2013). There are a number of points to consider when commissioning, conducting or reading rapid reviews[10].

Firstly, if rapid reviews are not systematic, to what extent do they suffer from the weaknesses described in Part 1 of this guide? This will affect how much reliance can be placed upon their findings. Secondly, how 'rapid' is the review, that is, how does it compare to a full systematic review. There is no set length for a rapid review – some may be more comprehensive than others. Thirdly, if reviews are relatively rapid, how was the rapidity achieved? It could be by: increasing staffing (so shortening the time but not the costs); narrowing the question (which may limit its fitness for purpose); limiting the search for studies or other stages of the review process (which may lead to a shallow review that is subject to bias); reducing quality assurance processes (such as double coding of data that increase error (Buscemi et al. 2005); or applying recent advances in information technology to identify relevant studies. Fourthly, if a review manages to be fit for purpose in terms of managing to be both rapid and systematic is it still fit for purpose in terms of its intended use. There are systematic rapid reviews that are so constrained in focus (to enable them to be rapid) and where relevance has been so compromised for timeliness that they are of very little use in practice. In sum, rapid reviews can be useful research tools providing the limitations are minimized and/or taken into consideration.

---

[10] Guidance on carrying out rapid reviews of social policy can be found at:
http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment

Living reviews

Systematic reviews are not necessarily completed; they can be updated whenever a new study becomes available. '*Living maps*' can be created, where research activity defined by the original inclusion criteria is regularly updated by repeat or automated searches of the literature.

Living reviews can help to change the way in which new primary research is conducted. Researchers may consider not only what the new study can achieve on its own, but how it can contribute to a review of all the research in that field. A good illustration is the statistical power that a RCT needs in order to be able to identify a difference between the effects of two interventions. Statistical power is dependent upon sample size and the nature of the outcome measures. It is normally calculated on the basis of the RCT alone. However, if the RCT was to contribute to a living review, the statistical power calculation would be based on what would be required to update the review in a way that would help to differentiate whether the intervention was effective or not. Researchers should of course consider how their studies can contribute to the evidence base but this is more clearly identified as an issue in living reviews.

Reviews of reviews

Before making any decision or undertaking new research it may be sensible to see what is already known from research already undertaken. Systematic reviews provide an explicit way to do this, but in some areas of study there are also existing reviews and so the most useful strategy may be to systematically review these reviews. Such reviews of reviews may, like all other reviews, be asking any research question and may use various methods of review to do this (and may have configuring and/or aggregative components).

Examining previous reviews may be an efficient way to examine a research literature through both mapping reviews and synthesizing their findings. It is possible to synthesise the findings of previous reviews. However, this can be difficult, as reviews of the same topic are likely to involve slightly different questions, methods and time lines. Taken together they may not provide a completely clear appraisal of what is known. They can also identify gaps in knowledge of research available from existing reviews. One resource effective strategy is thus to combine a reviews of reviews with new review work on the gaps in review knowledge to provide a quick overview of an area of study (see also appraising the quality of reviews in Part 4 of this booklet).

Questions and answers

Q 3.1: How do reviews vary?

A 3.1: They vary considerably (for example, in type of question, theoretical and ideological assumptions, breadth and depth of inquiry, and review design and specific methods). This should not be a surprise as primary research has a broad variety of questions and methods and these are reflected in the secondary research of systematic reviews. The critical issue is the fitness for purpose and quality of the primary studies and the reviews of those studies.

Q 3.2: How do I know what sort of review I need?

A 3.2: If you want to understand a problem you want a review of studies that look at an issue from different angles and use them to build a more comprehensive or coherent picture that is possible from one study alone.

If you want to predict the effects of a particular policy, or identify factors causing a problem, you want a review of similar studies that combines the findings to provide a more precise measure.

Q 3.2: Why introduce unfamiliar concepts of aggregating and configuring approaches to research?

A 3.2: In our experience there is considerable confusion about the nature and purpose of research. Although the terms quantitative and qualitative research seem similar and seem superficially to be well understood, we believe that these terms create confusion as they do not describe the research activity or its purpose.

4 PLANNING AND USING REVIEWS

Planning and investing in reviews

Reviews are major pieces of research and require a substantial investment of money and time equal to the investment required in much primary research. They therefore require careful commissioning which must take account of a wide range of factors.

Short and long term needs

What short term needs does the review need to fulfill? The review may be needed to help answer a specific question or assist with a specific decision. What perspectives need to be included in developing the review question? Involving a broad range of views may help ensure that all important issues are considered, even if they cannot all be included in the final review question.

What longer term needs can or need to be met by the review? Planning for a review should also involve consideration of how it fits into longer term strategic issues about research in an area. There may be a need to build up evidence in an area to be able to respond at short notice to future questions. There are many existing reviews in health research and so a quick, partial response can be given to many research questions. There are fewer existing reviews in social research. Consideration needs to be given to developing capacity and/or infrastructure to support evidence production, synthesis and use. On the other hand, care must be taken to avoid duplication of reviews (as happens frequently in primary research) so care needs to be taken to identify reviews completed or in progress (and a number of databases and journals are now listing protocols of planned reviews to help with this).

Staff capacity

The skills and capacity required to undertake a review may include:

- Perspective expertise (on ways of understanding the topic): to assist with question formulation, decisions at every stage of the review process and interpretation of review findings.
- Academic expertise (on ways to study the topic): to assist with question formulation, decisions at every stage of the review process and interpretation of review findings.
- Systematic review expertise: systematic reviews are a relatively new method and are not taught on most research methods courses. Many academics do not have such specialist training and skills. Particular parts of the review process may need further specialist skills such as in information science (for searching), text mining (for screening), statistical meta-analysis, and configurative methods.
- Commissioners and managers of reviews require the skills to ensure internal and external quality assurance and relevance processes are adequate and applied.

Information management and analysis

Reviews are complex pieces of research. They are difficult to undertake using paper-based management processes. Records need to be made of: studies found, abstracts screened, requests made for full copies of papers, receipt of full papers, full papers screened, coding, analysis and methodological decisions taken. In addition, the analysis and synthesis may require specialist software for thematic coding or statistical analysis. For this reason, reviewers often use a number of different proprietary software packages or specialist review software such as EPPI-Reviewer.[11]

Using reviews

Appraising the quality of reviews

Readers can appraise the quality and relevance of reviews using similar tools to those used to appraise the quality and relevance of primary studies. The analogous questions are: (i) Does the focus of the review suit the purposes the reader has in mind?; (ii) Did the review method chosen suit the review questions posed? and (iii) Was the review undertaken well within the expectations of that particular review method? As with primary research[12] there are systems for appraising the quality of reviews. AMSTAR[13] can be used to assess the methodological quality of aggregative approaches of the effectiveness of interventions. In addition, the Cochrane Collaboration has developed methodological standards for Cochrane reviews (MECIR[14]). There is some overlap between systems to evaluate reviews and standards for reporting reviews described in Part 2 of this guide.

Interpreting and applying review findings

Ultimately, a systematic review is only useful if people (including those commissioning the review) can make use of it. Studying and understanding how research is used is itself becoming an important area of investigation with academic journals, texts and research networks[15].

Research findings can inform decision making. The findings do not have inherent meaning; they need to be interpreted by the user of that research. In some cases this may be relatively straightforward; for example, where findings show an intervention thought to be helpful to either have no effect or to be harmful and thus further enquiry is required before further use.

One famous illustration of this is the 'Scared Straight' programme developed in the US in the 1970s to deter young people from future crime actually does more harm than good. The idea is that participants visit inmates, observe first-hand prison life and have interaction with adult inmates, and are scared off crime. These programs are popular in many areas of the world programmes but the evidence is that such programmes increase delinquency, relative to doing nothing at all to juveniles in similar circumstances[16].

---

[11] http://eppi.ioe.ac.uk/cms/er4
[12] For primary research, see for example, http://www.equator-network.org/
[13] http://amstar.ca/
[14] http://www.editorial-unit.cochrane.org/mecir
[15] See for example, the journal Evidence and Policy: http://www.policypress.co.uk/journals_eap.asp; the text Nutley, S. M. , Walter, I. C. & Davies, H. T. O. (2007) Using Evidence: How Research Can Inform Public Services. Bristol: Policy Press; and the EIPPEE education research use network: http://www.eippee.eu
[1616] http://www.cochrane.org/features/scared-straight-and-other-juvenile-awareness-programmes-preventing-juvenile-delinquency

Another example is the use of micro finance to support poor people in international development. A recent review found that such schemes could in some circumstances have positive effects on health and access to food but could also make people poorer and reduce the access to education for some children[17].

In practice there may be many reasons why a piece of research (or a review of research) may not be relevant or be difficult to interpret.

- Research question: this may differ from the interests of the user in a gross way or in detail of the perspectives (ideological and theoretical assumptions) within the question.
- Context: even if the research question is highly relevant to the user, the context in which the research has been undertaken may not be.
- Execution of the research: the research may not have been undertaken well or have some problems that limited its relevance in practice.

If the perspective of the user of the research is not well represented in the setting of the research question, the contexts studied and the execution of the study then the findings of the research may not be very relevant to that user. These gaps between the focus of interest of the research and the potential reader of the research create challenges for the interpretation of the research. If the gap is wide then the user may need to rely on other evidence to assist in interpretation.

The challenge for interpretation may also be affected by the strength of the research findings. If, for example, a review of research on the impact of an intervention in another country found that most people died after receiving the intervention then, that finding would likely be of concern to those in other contexts. In many areas of social policy, findings are not so clear cut. In addition, reviews of research provide a summary of findings across different contexts and may not describe what the contexts are or link the findings to particular contexts.

In addition to interpreting the meaning of research findings, there are also a range of further factors that may be involved in applying the interpretation of findings in practice. Research may suggest that a particular course of action should be undertaken but there may be problems of operationalizing it on the ground. It may be that there are not the human or material resources to apply the findings in a coherent way. A further problem is that reviews of research may not provide data to assist in such decisions about application of findings in different contexts.

An example of the challenges of interpretation and application is the use of social policy programmes developed with very strict schedules to ensure intervention fidelity. Not only may the findings of the research be applicable only to the context into which it is developed, but the attempts to ensure fidelity of application may prevent the adaption of the research to other contexts. In these circumstances, a review that informs decision makers about mechanisms of change may be equally if not more informative as those that simply report on levels of impact.

Models of research use

The gathering together of relevant research findings to inform decision making is in contrast to the 'cherry picking' of research findings to support prior decisions. The use of research findings to make specific decisions assumes that decisions are based on facts but the effect of research may be through the ways that problems are conceptualized and understood.

---

[17] http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2965

Weiss (1979) described this as the enlightenment model of research use where research helps develop theories and concepts that may be applied in practice.

Undertaking research where the data or concepts may be useful subsequently for decision makers is a production or 'push' model of research use. However such research findings may not be easily available to potential users of research who, being busy people, have neither the time nor inclination to sift through an enormous amount of research which is often written in academic language in academic publications. Reviews of research can assist by mapping and synthesizing research, particularly if the information is made available in user friendly summaries with access to the full details of the review as necessary.

Although making research available is a necessary condition for its use, it may not be sufficient. We know from studies of research use that the availability of information alone does not normally lead to its use. As well as the 'push' from the research being produced, there also needs to be demand or 'pull' from those who might use the research. This demand can be built in from the start through user involvement in setting priorities, framing research questions and undertaking the research, including reviews of research. Also, being aware of the need for research uptake can inform the development of strategies and advice to achieve this (see for example, Oxman et al. 2009).

Policies, processes and infrastructure for research production and use

National governments and institutions, to varying extents, develop policies for the production and use of research and invest in processes and infrastructures to enable this. Most investment is usually directed towards the production of primary research with relatively little invested in reviews of research or capacity to enable its use. In recent years, several governments and research collaborations have begun to increase their investment in two types of knowledge centre to support the use of research evidence.

One type of centre is mainly concerned with making research findings available through producing and communicating the findings of research reviews of the impact of interventions. Examples of such initiatives are the Cochrane Collaboration in health, the Campbell Collaboration in other areas of social policy, clearing houses such as the federally funded What Works Clearinghouse in education in the United States, and What Works Centres by the UK government and other institutions[18].

These organizations aim to make better use of research by making research available. The reviews of research bring together what is known and also indicate what is not known and thus help inform future research. Also, the gaps in knowledge from research not having been undertaken or not being well executed or well reported provides pressure to improve the quality of future research. The reviews are driven by academic or other research user interest but their emphasis is often on the availability of existing research so they can have rather an emphasis on research production (or 'push' focus).

Some other organizations start with more of a decision making focus. They start with specific issues where decisions need to be made and research and other evidence is then sought to assist to support that decision. These centres have a clearer demand (or 'pull') focus as being based on developing advice and guidance to support decisions.

18

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/136227/What_Works_publication.pdf

A well-known and very well developed example is the National Institute for Health and Care Excellence (NICE)[19] that produces guidance and quality standards for service delivery. NICE starts with a decision to be made by government or broader society, advertises and invites potential stakeholders to engage in a consultation exercise on how to focus the issue and forms a group with stakeholder representation to manage the guidance development. The guidance development committees then commission reviews of research and other evidence including information on practical contexts to develop guidance for health and social care services. This provides a transparent democratic process for developing research informed guidance for public services.

Questions and answers

Q 4.1: Is a systematic review the only or always the best approach to reviewing evidence?

A 4.1: All research needs to be fit for purpose. If the requirement is for a rigorous piece of work that uses research methods to bring together findings of research, then a systematic review is appropriate. If, however, there is no need for such rigour and quality standards then other methods may be relevant. The same is true of primary research which has more rigour and different purposes, standards and timescales to, for example, a newspaper article on today's news.

Q 4.2: What do you need to know before undertaking or commissioning a review?

A 4.2: Many things including: (i) Is the review question and its conceptual framework and methods appropriate for the issue to which the review is to be put; (ii) are there appropriate resources including perspective, topic, and reviewing skills available; (iii) how will the review findings be used in practice and what is required to achieve this?

Q 4.3: What part do systematic reviews play in a research strategy?

A 4.3: Reviews: (i) identify what is known and how and therefore suggest what is not known and how it could be known; (ii) should be the first thing that is done – before undertaking and potentially duplicating prior research – and so should inform primary research; (iii) are of equal importance to primary research and so require similar levels of investment in training and resources: (iv) are part of a developmental strategic approach that may involve many components of primary research and single and multi-component reviews over time; (v) need user input to enable them to be a demand ('pull') rather than a supply ('push') product; (vi) need to be undertaken and used within an understanding of research uptake; and (vii) may be more sustainable and useful if supported by infrastructures of quality standards and information systems.

---

[19] http://www.nice.org.uk/

REFERENCES

Anderson, L.M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D., and Tugwell, P. (2011) Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods*, 2(1):33-42.

Barnett-Page E., and Thomas, J. (2009) Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 9:59.

Brunton G., Stansfield, C., and Thomas, J. (2012) Finding Relevant Studies. In Gough, D., Oliver, S. and Thomas, J. *An Introduction to Systematic Reviews*. London: Sage.

Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2005). Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*, *59*, 697–703.

Cabinet Office (2013) *What Works: Evidence Centres for Social Policy*. London: Cabinet Office.

Campbell Collaboration International Development Coordinating Group (2012) *International Development Coordinating Group Guidance for establishing and managing Review Advisory Groups*. http://www.campbellcollaboration.org/artman2/uploads/1/IDCG_Advisory_Group_Guidance_final.pdf [Accessed 5 May 2013].

Carroll, C., Booth, A., Cooper, K. (2011) A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Med Res Methodol* 11:29.

Chalmers, I. (2006) From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med* 99:337-341.

Dickson, K., Sutcliffe, K., Gough, D., Statham, J. and Morris, M. (2010) *Improving the emotional and behavioural health of looked-after children and young people*. London: Centre for Excellence and Outcomes in Children and Young People's Services.

Glass, G.V. and Smith, M.L. (1979) Meta analysis of class size and achievement. *Educational Evaluation and Policy Analysis*. 1 (1) 2-16.

Goldacre, B (2012). *Bad Pharma: How drug companies mislead doctors and harm patients.* Harper Collins Publishers. Kindle Edition.

Gough, D.A., Kiwan, D., Sutcliffe, K., Simpson, D. and Houghton, N. (2003) *A Systematic Map and Synthesis Review of the Effectiveness of Personal Development Planning for Improving Student Learning.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Gough, D. (2007) 'Weight of evidence: a framework for the appraisal of the quality and relevance of evidence ', *In J. Furlong, A. Oancea (Eds.) Applied and Practice-based Research. Special Edition of Research Papers in Education* 22, (2), 213-228.

Gough, D. (2013) Meta-narrative and realist reviews: guidance, rules, publication standards and quality appraisal, *BMC Medicine*, 11:22.
http://www.biomedcentral.com/1741-7015/11/22

Gough, D., Martin S., and France, J. (2012a) *Meta-evaluation of the Impact and Legacy of the London 2012 Olympic Games and Paralympic Games. Developing Methods Paper.* London: Department of Media Culture and Sport.

Gough, D., Oliver, S. and Thomas, J. (2012b) *Introduction to systematic reviews.* London: Sage.

Gough, D., Thomas, J. and Oliver, S. (2012c) Clarifying differences between review designs and methods. *Systematic Reviews*, 12, 1-28.
http://www.systematicreviewsjournal.com/content/1/1/28

Grayson L, Gomersall A (2003) A difficult business: finding the evidence for social science reviews. *UK Centre for Evidence Based Policy & practice Queen Mary University of London, Working Paper no 19 Oct 2003 21 pages.*
http://www.kcl.ac.uk/sspp/departments/politicaleconomy/research/cep/pubs/papers/paper-19.aspx.

Green, S. and Higgins, J.P.T. (Eds.) (2011). Chapter 2: Preparing a Cochrane review. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P. and Kyriakidou, O. (2004) Diffusion of innovations in service organisations: systematic literature review and recommendations for future research. *Milbank Q* 82:581-629.

Hammerstrøm, K., Wade, A. and Jørgensen, A.M.K. (2010) Searching for studies: A guide to information retrieval for Campbell Systematic Reviews *Campbell Systematic Reviews* 2010: Supplement 1 DOI: 10.4073/csrs.2010.

Harbord, R.M., Deeks, J.J., Egger, M., Whiting, P. and Sterne, J.A.C. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8: 239–251.

Konnerup, M. and Sowden, A. (2008) User involvement in the systematic review process *Campbell Collaboration Policy Brief.*
http://www.campbellcollaboration.org/artman2/uploads/1/Involvement_in_review_process.pdf
[Accessed 5 May 2013].

Lefebvre, C., Manheimer, E. and Glanville, J. (2011) Chapter 6: Searching for studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. The PRISMA Group (2009).
http://www.bmj.com/content/339/bmj.b2535.

Noblit, G. and Hare, R.D. (1988) *Meta-ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: Sage.

Oakley, A. (2000) *Experiments in Knowing: Gender and method in the social sciences.* Cambridge: Polity Press.

Oliver, S., Peersman, G., Harden, A. and Oakley, A. (1999) Discrepancies in findings from effectiveness reviews: The case of health promotion for older people in accident and injury prevention. *Health Education Journal*, 58(1): 66–77.

Oliver, S.R., Rees, R.W., Clarke-Jones, L., Milne, R., Oakley, A.R., Gabbay, J., Stein, K., Buchanan, P. and Gyte, G.(2008) A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations* 11 (1): 72–84.

Oxman A, Hanney S, editors. SUPPORT Tools for evidence-informed health Policymaking (STP). *Health Research Policy and Systems* [Internet]. 2009; 7 Suppl 1:[S1-S18 pp.]. Available from: http://www.health-policy-systems.com/supplements/7/S1.

Paterson, B., Thorne, S., Canam, C. and Jillings, C. *Meta-Study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis*. Thousand Oaks, California, USA: Sage Publications; 2001.

Pawson, R. (2006) *Evidenced-based Policy: A Realist Perspective*. London: Sage.

Rees, R. and Oliver, S. (2007) An example from health promotion. In Coren, E (ed) (2007) *Collection of examples of service user and carer participation in systematic reviews*. London. Social Care Institute for Excellence; 15 – 34.

Rees, R. and Oliver, S. (2012) Stakeholder perspectives and participation in systematic reviews. In: Gough D, Oliver S, Thomas J *An Introduction to Systematic Reviews*. London: Sage. 17-34.

Rogers, P.J. (2008) Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1):29-48.

Rycroft-Malone, J., McCormack, B., Hutchinson, A.M., DeCorby, K., Bucknall, T.K., Kent, B., Schultz, A., Snelgrove-Clarke, E., Stetler, C.B., Titler, M., Wallin, L. and Wilson, V. (2012) Realist synthesis: illustrating the method for implementation research. *Implement Sci.* 2012 Apr 19;7:33.

Sandelowski, M.,Voils, C., Leeman, J. and Crandell, J. (2012) Mapping the Mixed Methods–Mixed Research Synthesis Terrain. *J Mixed Method Res* 6: 317-331.

Scriven, M. (1969) An Introduction to meta-evaluation, *Educational Products Report*, 2, 36-38.

Stansfield, C., Kavanagh, J., Rees, R., Gomersall, A. and Thomas, J. (2012) The selection of search sources influences the findings of a systematic review of people's views: a case study in public health. *BMC Medical Research Methodology*, 12-55.

Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R., Brunton, G. and Kavanagh, J. (2004) Integrating qualitative research with trials in systematic reviews: an example from public health. *British Medical Journal* 328: 1010-1012.

Thomas, J., McNaught, J. and Ananiadou, S. (2011) Applications of text mining within systematic reviews. *Research Synthesis Methods*. 2(1): 1-14.

Thomas, J., Newman, M. and Oliver, S. (2013) Rapid evidence assessments of research to inform social policy: taking stock and moving forward. *Evidence & Policy: A Journal of Research, Debate and Practice*, 9 (1):, 5-27.

Ueffing, E., Tugwell, P., Welch, V., Petticrew, M. and Kristjansson, E. for the Cochrane Health Equity Field. (2009). *C1, C2 Equity Checklist for Systematic Review Authors*. Version 2009-05-28. Accessed at http://equity.cochrane.org/sites/equity.cochrane.org/files/uploads/Equity%20Checklist%20-%20oct%204%202012%20.pdf on [*accessed 11:46:11*].

Weiss, C. (1979). The many meanings of research utilisation. *Public Administration Review* 1979 29: 426–431.

Wong et al. (2013a) RAMESES publication standards: meta-narrative reviews. *BMC Medicine,* 11:20.

Wong et al. (2013b) RAMESES publication standards: realist synthesis. *BMC Medicine*, 11:21.

Woodman, J., Thomas, J. and Dickson, K. (2012) How explicable are differences between reviews that appear to address a similar research question? A review of reviews of physical activity interventions. *Systematic Reviews*, 1:37.
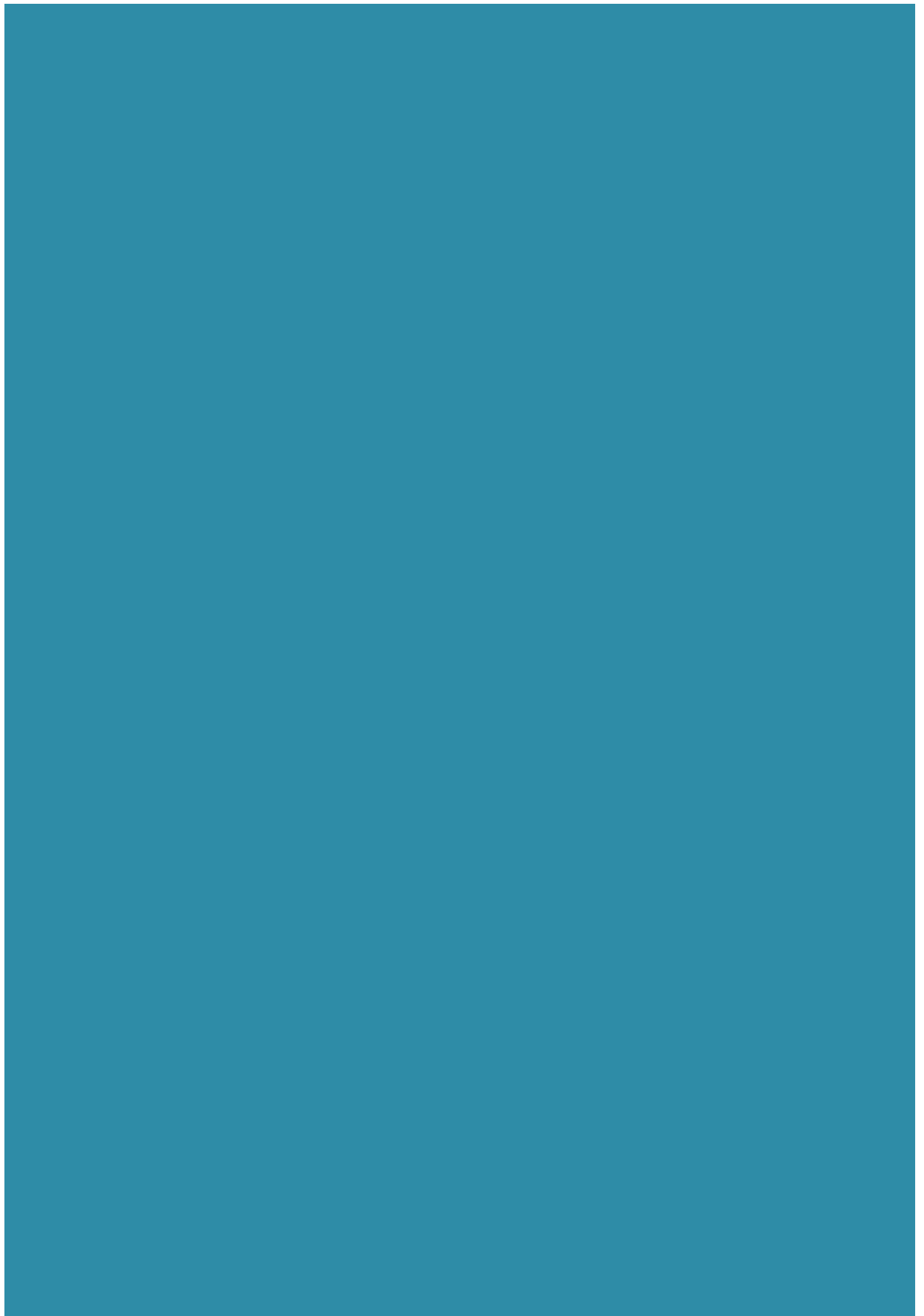
Alliance for Useful Evidence

1 Plough Place
London EC4A 1DE

www.alliance.4usefulevidence.org

December 2013