THE PRESIDENCY
REPUBLIC OF SOUTH AFRICA

DEPARTMENT: PERFORMANCE MONITORING AND EVALUATION

DPME Evaluation Guideline No 2.2.15

Guideline on Evaluation Synthesis

Created 20 March 2014

Addressed to	Government departments who are undertaking evaluations (programme managers and M&E staff) as well as evaluators of government programmes and policies.
Purpose	The purpose of this Guideline is to provide technical guidance on undertaking or managing an Evaluation Synthesis
Policy reference	This guideline should be read in conjunction with the National Evaluation Policy Framework approved by Cabinet on 23 November 2011 (available on the DPME website).
Contact person for this guideline	Jabu Mathe, Evaluation and Research Unit (ERU) E-mail: jabu@po-dpme.gov.za Tel: 012 312 0158

1 Introduction

This Guideline is designed to assist government departments to effectively plan and manage evaluation synthesis. The Guideline provides a definition and description of evaluation synthesis, followed by key issues to be considered, key questions that should be addressed, and typical evaluation methods. This is a broad guideline that can be applied in different contexts. These are broad guidelines that can be applied in different contexts. It is focused on providing an overview for government staff managing evaluations and is not targeted as a manual for an evaluator on how to undertake an evaluation synthesis. Hence it does not go into detail into the different tools that are introduced. Note the word programme is used here but the evaluation could equally apply to a policy, or plan – we use intervention to cover any of these.

The concept of evaluation synthesis draws from the concept of research synthesis. Research synthesis is the practice of systematically distilling and integrating data from a variety of sources in order to draw more reliable conclusions about a given question or topic. Evidence is not limited to any one type of literature. The type of evidence included in a synthesis should depend on the central question that the synthesis aims to address. Reviewers need to determine (a priori) what types of evidence are (and are not) “fit for purpose” for a particular review.

Systematic research synthesis attempts to identify research reports on a given topic and to examine them in explicit and standard ways so as to produce accessible, reliable and useful summaries of research findings (these are sometimes called 'systematic reviews'). As at December 2013 the evaluation repository held by DPME already held 70 evaluations with 40 to 50 to be added in 2014. This provides an opportunity to draw out lessons across these evaluations, as well as potentially from evaluations on similar topics in other countries with similar contexts.

2 Definition of evaluation synthesis

The National Evaluation Policy Framework (NEPF) describes evaluation synthesis as the process of synthesizing results of a spectrum of evaluations, in order to generalize findings across government (NEPF, 2011). Evaluation synthesis brings together existing studies, assesses their

relevance and reliability, and draws together their data to answer specific questions. An evaluation synthesis aims to bring together what is known about a particular type of programme or a particular issue in service delivery. It can explain how, and under what conditions, what type of programmes does (and does not) work. It is more than a listing of the results of individual studies and may take a variety of forms (as outlined below).

3 Purpose of an evaluation synthesis

The purpose of evaluation synthesis is to systematically distil and integrate data from a number of evaluations as well as other sources of evidence in order to draw more informed conclusions about a given question or topic.

An evaluation synthesis can inform programme and policy decisions by identifying feasible, appropriate, meaningful and effective programme models and service delivery practices. It can inform the planning of an evaluation by identifying what is currently known, the gaps in knowledge, and appropriate methods for measurement and data collection.

Conducting evaluation synthesis is important for several reasons. First, the synthesised findings from several evaluations can help decision-makers make better judgments by bringing together available evidence that can be critically appraised (for quality) within the decision-making timeframe. Second, evaluation synthesis makes use of existing evaluation and research about programmes and issues in service delivery, so it draws on more data than a single evaluation. This brings a broad perspective. Third, evaluation synthesis integrates evaluation findings, establishes an accessible base of knowledge and identifies knowledge gaps or needs for a specific topic or theme, which can then guide future evaluations.

Because a synthesis evaluation uses existing data, this can mean producing information for policy decisions more quickly and cost-effectively than doing a new evaluation. However, this is only the case when it is possible to locate and access good quality and relevant evaluations and other evidence that provide detailed information. In many cases, evaluation reports are hard or impossible to locate and retrieve, are not of sufficient quality, are not relevant (for example, evaluations in countries with different social contexts to South Africa) or do not have detailed data.

Evaluation synthesis can be an effective method to assess early approaches for potential projects and policy outputs, outcomes and impacts, using knowledge already generated about particular topics. However, it often does not generate new knowledge but rather summarizes existing knowledge.

Evaluation synthesis can take place at any point as long as relevant evaluations or research exist. These studies may exist within or outside of South Africa depending on the nature of the evaluation question. These could be of implementation (questions around the experience of operating certain types of interventions), or impact.

4 Evaluation synthesis questions

Evaluation synthesis can build on diagnostic, implementation, impact or economic evaluations (see Guidelines 2.2.10, 2.2.12-14) and so use similar questions, providing such data exists. However there is also the possibility of exploring cross-cutting issues. For example evaluation synthesis could review evaluations of a number of different sectorial interventions and explore questions around cross-cutting issues such as the effectiveness of government supply chain management, or how provincial departments integrate appropriate gender messages.

5 Different Forms of Evaluation Synthesis

There are a range of forms of synthesis evaluations which can be undertaken from a literature review to a review of a number of existing systematic reviews, with different implications for the degree of confidence in the findings (see Figure 1) and the time they take (see Table 1)1. A key issue is whether they have a systematic and replicable methodology for identifying and synthesising studies, minimising bias. The use of systematic procedures alone is insufficient for credible research synthesis. The explicit purpose of these procedures should be to reduce bias and error at each step in the review process. That means that some "systematic" approaches are valid and others are not. For example, reviews that only include published studies are likely to produce biased results.

Table 1: Some forms of evaluation synthesis, their nature and the time to complete

Type of synthesis	Nature	Time required to complete
Traditional literature review (sometimes called narrative review)[2]	Narrative, selective review (not systematic); collates relevant studies and draw conclusions from them. Does not use a systematic method to identify studies. Often also don't have a systematic method for synthesizing or prioritizing studies and dealing with conflicting findings.	1 week to 2 months
Quick scoping review	Quick overview of research undertaken on a (constrained) topic This could be systematic, but because it is quick, it is unlikely to be comprehensive or detailed	1-2 weeks to 2 months
Rapid evidence assessment (REA)	Systematic search but a quick overview of existing research on a topic Synthesis of evidence provided by these studies to answer the REA question	2 to 6 months (quicker than systematic review)
Full systematic review	Systematic: A broad review of existing research on a topic and synthesis of the evidence provided by these studies to answer the review question.	8 to 12 months minimum
Multi-arm systematic review	Systematic: Full map and synthesis of different types of evidence to answer the review question	12 months minimum
Review of reviews	Systematic: Same as any of the above methods but only includes reviews of other reviews.	Often quicker than other types of full systematic review

[1] Source: http://hlwiki.slais.ubc.ca/index.php/File:Evidence-review-types.jpg)
[2] Narrative review can get confused with narrative synthesis which can take place within a systematic review

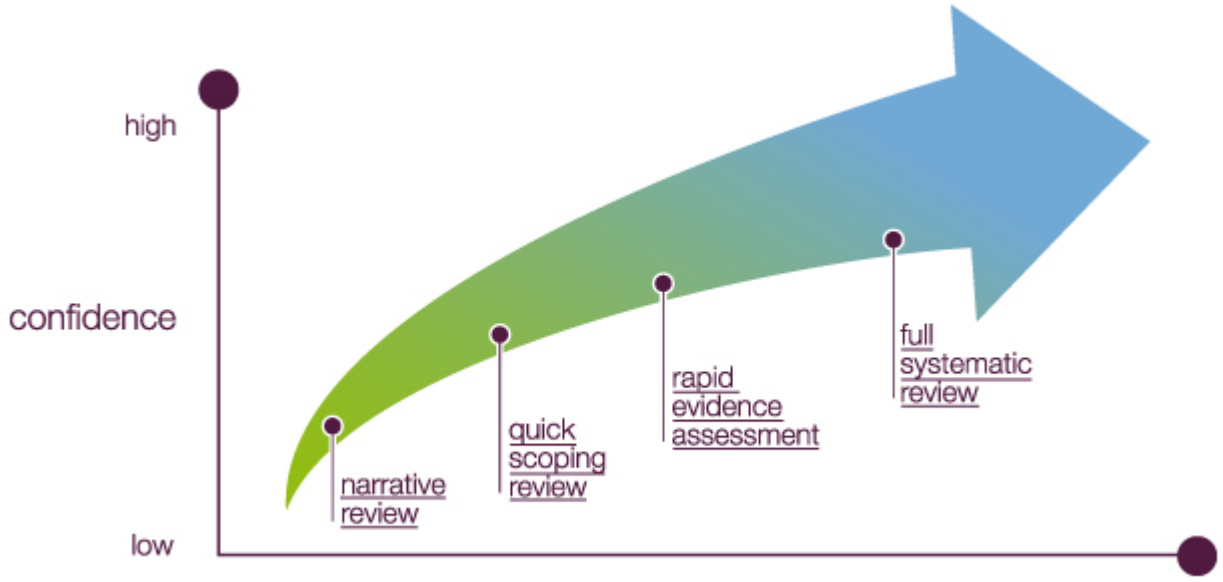


Figure 1: Range of syntheses and degree of confidence versus time to undertake

This guideline focuses on four different forms of evaluation synthesis: systematic reviews, rapid evidence assessment, iterative best evidence synthesis, and thematic analysis (described below). All forms of evaluation synthesis involve a process for identifying relevant material, extracting data, and synthesizing it. The choice of which form of evaluation synthesis to conduct or commission should be based on your question, timeframe and budget, and what your evaluation users would consider credible evidence. To undertake such synthesis requires the ability to use sector databases to search for relevant studies, as well as an understanding of public policy issues.

Systematic reviews take a longer time, as they aim to be comprehensive in their search, and replicable in their process. Rapid evidence assessment is intended to provide quick findings, but is not as comprehensive. Iterative best evidence synthesis is a participatory approach that checks interpretations with policy makers and service deliverers. Thematic analysis is most like a literature review, but more systematic and transparent in its searching and focused on policy questions rather than research questions.

It can be useful to develop a “theory of change” or “logic model” before you begin your review. This helps identify the precise scope of the review, and assists with the searching for evaluation studies that meet this scope. A theory of change also helps to structure the analysis of the evidence, and the presentation of findings to policy makers and other users of evidence. Annex 4 discusses theory of change.

Table 2: Questions and methods for different synthesis evaluations

Purposes	Synthesis focus	Common evaluation methods and approaches
Type 1: Systematic review	Studies which may have a significant effect size (experimental or quasi-experimental studies) and may use meta-analysis to produce a summary in the form of an overall conclusion about effectiveness	• Systematic review
	What overall model can be developed from how concepts are understood across different qualitative studies?	• Meta-ethnography

Purposes	Synthesis focus	Common evaluation methods and approaches
	What works for whom in what circumstances?	• Realist synthesis • Moderator synthesis
	Wide range of evidence (including single cases where relevant), weights evidence more if it demonstrates long term effects and large effect sizes), and recognizes the importance of programme context	• Iterative best evidence synthesis
Type 2: Rapid evidence assessment	Focus on fairly comprehensive electronic searches of relevant databases and some searching of print material. They do not have the exhaustive database searching, hand searching of journals and textbooks or searches of the grey literature of systematic reviews	• Rapid evidence assessment
Type 3: Thematic analysis	Focus on a specific theme and refers to evaluations that cross multiple programmes and projects. Because of the mixed nature of these studies (e.g. they come from different fields or focus on multiple topics), this approach is often used to analyze qualitative research	• Thematic analysis

5.1 Systematic Reviews

While there are different ways of conducting a systematic review, they all use an explicit, systematic process involving six steps. The steps are:

(1) Identifying a specific and answerable review question
(2) Developing and implementing an explicit search strategy, including strategies to locate unpublished research and evaluation studies
(3) Selecting sources based on explicit inclusion and exclusion criteria
(4) Assessing the relevance and quality of studies
(5) Extracting data to help you address your review question in a standardized format
(6) Summarizing and synthesizing the results.
(7) Interpreting and disseminating the results.

Although all systematic reviews use formal, explicit methods to describe and synthesise evidence, they can vary considerably in the types of questions they aim to answer. As a result, different types of evidence will be suitable for answering different questions and different methods will be appropriate for describing and synthesising different types of evidence.

Useful Tip

There are several examples of systematic reviews and their findings in the areas of education, health care, crime and justice, social welfare, environmental evidence and international development. These can be found at the following websites:

- Collaboration for Environmental Evidence (www.environmentalevidence.org/)
- Evidence Encyclopedia (http://www.bestevidence.org/)
- EPPI-Centre Evidence Library (http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=56)
- Cochrane Collaboration (www.cochrane.org)
- Campbell Collaboration (www.campbellcollaboration.org)
- 3ie Systematic Reviews Database (http://www.3ieimpact.org/en/evidence/systematic-reviews/)

A **systematic review** can focus on other questions besides "what works" – in fact, "what works" is a difficult question for a systematic review! Systematic reviews have focused on correlations (e.g., gender differences), trends (e.g., incidence/prevalence of various conditions over time), comparative effectiveness, and moderators of intervention effects. They often focus on studies which have an effect size (experimental or quasi-experimental studies) and may use meta-analysis to produce a summary in the form of an overall conclusion about effectiveness (see the later methods section). They can sometimes identify differences in effectiveness across types of programmes or implementation contexts.

Key Point – What is the difference between a systematic review and a meta-analysis?

A **systematic review** uses transparent and replicable procedures to minimize bias and error at each step of the review process.

A "**meta-analysis**" is one way of summarizing the data from a systematic-review. It produces a weighted average of the effect of an intervention, taking into account the strength of results and the sample size of studies which have met inclusion criteria.

A **meta-ethnography** focuses on answering questions about how concepts are understood, by translating concepts across different studies to produce an overall model. For example a meta-ethnography of factors associated with adherence to treatment for tuberculosis (Munro, Lewin, Smith, Engel, Fretheim , & Volmink, 2007)[3] reviewed 44 published papers that met their predefined inclusion criteria, identified eight major factors associated with adherence to treatment, and developed a simple model that proposes that adherence to tuberculosis treatment is influenced by four interacting sets of factors—structural factors (including poverty and gender discrimination), social context factors, health service factors, and personal factors (including attitudes towards treatment and illness).

A **realist synthesis** focuses on answering questions about "What works for whom in what circumstances" and draws on a range of evidence to build and test a model of how causal mechanisms work in particular contexts. [4]

An **iterative best evidence synthesis** actively engages with stakeholders to optimise the usefulness of the synthesis and encourage ownership by involving them in identifying what would constitute an improvement in outcomes, and implications for practice. An iterative best evidence synthesis draws on a wide range of evidence (including single cases where relevant), weights evidence more if it demonstrates long term effects and large effect sizes), and recognises the importance of programme context.

While the evaluation question will inform the specifics of how a systematic review is undertaken (e.g. methods chosen), all systematic reviews should explain how evaluation studies are selected for inclusion and describe how and why they were selected. The synthesis (i.e. findings) is presented as a structured narrative, summary tables and/or a statistical analysis of impact.

The review needs to provide sufficient detail so that the reader can understand the context of the programme or policy implemented and with what participants and include a clear detailed link

[3] The review is available here: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0040238.
A discussion of the methods used is available here: http://www.biomedcentral.com/1471-2288/8/21 Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC medical research methodology, 8(1), 21.

[4] This presentation provides a nice summary of realist synthesis. http://www.ncchpp.ca/docs/Sridharan_JASP_realistAN_nov08.pdf

between the synthesis and the reviewers' conclusions. The final reports should take into consideration that what works in one place for specific participants may not work everywhere. [5]

5.2 Rapid Evidence Assessment

Where decisions need to be informed by empirical data and time is a pressing factor, a Rapid Evidence Assessment (REA) should be considered. REAs employ a variety of methodologies and vary in terms of the depth of methods used to make the processes rapid. Traditional systematic reviews typically take at least 12 months or longer to conduct (see Table 1). REAs streamline traditional systematic review methods in order to synthesize evidence within a shortened timeframe, sometimes being conducted in a few weeks. However there is still great variation in the length of time for a 'rapid' study. While overall findings from REAs may not vary greatly from Systematic Reviews (detailed in Section 3.1), the latter are more likely than REAs to provide greater depth of information.

Key Point – How is a Rapid Evidence Assessment different from a Systematic Review?
A Rapid Evidence Assessment is a pared down version of a systematic review and draws on the same principles. It is however a 'lighter touch' approach that enables the review to take place more quickly (UK Government Social Research Service, 2007).

REAs draw their data from fairly comprehensive electronic searches of relevant databases and some searching of print material. They do not have the exhaustive database searching, hand searching of journals and textbooks or searches of the grey literature that are used in full Systematic Reviews. Consequently, the findings from an REA have to be more tentative than those from a fully developed systematic review. REAs may also make use of snowball sampling to identify key stakeholders who can inform the identification of relevant studies (HM Treasury, April 2011). A challenge when using an REA approach is to achieve a balance between the 'rapidness' of the study and the trustworthiness of the findings.

Useful Tip
The United Kingdom (UK) Civil Service has a useful toolkit for understanding what REAs can offer policymakers, and how they can be undertaken. The toolkit can be accessed at: http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment.

5.3 Thematic analysis

Thematic analysis focuses on a specific theme and refers to evaluations that cross multiple programmes and projects. Because of the mixed nature of these studies (e.g. they come from different fields or focus on multiple topics), this approach is often used to analyse qualitative research. Thematic synthesis has three stages: the coding of text 'line-by-line'; the development of 'descriptive themes'; and the generation of 'analytical themes'. While the development of descriptive themes remains 'close' to the primary studies, the analytical themes represent a stage of interpretation whereby the reviewers 'go beyond' the primary studies and generate new interpretive constructs, explanations or hypotheses. One aspect of credibility involves the extent to which different reviewers agree on line-by-line coding and development of themes. Some inter-rater agreement checks should be performed and reported.

Example of Thematic Analysis Findings (Methods for the thematic synthesis of qualitative research in systematic reviews by (Thomas & Harden, 2008)**)** [6]

[5] For further detail on Iterative Best Evidence Synthesis go to http://www.educationcounts.govt.nz/publications/series/2515
[6] This paper can be accessed online at http://roar.uel.ac.uk/1635/

"…A thematic synthesis [combined] the studies of children's views and identified key themes to explore in the intervention studies. Most interventions were based in school and often combined learning about health benefits with 'hands-on' experience. The studies of children's views suggested that fruit and vegetables should be treated in different ways, and that messages should not focus on health warnings. Interventions that were in line with these suggestions tended to be more effective. Thematic synthesis enabled us to stay 'close' to the results of the primary studies, synthesising them in a transparent way, and facilitating the explicit production of new concepts and hypotheses."

To see additional examples look at UNICEF's Evaluation of Humanitarian Action report, which can be accessed at: http://www.unicef.org/about/execboard/files/Colin_Kirk-EHA_synth_paper_19_May.pdf

6 Typical methods involved in evaluation synthesis

The typical evaluation methods are described in four areas: (1) locating potentially relevant studies, (2) developing a theory of change, (3) assessing the quality of data sources, and (4) summarising data.

6.1 Locating potentially relevant studies

6.1.1 Searching databases

Synthesis evaluations often search databases of research articles using keywords to find potentially relevant studies. In addition citation searches (e.g. Google Scholar) can be useful – these search for any article or report which has cited a particular reference. Finding copies of reports can be difficult, especially without access to an academic library with journal subscriptions, or for unpublished reports. A main issue is obtaining an unbiased set of potentially relevant studies. This requires specific strategies to counteract publication and dissemination biases. Publication status should never be used as an inclusion criterion.

6.1.2 Key Informant Interviews

Leading researchers, policymakers, and practitioners can provide valuable advice about relevant research and evaluation studies and might even be able to facilitate access to copies of material. If significant amounts of their time are required, this should be factored into the budget of the synthesis evaluation.

6.1.3 Research mapping

Research mapping (sometimes referred to as evidence mapping) sets out the available research evidence. This process can include identifying studies by applying the chosen inclusion and exclusion criteria and then identifying them with a detailed coding. For example a study could be included or excluded based on its methods or its focus, and then once it is decided to include the study in the synthesis, it is then coded (i.e. given descriptive labels) that help to organise and analyse the data. 3ie also refer to gap maps, which shows where there are gaps in the evidence.

6.2 Assessing the quality of data sources

There are several methods for assessing quality of data sources. A hierarchy of evidence approach prioritises evidence from experimental research (using a randomly assigned control group) or quasi-experimental research (using a credible created comparison group). For an example of this see the What Works Clearinghouse (WWC) Guide which can be found at http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_draft_standards_handbook.pdf. The WWC hierarchy applies to a narrow set of questions about programme impact. Other tools include the Cochrane risk-of-bias framework and the Newcastle-Ottawa scale (the latter is for nonrandomized studies). The Mixed Methods Appraisal Tool assesses the use of

qualitative and quantitative data used in studies. It can be accessed here http://mixedmethodsappraisaltoolpublic.pbworks.com/w/page/24607821/FrontPage

6.3 Summarising data

There are three main approaches for summarising data. First, a meta-analysis is a statistical method for combining numeric evidence from experimental (and sometimes quasi-experimental studies) to produce a weighted average effect size (EPPI Centre, 2007:14-16). It can be used to compare alternative interventions or to compare an intervention to a control or comparison group. A Forest Plot shows the information from the individual studies that went into the meta-analysis. The plot allows readers to see the information from the individual studies that went into the meta-analysis at a glance. It provides a simple visual representation of the amount of variation between the results of the studies, as well as an estimate of the overall result of all the studies. Annex 2 provides an example of a Forest Plot. Meta-analysis is also used with purely observational data (e.g., correlations, rates, trends). Other important issues are: variation among studies (heterogeneity) and potential reasons for those differences (moderators and mediators). Meta-analysis is not limited to comparing alternative interventions or comparing intervention to a control or comparison group.

Second, n*arrative empirical synthesis* brings together the results of empirical research that are in a narrative (i.e. words) and tells a story. For example, while a meta-analysis presents statistics, a narrative empirical synthesis is descriptive. It is appropriate for use with diverse evidence.

Third, c*onceptual synthesis* brings together different understandings to create a new concept or theory ((EPPI Centre, 2007, pp. 14-16). The mixed methods tool mentioned above also provides guidance on combining the synthesis of qualitative and quantitative studies – either sequentially or convergent.

7 Critical issues when planning and managing an evaluation synthesis

This section covers particular challenges that may be encountered in relation to impact evaluations. It draws on the quality criteria set out in the NEPF, specifically: relevance and timeliness, legitimacy, credibility, ethics and trade-offs.

7.1 Relevance and timeliness

When planning a synthesis evaluation, time is a factor both in how long it may take to implement the evaluation and when answers are needed to respond to management questions. For example, Iterative Best Evidence Synthesis may be more time consuming than a Systematic Review because of its participatory and iterative processes, however its inclusive process may result in a more relevant and credible study. While a Rapid Evidence Assessment might provide timely data, the type of in-depth findings provided by a Systematic Review may provide more relevant findings.

Relevance also has to be determined in terms of whether the scope of an evaluation synthesis coincides with the scope of the policy or practice issue(s) in hand. In the evaluation synthesis world, scope for a review is defined in terms of the acronym PICO: What is the population (or sub-groups), the intervention in question, the comparison groups (where appropriate), and the outcomes being assessed? This helps to determine the external validity of the reviews. In other words does the evidence that has been synthesized apply to the ‘real’ world context in which a policy or practice issue takes place? Also care has to be taken as Interventions that are effective in improving some outcomes may have unanticipated adverse effects or no effects on other important outcomes.

7.2 Legitimacy

Engaging key stakeholders who need to use the findings or believe in their credibility is important to ensure use of the synthesis. The evaluation process should address appropriate levels of participation at key points during the process. This may include involving key stakeholders in identifying key questions, participating in the development or review of the design, or engaging relevant stakeholders in identifying studies for consideration in the study or in the selection criteria.

7.3 Credibility

There are several decision points that will affect the credibility of the synthesis. First, decisions around what studies to include and exclude (especially if non-experimental research is automatically excluded or even quasi-experimental research) is a critical decision that needs transparency. Second, the study needs to consider the extent that it addresses contextual variation. In other words, for whom and in what circumstances does 'x' work, or does it not work? Further, it is important to understand how the synthesis addressed the implications of variations by understanding how the interventions chosen for the study were actually implemented. That is to say the synthesis needs to explore implementation aspects as well results. Credibility is especially important with qualitative synthesis, where it is an alternative measure of quality to internal validity.

There are several potential methodological challenges for evaluation synthesis reports. First, there may be challenges to the internal validity (i.e. can one draw conclusions about cause and effect). Second, there may be questions regarding the synthesis study's statistical conclusion validity (i.e. are the inferential statistics capable of detecting cause-and effect relationships). Synthesis is the combined findings of what all the different measurement tools actually represent or measure the construct being investigated. Third, the external validity needs to be considered, i.e. whether one can generalize results to particular populations, settings, or time periods (Shadish, Cook, & Campbell (2002). The newer version of this book has a chapter on research synthesis.

The programme manager should ensure that the report has high quality. High quality is about external and well as internal validity. High quality requires certain tests of evaluation design, execution, analysis and reporting. This would include ensuring the trustworthiness of the results (methodological quality), the appropriateness of the use of the study design (methodological relevance); and appropriateness of the focus of the research (topic relevance).

7.4 Ethical issues

DPME guidelines on ethics clearance and other ethics procedures for evaluation should be followed at all times. When an evaluation synthesis uses reports that are already in the public domain, it does not need ethical review. However if the synthesis involves re-analysing detailed data from previous evaluations, then attention is needed to issues of anonymity and confidentiality. The data can: (1) be anonymised either by being stripped of identifying information or by never having contained identifying information; (2) have had identifying information removed and replaced with a code, and the research team does not have access to the code, (3) contain information that could potentially be linked together to identify participants, or (4) contain identifying information on participants. Research that involves data that have been anonymised or for which the researcher does not have access to identifying information does not require an ethics review.

7.5 Trade-offs

More comprehensive evaluation synthesis will take more time, especially to locate and access original evaluation reports, and to review them for quality and relevance, and then extract data. More participatory approaches, involving active end-user groups or the iterative best evidence synthesis form will also require additional time.

Traditional systematic reviews are sometimes seen as following an intentionally inflexible methodology with the aim of assuring high reliability (Pawson, Greenhalgh, Harvey, & Walshe, 2004). However the SR methodology is flexible and protocols lay out plans that can and should be changed if better concepts or approaches are developed along the way (see the Cochrane Handbook).

Social interventions are complex and attempts to measure what works for whom, when and where using a rather rigid systematic review may result in a homogenised answer of 'to some extent'. While the study may meet high standards for quality and rigour, the review may not provide specific details of what works in what circumstances (Pawson, Greenhalgh, Harvey, & Walshe, 2004)). However moderator analysis in meta-analysis can assist to answer the question "what works best for whom in what circumstances?"

If time is a concern, then Rapid Evidence Assessments may well be the best option, with some trade-off between rigour and time.

8 Typical costs

This section provides some basic guides for helping to determine the size of an evaluation budget. However each context will be unique and require specific budgeting discussions and decisions.
The programme manager has a key role in ensuring that the scope of what is promised by evaluators, or expected by the programme manager, is realistic for the amount budgeted; as over ambitious and under budgeted scope of work is likely to yield a weak base of evidence and an unused report.

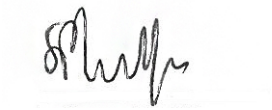
Budgeting for an evaluation is dependent on numerous factors. A general 'rule of thumb' is that an evaluation should be between 0.1% to 5% of an intervention's budget. However this depends on many variables, such as the size of budget, with large programmes needing proportionally less, the amount of credible data already collected, the timeline to collect data, the amount of field work that needs to be done, and other contributing cost factors.

Another common guidance is that 5% and 15% of total programme budget should be set aside for monitoring and evaluation (M&E), or 3-10% of annual budget. However this refers to M&E rather than evaluation specifically; and it is likely in many programmes that routine monitoring will consume most of the M&E budget.

Broadly speaking, the following factors weigh heavily on cost:

- The scope of the information requirement and wideness of the search review required, e.g. how focused is the search;
- The degree of credibility required of the information results (which influences the spread of the search required) versus time available.

Signed



_____

Dr Sean Phillips
Director-General
The Presidency: Department of Performance Monitoring and Evaluation
Date: 31 March 2014

Annex 1: Glossary

- **Research Synthesis** - is the practice of systematically distilling and integrating data from a variety of sources in order to draw more reliable conclusions about a given question or topic.
- **Systematic Reviews** - uses transparent and replicable procedures to minimize bias and error at each step of the review process.
- **Meta-analysis** - summarizes the data from a systematic-review. It produces a weighted average of the effect of an intervention, taking into account the strength of results and the sample size of studies which have met inclusion criteria.
- **Moderator Analysis** in meta-analysis can assist to answer the question "what works best for whom in what circumstances"
- **Narrative Empirical Synthesis** brings together the results of empirical research that are in a narrative (i.e. words) and tells a story.
- **Conceptual Synthesis** brings together different understandings to create a new concept or theory
- **Meta-ethnography** - focuses on answering questions about how concepts are understood, by translating concepts across different studies to produce an overall model.
- **Realist Synthesis** - focuses on answering questions about "What works for whom in what circumstances" and draws on a range of evidence to build and test a model of how causal mechanisms work in particular contexts
- **Iterative best Evidence Synthesis -** actively engages with stakeholders to optimise the usefulness of the synthesis and encourage ownership by involving them in identifying what would constitute an improvement in outcomes, and implications for practice
- **Rapid Evidence Assessment** - a pared down version of a systematic review and draws on the same principles. It is however a 'lighter touch' approach that enables the review to take place more quickly
- **Thematic analysis** - focuses on a specific theme and refers to evaluations that cross"- multiple programmes and projects.

Annex 2: Useful web resources

Databases of systematic reviews (and in some cases REAs):

- Collaboration for Environmental Evidence (www.environmentalevidence.org/)
- Evidence Encyclopedia (http://www.bestevidence.org/)
- EPPI-Centre Evidence Library (http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=56)
- Cochrane Collaboration (www.cochrane.org)
- Campbell Collaboration (www.campbellcollaboration.org)
- 3ie Systematic Reviews Database (http://www.3ieimpact.org/en/evidence/systematic-reviews/

- Evidence for Policy and Practice Information and Co-ordinating Centre:

http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=62
ADD the Collaboration for Environmental Evidence, and Cochrane Library

- Cochrane MECIR standards and IOM standards:

http://www.editorial-unit.cochrane.org/mecir
http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews/Standards.aspx

Annex 3: References

EPPI Centre. (2007). *EPPI Centre Methods for Conducting Systematic Reviews.* London: EPPI Centre.

Munro, S. A., Lewin, S. A., Smith, H. J., Engel, M. E., Fretheim , A., & Volmink, J. (2007). Patient adherence to tuberculosis treatment: a systematic review of qualitative research. *PLoS medicine 4 (7), e238.*

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: An Introduction.* Retrieved December 24, 2013, from CCSR: http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, pp. 8(45), pp. 1471-2288.

UK Government Social Research Service. (2007, October). *The Magenta Book: guidance notes for policy evaluation and analysis.* Retrieved December 24, 2013, from UK Civil Service: http://www.civilservice.gov.uk/wp-content/uploads/2011/09/the_complete_magenta_book_2007_edition2.pdf

.

Annex 4: Theory of change

The Theory of Change (TOC) or programme theory describes the causal mechanism of how activities and outputs (such as meals delivered to needy school children) will result in the anticipated outcomes (eg improved concentration in school), and impacts (eg improved grades) and the assumptions involved. There can be multiple TOCs that describe the programme. For example different theories can show how the intervention works in different contexts, or at different stages of the intervention, or even for different intended impacts (Interaction: 6-7). The ToC should be established during the early planning stages of a policy or programme.

A ToC can help to identify which impacts are likely to be achieved during the timeline of an evaluation, and what else should be examined in the evaluation – activities, context, and intermediate outcomes. Also the ToC helps to identify what needs to be in place – people, agencies, activities, mechanisms, resources – for the impact to be achieved. It can also be used to analyse the evaluation results. If a programme has not worked, the ToC can help to identify whether this is due to failures in implementation or because the theory of change does not work. If a programme has worked, the ToC can help to identify what is needed to repeat this success at another time or another site.

Developing a TOC is best done through a combination of a desk review of existing documentation, a literature review of research and evaluations of similar programmes including systematic reviews, observing the programme (if it is already running) or similar programmes, and talking with stakeholders about how they think it works. It often involves an iterative, participatory process with programme developers and/or staff and other relevant stakeholders.

The Theory of Change can be represented in the form of a logframe, a results chain or an outcomes hierarchy.

Outcome Mapping is a particular approach to developing a Theory of Change which is particularly suitable when a programme does not directly produce the intended results but works through influencing the behaviour of people in another organisation[7].

Useful Tip
For more information on different approaches for developing a TOC, check out
http://betterevaluation.org/plan/define/develop_logic_model

For evaluations under the National Evaluation Plan, the theory of change should also be expressed in the form of a logframe. Annex 1 of Guideline 2.2.3 on Planning Implementation Programmes provides an example of a theory of change for the National School Nutrition Programme, as well as a model of a logframe. The logframe should include:

i. Indicators at different levels, baselines and SMART targets, where appropriate, as part of the logical framework;
ii. The key assumptions and risks which underlie the results chain;
iii. Key outputs and related activities required to achieve the desired outcomes;
iv. A summary of the human and financial resources (inputs) needed to achieve the outcomes and impacts.

If a theory of change and logframe does not exist, then one of the first activities in the implementation evaluation should be to derive one, based on the understanding of how the programme or policy was established. As part of the evaluation, changes to the TOC and logframe may be recommended.

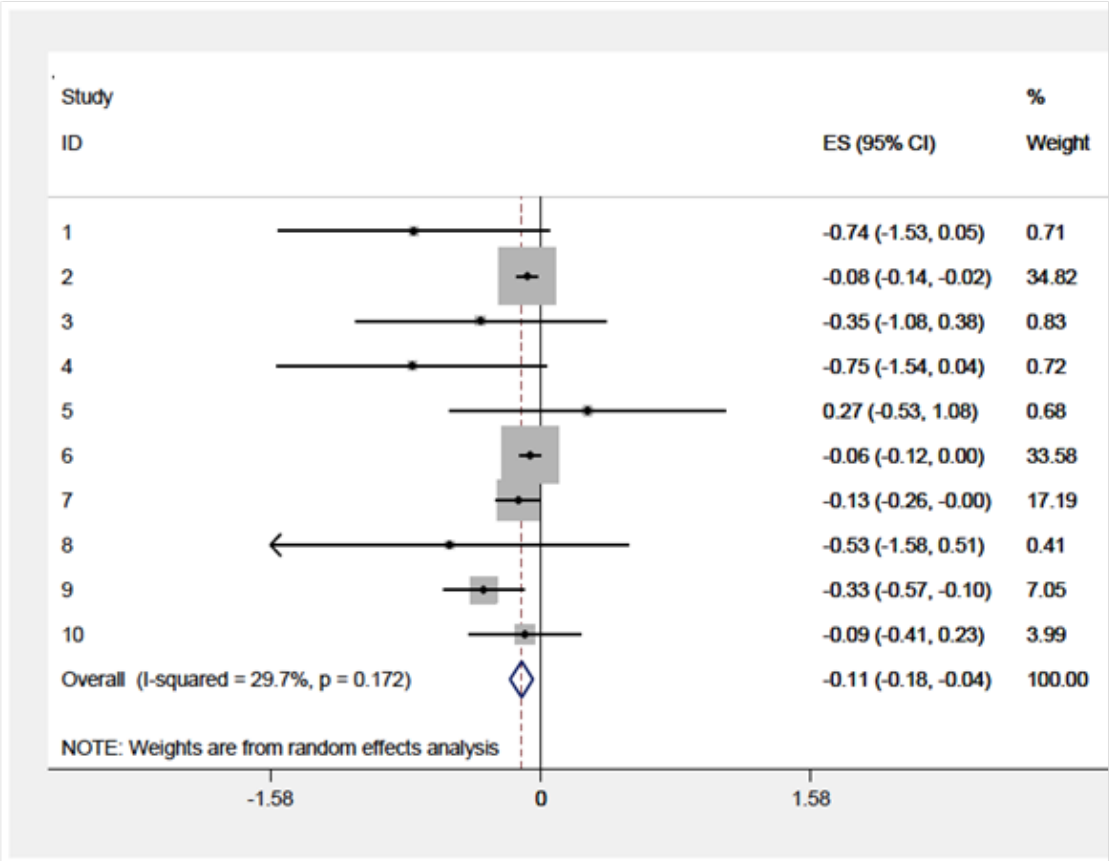
[7] For further information go to http://www.outcomemapping.ca/

Annex 5: Example of Forest Plot

An Example of a Statistical Meta Analysis

The Effect of Conditional Cash Transfers on Child Labour

The diagram below is called a *Forest Plot*. It presents the results of 10 controlled evaluations of the effectiveness of conditional cash transfers to poor households in low and middle income countries in terms of reducing the incidence of child labour. These 10 studies have been tested for *homogeneity* (i.e. being sufficiently similar to be truly comparable), which allows their results to be aggregated into an overall effect estimate for all of the ten studies combined.



Source: Kabeer, N., Piza, C, and Taylor, L., 2012, *What are the economic impacts of conditional cash transfer programmes?* A systematic review of the evidence, London, 3ie.

The forest plot presents the average effect of the conditional cash transfer of each evaluation (the dot in the middle of the horizontal lines), its standard error, and the corresponding confidence interval (the length of the line either side of the dot representing the average effect). The solid vertical line crosses the horizontal axis at zero, indicating no intervention effect. Nine of the evaluations indicate a positive effect (i.e. the average effect is to the left of zero), and one study (number 5) reports a negative effect (i.e. the average effect is to the right of zero).