



Background Note

June 2013

Adapting systematic reviews for social research in international development: a case study on child protection

By David Walker, Gina Bergh, Ella Page and Maren Duvendack

In the past decade, policy-makers and practitioners have, increasingly, turned their attention to ensuring that thinking and practice around international development are based on reliable and broad caches of evidence. Confronted with more primary and secondary evidence stocks than ever before, they are looking for ways to take advantage of this situation, while preventing the potentially stultifying and debilitating effects of having contradictory, varied and excessive evidence stockpiles. Fuelled mainly by this rapid expansion of the evidence base, systematic reviews (SRs) have emerged for many as a promising answer to the otherwise confusing and insurmountable data assimilation challenge. For others, the SR methodology raises more questions than it answers. While an emerging body of evidence discusses these issues, there is little evidence that goes into significant detail about the SR method's relevance to qualitative and complex social topics or interventions.

This paper aims to close this gap by discussing lessons from an 'adapted systematic review' on the links between approaches on child protection and income poverty. The review, conducted by the Social Development programme at ODI, and supported by the Oak Foundation as part of a larger programme of work, was confronted by a series of specific obstacles relating to the prevalence of integrated programming in these sectors, as well as a relatively fragmented literature base – itself often built on variable and mixed-methods foundations. Together these obstacles present three core challenges:

1. managing disciplinary diversity and evidence gaps
2. handling quantitative, qualitative and mixed data and an overly stringent application of guidelines
3. the need for a suitably adapted method of synthesising evidence.

Each of these are addressed individually in this paper, which culminates in an assessment of the broader implications for different actors in the international development community.

Understanding the processes of systematic reviews

The interest in SRs began, for the most part, with bilateral donors such as the UK Department for International Development (DFID) and AusAID, as well as 'knowledge brokers' such as 3ie, and is only likely to expand. The recent development, for instance, by 3ie (the International Initiative for Impact Evaluation) of a summary database of 200 SRs is testament to the fact that SRs are likely to become crucial currency in a variety of evidence and policy debates. As the 'first wave' of SRs passes and associated lessons are incorporated into institutional research procedures and the wider international development community, those without a critical respect for SRs may be likely to use or undertake SRs themselves that are at odds with a research community intent on following strict SR procedures. At worst, ignorance of emerging lessons may even harm development outcomes – that is, if policy decisions and negative impacts occur on the basis of faulty impressions of interventions provided by ill-treatment of an evidence base.

Given the likelihood that SRs are only likely to expand and formalise in the coming years, there is a clear need to raise healthy awareness about their

The Overseas Development Institute is the UK's leading independent think tank on international development and humanitarian issues. ODI Background Notes provide a summary or snapshot of an issue or of an area of ODI work in progress. This and other ODI Background Notes are available from www.odi.org.uk

Table 1: Timeline of SR development in relation to uptake in international development research

1980s	1993	2000	2007	2008	2010
<p>Early work on systematic review in field of psychology</p> <p>Aimed to minimise reviewer bias through methodological explicitness and transparency.</p> <p>General guidance and methods developed as antidote to traditional narrative review by content experts, which was 'subject to criticism for its lack of transparency'.</p>	<p>Cochrane Collaboration established</p> <p>Building on work of Archie Cochrane.</p> <p>Comes to dominate medical sciences research.</p> <p>Evidence-based systematic review, relying primarily on randomised control trials (RCTs).</p> <p>Translates early general SR requirements into specific 'methodological rules'.</p>	<p>Campbell Collaboration established</p> <p>In partnership with Cochrane Collaboration; applies their methods and principles to social / behavioural sciences.</p> <p>Conducts and supports SRs of interventions in education; criminal justice; social work.</p>	<p>International Initiative for Impact Evaluation (3ie) established</p> <p>In the US, hosted by Global Development Network.</p> <p>Applies SR methods to international development research.</p> <p>Builds on the work of the Evaluation Gap Working Group (Center for Global Development) responding to increasing donor attention to evidence.</p>	<p>International Development Coordination Group established</p> <p>A Campbell Collaboration group, aiming to register and support systematic reviews in international development.</p> <p>Subsequently (2009), Chris Whitty is appointed Director of DFID Research; with his background in clinical research, he launches the systematic review programme at DFID.</p>	<p>DFID, AusAID and 3ie commission pilot systematic reviews</p> <p>Institutions acting as quality assurers for pilot SRs:</p> <ol style="list-style-type: none"> 1. Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre). 2. International Initiative for Impact Evaluation (3ie). 3. Collaboration for Environmental Evidence (CEE).

purposes and pitfalls amongst the non-specialist audiences likely to draw on the emerging SR evidence base, be they legislators, NGO programme managers, executives within donor organisations or others. These audiences will be increasingly pressured to commission or reference such reviews for their ongoing programming, practice or policy work. They will need to be confident in constructing relevant terms of reference, managing outputs, and perhaps most importantly, asking the appropriate core research questions in the first place.

What are SRs?

Essentially, an SR is both a research process and methodology, designed to address all the evidence on a given question (usually regarding intervention impacts) without bias or unforeseen exclusions, with a focus on replicability, accountability and transparency (see Box 1 for interpretations). SRs can, therefore, be viewed as a 'public good' in that they promise consummate assessments of the evidence (DFID, 2011). As a result of this broad definition, SRs have earned a reputation as being authoritative assessments that assemble and synthesise all available research evidence (Waddington et al., 2012).

Emerging from the health sector and the drive for evidence-based medicine in the 1980s and 1990s, and only recently making the 'jump' to international development research (see Table 1), a completed SR would demonstrate an exhaustive list of literature on a given question. The SR would then provide comprehensive detail of the databases and journals searched and accessed, and a full justification for those purposively included and excluded. The assembled information is then assessed for eligibility in relation to the core research question using one of a range of possible quality assessment standards. Summary and synthesis methods – selected accord-

Box 1: Selected interpretations of SRs

- A systematic review summarises 'what works, where and for whom' (Gasteen, 2010).
- A systematic review is not a mechanical process whereby policy prescriptions that can be applied across contexts will emerge (Gasteen, 2010). It represents a robust, high-quality technique for evidence synthesis (DFID, 2013).
- Systematic reviews should be viewed as a way to find a robust and sensible answer to the research question' (Hagen-Zanker et al., 2012)
- Systematic reviews aim to be as objective as possible about research to remove bias; too many systematic reviews are too long and too complex (Boaz et al., 2002).
- The challenge of systematic review is vital to evidence-based policy development (Pawson, 2006).

ing to their suitability to the subject matter – would then be applied to the evidence that qualifies.

The SR approach within the social and behavioural sciences

The advantages and disadvantages of systematic reviews have been well reviewed (White and Waddington, 2012), but the question of how the method might apply to generic international development concerns has been addressed to a far lesser extent (Mallett et al., 2012). The sector needs, therefore, to further appreciate how the SR research method applies to international development issues particularly its social and behavioural aspects.

For instance, a number of SRs commissioned or piloted by DFID, AusAID and 3ie have yielded disappointing results. This suggests that SRs have proved highly effective and relevant to the medical sciences, particularly for collating and cross-comparing highly technical quantitative studies, such as randomised

control trials (RCTs), but the same methods may not work when accessing and synthesising the evidence on questions around social policy, which are based largely on qualitative or mixed-method forms of enquiry. As a result, implementing teams face three central challenges to the development of adapted systematic reviews in this area.

The first challenge that is more specific to the social and behavioural side of SR approaches concerns the overly stringent application of quantitative guidelines for the research method and information synthesis (Mallett et al., 2012). Given that the prevalence of robust scientific evaluations in international development and the social sciences is more limited than in the health sciences where SRs originated, quantitatively-orientated rules become less important. Indeed, as we discuss later, the social sectors are relatively dominated by qualitative and mixed-methods research. This calls for a re-think of the application and functionality of SRs to social-science questions, especially in the social sectors.

A second major challenge is the inter-disciplinary tendency of international development and social science research. Here the concern is the balance and weighting not only of quantitative, qualitative, and mixed research, but also the relative value of different disciplines as they relate to the research question. Similarly, these concerns will vary according to the core goals, values and approaches espoused by the primary commissioning agent(s). An agricultural-science institute, for instance, will display certain disciplinary predispositions in relation to a commissioning agent whose primary interest is gender-based violence. These aspects include questions that relate to the inclusion and exclusion criteria adopted in an SR – which, during implementation, can cause vociferous and/or distracting debate if not clearly constructed and communicated to the core research team. However, this challenge presupposes that there is written or readily available evidence to review in the first place – a concern particular to low- and middle-income countries, given the limited capacities and resources made available to them, and one that must be accounted for somehow.

A third major challenge specific to social research in international development centres on the need for a suitably adapted evidence-synthesis method. Although synthesis methods vary, they provide a conceptual and practical basis upon which to construct new insights from established evidence, rather than focusing merely on comparing and summarising evidence. The greater diversity of contexts, outcomes and intervention designs in some areas of social research renders traditional SR synthesis methods problematic and robust cross-comparison

unfeasible. On the one hand, many evaluations suffer from limited or poorly applied programme theory that disables attempts to determine the wider context and determinants of causalities. On the other hand, many SRs themselves overlook or fail to provide a detailed description of interventions and how they were implemented.

As highlighted by Mallett et al. (2012) this is exactly the kind of information that is crucial for thinkers and practitioners in international development: without an understanding of the specific drivers and enabling environment that support an intervention, there are limits to how they can be reapplied. In addition, if a series of interventions have identical or very similar outcomes, it will be impossible to undertake further comparison via a cost-benefit analysis without more detail of the allocated resource inputs (Snilstveit, 2012).

A final challenge is one that applies to all SRs, but that is nevertheless rarely voiced. It concerns the up-stream activity of research up-take, where all stakeholders have a role in adjusting the final message whether they recognise that or not. A somewhat polarised debate has emerged in the blogosphere among an expanding cadre of so-called ‘truth junkies’ on one side of the development-evidence debate, and a gathering of ‘politics junkies’ on the other. This debate is characterised by the extent to which one promotes SR evidence as a compression of knowledge that is relatively free from the constraints of political interests, or whether one promotes SRs as part and parcel of a deeper ideological landscape.

Thinkers and practitioners should recognise a grain of truth in both – and consider this in their negotiations with commissioning parties when designing SRs. However, they should also realise that both sides of the debate have failed to realise a broader concern: SR results, particularly those examining social and behavioural dynamics, attempt to synthesise ‘what works’ without an appreciation of complex systems. A focus on either simple or complicated development interventions often obscures the fact that complexity is a separate and important issue (Ramalingam and Jones, 2008), one that emphasises indecipherable patterns and unpredictable feedback loops that cannot be captured in SRs. This is a persistent challenge that needs to be recognised in the construction and communication of SRs, as well as in the international development community as a whole.

The SR approach applied to child protection

The issues discussed above come to the fore in SR projects that address relatively marginalised and sensitive subjects such as child protection, maltreat-

ment prevention, rights and social institutions. This case study SR on child protection was no exception, and sought to examine two related questions. First, to what extent do child-protection interventions have an anti-poverty focus? Second, what contribution do anti-poverty components make to the quality of child-protection interventions?

The ultimate programme goal is to enhance the impact of global efforts to tackle the alarming scale of child-protection violations in developing-country contexts, promote understanding about the role of poverty as a driver of child ill-being, and inspire action to improve children’s well-being and capabilities. This goal recognises that while child protection and poverty are intimately linked, child protection remains marginalised in mainstream development dialogues and action. As a result, few actors appreciate the importance of integrated interventions in tackling child-protection deficits – in part because of a limited and fragmented knowledge base. The role of SRs is key in developing a verifiable resource base to support decision-making in child-protection activities.

Given that the evidence base on child protection largely reflects the major SR challenges already outlined, the programme team chose to undertake an ‘adapted systematic review’ (ASR) with a focus on innovative methods to assess the diverse and fragmented evidence base in question. Naturally, there is a question of chronology: how and when is it decided that an ASR, rather than a SR, is required? Expert knowledge and existing literature reviews can help to inform this decision. As discussed later, this case study included a literature review that indicated to researchers the quality and quantity of the overall evidence base (see Figure 1).

Challenge 1: managing disciplinary diversity and evidence gaps

As stated earlier, international development and social-science research are often challenged by having to consider an enormous array of disciplines and their associated methods, methodologies and terminologies. The child-protection sector, for instance, involves both prevention and response-based approaches – each of which have their own focus on rule of law and legislative practice, the supply and demand of essential services, welfare services, psychosocial support, cultural change and much more. Consequently, a broad range of disciplines are correlated to these varied interventions, as are the researchers tasked with their evaluation. Given that this particular systematic review sought consciously to reach across ‘child protection’ and ‘child poverty’ communities to clarify linkages and

gaps, the issue of disciplinary diversity was all the more apparent.

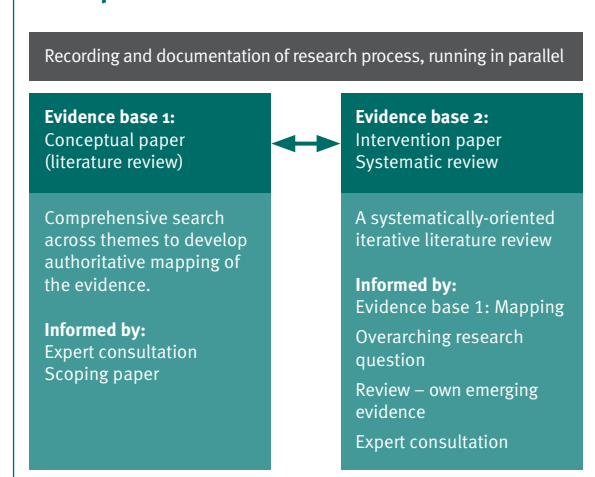
The issue of evidence gaps is of equal importance: in a meta-analysis (review of reviews). Mikton and Butchart (2009) assess the geographical distribution of evidence relating to outcome interventions to prevent child maltreatment. Findings show that there are very few studies of such interventions in low- and middle-income countries compared to high-income countries – constituting only 0.6% of the total available academic evidence-base. These issues are of critical concern, as a comprehensive SR process is obliged to take significant steps to confirm, recognise and account for the scale of multiple and disproportionate evidence-bases.

Responding to disciplinary diversity and gaps

The response from the implementation team was to develop a hybrid approach to the SR methodology, which drew firstly upon a more ‘traditional’ scoping review and conceptual paper, and secondly, parallel expert-panel inputs to inform findings and fill gaps. An initial scoping of the evidence was carried out to provide an indicator of the types of disciplines (and their associated journals), organisations and evidence gaps relating broadly to the topic. This scoping paper then informed a parallel and mutually-informative research process, which aimed to consider both the conceptual linkages between poverty and child-protection violations (evidence base 1) and the position and impact of anti-poverty elements within child-protection interventions (evidence base 2) (see Figure 1).

The two components outlined in Figure 1 inform each other. The conceptual paper allowed the

Figure 1: A representation of an adapted method used in the systematic review on child protection



implementing team to identify key themes of interest, while the SR drew on insights from this process to define search foci and potential information sources. Taking this dual approach offered findings of relevance across thematic areas of interest, and highlighted evidence gaps and areas for further research. This approach is not necessarily new: the stages of a Campbell Systematic Review include activities such as ‘hand-searching’ of key journals, checking bibliographic databases and undertaking personal communications with experts in relevant fields (Petrosino et al., 2001), but the level of depth and formality of the method as used in the intervention-focused paper is not made clear as a formal exercise in itself in current SR guidelines.

The development and engagement of the expert panel is the second area of promising practice that was developed in parallel to the scoping and literature-review papers. In the case of the child protection ASR, the implementation team was committed to including the insights and recommendations of the broad variety of programme, policy and practitioner individuals that work across child-protection and child-poverty communities. As established SR methods do not cater explicitly for this process, the implementation team sought to use the expert-panel input method promoted by the Canadian Health Services Research Foundation (CHSRF). This approach provides a structure through which an expert panel can feed their professional perspectives routinely into a research process.

While these collected viewpoints are not peer reviewed to scientific standards, they do offer a richer picture of a very diverse reality, while also catering for potentially under-studied research areas where established scientific evidence may be insufficient or inappropriate. Of course, these processes themselves are not immune to bias: politics and personal networks based on disciplinary focus can adjust outcomes. The crucial issue is the maintenance of transparency and accountability by capturing the justification of decisions and processes wherever possible.

While the Mixed Methods Assessment Tool (MMAT) classification system does not cater for participatory inputs in the inclusion phase itself, it can be argued that there is no harm in including direct testimony from stakeholders in ‘adapted’ SRs to assist search protocols – provided this testimony is organised appropriately. The expert-panel consultation process included in the CHSRF method does, for instance, attempt to circumvent input bias through the development of deliberative structures chaired by a management committee.

The role of the committee in the CHSRF approach would be to govern and judge the process by ensuring that the balance of views and roles of the various

actors in the process are clearly-defined, transparent and accountable. The process is managed through consensus-building methods and facilities that can negotiate value disagreements between stakeholders – whether they are researchers, intermediaries or end-users (policy-makers). Although this is not the prescriptive process found in SR approaches, it can be argued that synthesis reached via consensus of different parties is far more objective than synthesis based on the inclusion and exclusion determined by a small team of researchers alone – as is often the case with SRs.

In contrast to classic SRs, CHSRF expert-panel methods do not promise the same results for similar initiatives. By allowing the incorporation of context and causal mechanisms during interviews and deliberation, expert-panel inputs help to redress some of the challenges discussed earlier. CHSRF-inspired panel inputs give more weight to context-making and, therefore, provide an excellent complementary service to SRs processes – particularly those less interested in whether an intervention ‘worked’, and more interested in why it worked (provided there is scope to reach the implementation personnel in question).

It must be noted however, that the CHSRF expert-panel consultation process requires committed oversight and skilled management as it demands the facilitation of various types of knowledge, as well as the oversight of the parties that are presenting them. For example, deliberative processes can quickly become controversial and eventually reach an impasse based on intractable values and beliefs (van Eeten, 1999). Reports reveal participant numbers that often range from 9 to 18 persons/actors (McGlynn et al., 1990) but there are cases of up to 30 actors being involved (e.g. National Institute of Health and Clinical Excellence processes). This deliberative process may itself (re)reveal the embedded nature of the impasse, rather than attempt a realistic resolution. Where the group is too large, too much time may be spent reconciling diverging viewpoints and interests.

To the limited extent that time and resources were available to incorporate expert input along the lines above, these pros and cons were equally evident. A panel of 8-12 sector professionals was established, all of whom were selected on the basis of their varied disciplines, thematic foci and organisational positioning. In the early stages of the ASR process, the implementation team found that, perhaps unsurprisingly, the level of commitment and participation was higher.

As the study progressed, requests for input required more intensive engagement processes and resources, which were increasingly tied up in the management of already existing and emerging outputs. In addition, to augment the expert-panel inputs, it was decided at the outset that an e-survey

would be undertaken to gain wider perspectives – primarily from professionals in low- and middle-income countries. Ultimately, the guidance, triangulation and contextualisation from the original core expert panel proved most important in the early stages of the research.

As the findings from the e-survey emerged, and the implementation team moved forward onto the systematic review itself, there was little time available to provide full feedback to the expert panel. This is not to say that other SR projects should not use this model – only that they should be aware that the full and continuous engagement of an expert panel requires significant resourcing and management of expectations, particularly if the expert panel is also required to feed into the lengthy and complex synthesis stage.

Challenge 2: handling quantitative, qualitative and mixed data and overly stringent application of guidelines

SR methods have significant limitations when encountering a methodologically diverse evidence-base, in particular, the lack of a clear protocol to deal with qualitative evidence. The fact that the classic SR method has few convincing answers on assessing the quality of qualitative evidence is perhaps unsurprising, given that the process has come from the natural sciences where quantitative evidence dominates. However, it is not for a lack of effort that the research community has been unable to make progress on this front. The issue is that qualitative and mixed-method assessment criteria relating to SR inclusion processes tend to vary widely, and they are often developed within the context of individual SRs, rather than centralised into sanctioning bodies, such as the Centre for Reviews and Dissemination.

Nevertheless, while the discussion about the quality appraisal of qualitative evidence is ongoing, the Cochrane Collaboration has set up a

Qualitative Methods Network to develop this area further. At this stage it is useful to consider which experimental frameworks and guidance notes provide promising guidance on the systematic review of qualitative evidence (see Table 2).

Additional notable examples include the ‘Quality in Qualitative Evaluation’ framework developed by Spencer et al. (2003), which examines mainly the development and implementation of social policy, and the approach discussed by Thomas et al. (2004), which focuses on combining qualitative and quantitative methods of evidence assessment.

The majority of SRs place a strong emphasis on assessing evidence on the basis of an ‘evidence hierarchy’. For the most part, there is ongoing consensus that such appraisal systems are highly problematic as they are based on methods designed primarily to assess quantitative evidence, for which hierarchies of evidence are clearly less problematic (Petticrew and Roberts, 2003). In recognition of the limited utility of evidence hierarchies for qualitative data, the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) advises the appraisal of each piece of evidence according to its Weight of Evidence (WoE) scale. This is a more holistic tool that incorporates a range of questions to test the quality of the evidence, including methodological rigour, relevance, and connection to the core research topic. While the WoE tool does incorporate hierarchical assessment to an extent, this is balanced by an intensive assessment process designed for highly-skilled researchers with extended or open timeframes. It can, therefore, assess multiple aspects of evidence provided that a given SR has extensive time and resources (Gough, 2007).

Responding to methodological variation and stringency

Given the methodologically diverse and fragmented evidence-base in question within the Child Protection ASR and the fact that four child-protection violation

Table 2: A selection of guidance papers on systematic review of qualitative evidence¹

Popay, Rogers and Williams (1998)	The primary question relates to the appropriateness of the methods used. This is followed by a detailed assessment of methodological soundness.
Critical Appraisal Skills Programme (1998)	Ten questions relating to rigour, credibility and relevance.
Quality Framework (2003)	Eighteen questions relating to nine key areas: findings; design; sample; data collection; analysis; reporting; reflexivity and neutrality; ethics and auditability.
Prompts for appraising qualitative research (2004)	Generic set of prompts relating to aspects of reporting and aspects of study design and execution.
Long and Godfrey (2004)	A tool to explore descriptive and evaluative elements of a study. Provides 34 questions across four key areas: phenomenon studied and context; ethics; data collection, analysis and potential researcher bias; policy and practice implications.
Walsh and Downe (2006)	Set of prompts relating to eight key areas: scope and purpose; design; sampling strategy; analysis; interpretation; reflexivity; ethical dimensions; relevance and transferability.

areas were being reviewed, the numbers of papers in question were above average (1,341). In addition, the search team consisted of a relatively expansive team of junior and mid-level researchers, with time and capacities that were not commensurate to the application of the WoE assessment scale. Consequently, the selected quality-assessment method was the Mixed Methods Assessment Tool (MMAT) developed by Pluye and colleagues (2011) at McGill University. The MMAT tool is a relatively new instrument designed by researchers working in health sciences and has been applied in relatively few published studies. Pace et al. (2012) found that the tool offered a high level of reliability, and that disagreements between researchers were easy to resolve.

The MMAT tool was also selected because it was designed to deal with quantitative, qualitative and mixed-methods research within the same assessment scale. The scale goes from ‘zero stars’, awarded where a paper’s methodology is unclear or not explained, to ‘four stars’, which applies if the evidence reports clearly that it is based on high-quality research methods. In experimentation exercises, the child protection ASR research team found the MMAT easier to apply than any of the quality-assessment scales discussed above, and noted that the specific inclusion criteria for assessing mixed-methods research were particularly relevant to the type of evidence we expected to find. The team also felt that applying the MMAT tool to all papers made the discussion of the quality of evidence across disciplines and methodologies easier.

The MMAT takes researchers through a series of questions about the research and a note of their judgements is recorded in a table. Quality assessment was completed in an independent and ‘blind’ fashion by two researchers per assessment paper, with any disagreements between the two reconciled through a discussion, based on which a final code was reached. This was recorded in the software. A simplified version of the assessment criteria is presented in Box 2.

In our review of child-protection violation interventions in four areas (violence, sexual abuse and exploitation, early marriage and inadequate care), a high number of papers assessed using the MMAT scale scored ‘zero stars’. This reflects the very limited or entirely absent explanation of the methods used to produce research papers. The research team also made a decision to exclude papers that scored ‘one star’ as these showed little reporting of the methodology used. Many discarded papers included ‘grey literature’ from NGOs, which may have proved useful. As discussed under challenge 1, it is advisable that research teams develop methods to assess the potential value of such discarded papers and then

consider ‘follow-up’ communications. Overall, only 20 papers, from a reduced total of 398 papers (5%) identified during the search process, scored the highest possible mark of ‘four stars’.

While the majority of the criteria were applicable to the type of research and papers identified, the few that focused on participatory methodologies were difficult to assess – suggesting that adaptations would be beneficial to include participatory research as a category in itself. The implementing team also felt that it was easier for quantitative papers to score highly on the MMAT, not as a reflection of any quality intrinsic to quantitative approaches, but of more standardised and frequent reporting of methods in comparison to qualitative papers. An additional observation was that papers marked as mixed methods-based often provided a strong quantitative description alongside a relatively poorly-described qualitative component – thereby excluding some papers despite their ostensible quantitative merits.

In theory, any gaps in the presentation of chosen methods should, according to the MMAT approach, be followed-up through direct communication. This would be the case, in particular, for promising high-impact evaluations and/or studies of ‘scaled-up’

Box 2: An applied interpretation of MMAT assessment criteria

In qualitative research:

- clear information about the selection of participants
- clear information about how data were recorded and analysed
- appropriate consideration for context, such as using multiple sites and taking the country context into account in the research design
- a consideration of researchers’ influence and reflexivity. We were looking for consideration of how the position of the researcher influences the research and evidence that this has been considered in the design of the research methods (or in the analysis).

In quantitative papers:

- clear description of the selection of participants – the sampling strategy, method of randomisation as appropriate to the research design
- discussion of selection bias, representation and comparability
- clear outline of instruments and measurements/indicators
- complete outcome data (more than 80%).

In mixed methods:

- a clear rationale for applying a mixed methods research design
- integration of qualitative and quantitative evidence to build a complete picture of outcomes.

interventions that have been reframed for non-technical audiences – an issue common to ‘grey literature’. Again however, fulfilling the MMAT criteria with this level of follow-up requires extensive resourcing and planning at the outset of the SR process.

The implementing team also found that mixed-methods papers scored consistently low on the scale. This was the result, primarily, of two separate issues. The first is the aforementioned matter of qualitative methods – in this case having lower reporting standards than quantitative methods. The second issue concerns the frugal reporting and valuation of mixed-methods approaches – such as the relative dominance of a quantitative versus qualitative approach used in the paper, and whether there was any sequencing involved (i.e. whether quantitative work fed into qualitative work and vice versa, or whether they ran in parallel).

Challenge 3: developing a suitably adapted method for evidence synthesis

Given the volume of evidence generated, potentially, by the methodology, the act of presenting findings through synthesis can pose some substantial practical and methodological challenges. A good synthesis of evidence helps to make review findings useful and accessible to policy-makers and other review users. For a systematic review to present results effectively, the findings across all included sources of evidence must be reconstituted cohesively – simply presenting conclusions drawn from an appraisal of individual studies is insufficient.

There are numerous approaches for the synthesis of evidence. One suitable approach depends heavily on the nature of the review questions and the type of evidence that is available. SR findings based on standard approaches to information synthesis, such as lengthy and output-focused ‘narrative review’ approaches, have been deemed to be too broad, incomparable, and excessively research-oriented to generate useful policy recommendations (i.e. findings have a tendency to be more academic than policy relevant) (Pawson, 2006). As shown in Table 3, there are numerous alternatives, each of which has advantages and disadvantages that must be weighed in accordance with the demands of the SR and available resources.

Selecting the synthesis method

The decision to select a particular synthesis method (or combination of methods) is important because the synthesis is the core of any SR methodology. It determines (or is determined by) the research field, research question, the needs of the commissioning body, the capacities of the implementing team and

the resources available. Given the broad range of evidence-types in the child protection ASR, as well as the varied intervention outcomes and the sheer volume of literature gathered, the implementing team selected a combination of narrative synthesis and realist synthesis.

Narrative synthesis was selected on the basis that it accommodates both quantitative and qualitative information, and is somewhat more descriptive than meta-analysis of the contextual and causal mechanisms behind results. It is the method used most commonly in the social sciences. It also goes further than the mere summative description common to narrative review and meta-synthesis. Moreover, narrative synthesis can be standardised and reproduced. It is based on a textual analysis of the evidence and suits a variety of studies by allowing an in-depth examination of their similarities and differences using a fixed format (Barnett-Page and Thomas, 2009). The approach thereby allows a good understanding of the context and characteristics of the studies included in the review.

Table 3: A selected summary of SR synthesis methods

Quant. data	Meta-analysis	A statistical quantitative method that performs calculations to reveal the ‘best buys’ in terms of successful interventions. Strongly biased in favour of randomised control trials.
Qual. / mixed-methods data	Meta-ethnography	An interpretive approach for combining the findings of ethnographic research. Relies on creating a list of themes or metaphors, juxtaposing them and determining how they are related.
	Narrative synthesis	Puts forward ‘exemplary cases’ in a narrative, textual analysis of the evidence. It presents a highly summarised picture of successes based on the quantitative and qualitative evidence across the range of studies.
	Meta-narrative mapping	Considers the commonalities and differences between various research contributions to topic(s). Identifies all the key dimensions of the problem(s) that have been researched; taking each in turn, gives a narrative account of the contribution (if any) made to it by each separate research tradition.
	Realist synthesis	Provides a review of an array of different programmes grouped and analysed according to their common underlying mechanisms. It draws lessons and analysis by establishing theories of how programmes succeed or fail, based on their consistencies and inconsistencies.
	Thematic synthesis	A method that is often used to analyse data in primary qualitative research. 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.
	Framework synthesis	A pragmatic method of qualitative evidence synthesis based on framework analysis. Offers a highly structured approach to organising and analysing data (e.g. indexing using numerical codes, rearranging data into charts, etc.).

The starting point for this method is most commonly the construction of a narrative summary of the findings of the various studies included in a review. That is, the studies have to be organised into logical categories using the data extracted about the intervention, study quality, contextual factors, and outcomes. The findings within each of these categories can then be analysed in terms of their relationships to each other and validated (Petticrew and Roberts, 2006; Popay et al., 2006).

However, the core challenge of narrative synthesis – as with many types of synthesis approaches – is that the varieties of interventions selected, as well as the categories used to unpack them, represent an excessive number of comparative options that can ultimately, jeopardise meaningful cross-comparison. The metaphor of comparing ‘apples with oranges’ or ‘chalk with cheese’ is often cited, quite rightly, to explain this concern, which is all the more prevalent in examining the type of integrated interventions that address child protection. The challenge of synthesising mixed-methods approaches effectively, and particularly qualitative-review results, is demonstrated by the extent of work being conducted to develop viable alternative forms of synthesis.

Realist synthesis, in contrast, seeks to understand why complicated programmes work or fail by overcoming the heterogeneity in intervention contexts, evaluative methods and outcomes. The method provides a way to assess an array of different programmes grouped and analysed according to their common underlying change mechanisms. It draws lessons and analysis by establishing theories of how programmes succeed or fail, based on a mapping of the causal chain from inputs to outcomes. This is a method routed in ‘theory-based evaluation’ promoted, for instance, by 3ie.

In addition to solving the ‘apples and oranges’ challenge, realist synthesis also encapsulates the principle of learning from both success and failure and reinterprets what it means to summarise or generalise from a range of evidence, based on the idea that cursory ‘case studies’ or ‘best buys’ do not offer enough depth to provide SR users or policy-makers with valuable learning (Pawson 2006; Boaz et al., 2002). Given the limited number and diversity of interventions addressed in the child protection ASR, however, even this approach failed to provide an adequate basis to suggest whether ‘programme x’ worked better than ‘programme y’.

Realist synthesis is, however, paired helpfully with narrative synthesis for two primary reasons. The first is that, contrary to most synthesis methods, realist syntheses cannot be either standardised or reproduced, as there can be no simple procedural formula to capture different decisions made by the implementing research team or their core ideologi-

cal or disciplinary biases. Second, best practice on this method remains under-developed. The area requires further experimentation and methodological solutions to become more useful to researchers. There are, at present, few experts on realist synthesis, and few studies that have applied this approach successfully with a transparent and systematic method. Narrative synthesis assists, therefore, in providing elements that realist synthesis lacks: reproducibility and, for the most part, a relatively established method (see Wong et al., 2013).

During the child protection ASR synthesis, analysis focused on the programme mechanisms and contextual factors of the included interventions. In order to get to the root of ‘what works, for whom, and in what circumstances’ the implementing team began by using the initial scoping paper commissioned for the project to identify the programme mechanisms that they expected to find in the included papers. As the search team uploaded the selected sources, descriptive coding information about the programme mechanisms involved in the intervention was applied. Once the search process was complete, and all papers had been fully coded, this information was extracted into a set of tables to facilitate further adjustments and analysis.

This method allowed the team to examine how programmes have been applied in a variety of contexts and expose some of the ‘how’ and contextual factors that affected the success (or more rarely, failure) of an intervention. However, the broad scope of this review of child protection and the limited team resources meant that, in most cases, the team was often reaching out to encompass the breadth of issues, rather than investing time in getting to the bottom of interventions across the selection of papers included in the review. In addition, the team’s overall ability to analyse contextual factors and the process of change in these interventions was constrained by the types of evaluation and assessment data that were available.

These challenges were, in turn, met with the types of integrated interventions common in child protection, which often combine a number of mechanisms to eventually improve child well-being. These may include, for example, combinations such as changes in teacher-training curricula, cash incentives for school attendance, universal cash transfers, positive-parenting programmes and changes in referral procedures. In these cases, the impacts of the individual elements of an intervention are rarely discussed, with little corresponding attention afforded to the mechanisms and target audiences that would be most effective in combating child-protection violations.

As a result, the implementing team was less able to answer the question on the impact of specific anti-

poverty interventions on child protection, although certain interventions, such as cash transfers, did demonstrate a strong contribution to positive child-protection outcomes. For the most part however, a detailed comparison of non-income versus income-based activities, such as the difference between parental education and adolescent skills training, was not an option, given the integrated nature of interventions and corresponding evaluative methods.

What could have been done differently, and why?

In retrospect, there are lessons to be learnt on the three challenges outlined in this paper: handling qualitative, quantitative, and mixed data; managing disciplinary diversity and evidence gaps; and making sense of synthesis. Each of these have implications for the different stakeholders involved in SRs – whether they be primary or secondary end-users, such as donors, NGOs, academic institutions or think tanks; intermediaries, such as knowledge-brokers and communication agents; or supply-side implementation teams from academia, research institutes or private-sector consultancies. Having said this, one overarching lesson is that a well-resourced and full capacity assessment of the situation and institutional and skills capacity (both internal and external) is critical in the initial developmental stages of any SR.

While the established and emerging evidence ‘out there’ will, in theory, always remain the same at a given point in time, the lenses through which knowledge is framed, re-framed, and synthesised will always depend on a range of factors. These include both macro- and micro-politics, disciplinary biases within the implementation team and commissioning agents, relative experience in the sector, the timeframes and, of course, the overall package of assigned resources (Jones et al., 2012).

The selection, for example, of a relatively large team of junior and mid-level researchers to search and code papers was both a reflection of the size of the child protection ASR and the resources available to address the wide number of papers gathered. As far as possible, steps were taken to ensure that the team exhibited relevant experience in child protection but, in some instances, the focus was shifted to hiring personnel with transferable skills – such as familiarity with broader social-development topics and processes. This was, ultimately an exercise of judgement, and to some extent, the chance availability of individual expertise. Of course, at a quality and technical management level, the requisition of experienced research staff is absolutely critical, with far less room for ‘generalists’ or those with transferable skills.

In addition to these considerations, there are two particular lessons related to the first challenge, of managing disciplinary diversity and evidence gaps.

- The first concerns the importance of the opening iterative stages that inform the grander ASR exercise itself. Working in a sector such as child protection presents particular challenges for the institutional remit of policy-makers and practitioners, as well as a historical tendency to focus on response mechanisms at the meso- and micro-scales. Getting an implementation team and expert management team to reach out beyond siloes and entrenched positions – as undertaken in the iterative stages of the child protection ASR – is a crucial exercise that sets the pace and frame of reference for the project as a whole. Researchers and SR commissioning agents must commit to these early phases and appreciate that scoping papers should be treated as evidence bases in themselves, and not a collection of hunches, expert reflections, and journal listings that constitute the ‘usual suspects’.
- The second lesson focuses on the need for a formal appreciation for expert-panel inputs as fully justifiable in certain types of SRs. As outlined above, drawing on methods espoused by the CHSRF can add vital perspectives on the context and timing of interventions – especially within SRs that aim to identify the important ‘how’ and ‘why’ factors of how an outcome occurred, rather than ‘if it occurred, ‘when’ or ‘where’. Expert-panel inputs, as well as crowd-sourced inputs such as e-surveys, do not address this issue on their own, but can fill gaps to some extent and direct attention to overlooked disciplines and findings. They can also go further by reinterpreting findings in the final synthesis stages of the SR. Of course, a series of unique skillsets and facilities are required to manage the quality and sustainability of this deliberative process. The possibility of ring-fencing expertise and funding for this exercise should be a standard topic of conversation at the outset of any SR process. If considered valuable, this exercise needs careful monitoring, given that expert inputs can easily reach stalemate or stagnate if deliberations are not well-facilitated to address issues related to entrenched values and beliefs.

The second challenge – handling methodological variety and stringent guidelines – reveals three overall lessons.

- The first concerns the realisation and acceptance that there is a wide variety of established and emerging methodological assessment tools that

relate to SRs. Many are housed in information clearing houses or knowledge-broker facilities, such as the Centre for Reviews and Dissemination (CRD), the Methods Network at the Cochrane Collaboration, and 3ie. Any SR project would be wise to familiarise itself with these assessment tools and reflect upon them critically before starting its work, and revisit these as the dominant evaluation methods on a regular basis.

- The second issue is that even if one or several evidence assessment tools are selected, it is very likely that alterations will be required given that this is an emerging area of specialist research, and that research questions vary enormously. The use of the MMAT tool in the child protection ASR for instance, while the best fit for the circumstance, lacked a comprehensive system to tackle the standards of research associated with non-RCT, mixed-methods and participatory processes. Implementation teams should simply be aware that hybrid assessment tools are inevitable, and should prepare for this eventuality.
- A third major lesson applies to the general research community in the medium and long-term. Acknowledging that only a minute fraction of intervention evaluations scored maximum points on the MMAT evidence assessment scale, with most scoring ‘one star’ or less, research practitioners are duty-bound to invest considerably more time in the critical reporting of methods used in their studies. This is particularly the case for qualitative, mixed and participatory-research outputs, which were thoroughly outclassed by quantitative studies in the child protection ASR. However, the MMAT should also focus on methodological reporting rigour without under-privileging papers that perform badly on methods discussions, but excellently on analytical strategies or overall presentation of results.

Finally, there are three lessons related to the third major challenge: undertaking evidence synthesis.

- It is well established in the emerging SR-based literature that the selection of a synthesis method is crucial, and must be closely associated with the contextual requirements of the study. However, as synthesis approaches are relatively experimental, alternative or adapted approaches are to be encouraged, as long as they are based on a solid reading of the SR literature and fully justifiable in the context of the study. The combination of narrative and realist synthesis used in the child protection ASR outlined above is one such example, but other authors have called for the adoption of innovative mechanisms to capture otherwise ‘external’ inputs,

such as project logframe documents and M&E data (Pawson, 2006). This experimentation might equally apply to the inclusion of expert-panel inputs, as outlined above, in realist synthesis. Again, undertaking realist synthesis to this extent can be a considerable drain on resources and management time, and need to be applied judiciously.

- Second, policy relevance should be a central concern in the synthesis stage. Many synthesis approaches rely too heavily on intricate detail and dedication to representation, rather than the compression and reconstitution of information into a format readily digestible to the commissioning agency, or indeed the policy and practice community as a whole. However, the issue goes beyond ‘communication’ toward ‘research-uptake’, which is a more purposive, multi-dimensional and mutually-constructive way to develop policy relevance for a SR. It may not be enough, for example, for policy-makers to know what intervention has an impact, in a certain context, with certain pre-conditions. The depth of knowledge required by policy-makers may stretch to a request for associated commercial costs that can, in turn, be weighed with benefits. SR implementation teams would do well, therefore, to clarify and ensure a shared understanding of ‘policy relevance’ in consultation with stakeholders throughout the SR project cycle to reduce the risk of confusing or misdirecting policy-makers and other end-users.
- Third, stakeholders in the SR process need to recognise the need for, but limits of, theory-led synthesis approaches. In the short-term, therefore, the responsibility lies with the SR implementation team and commissioning agents to recognise and adjust for sectoral idiosyncrasies, such as a prevalence of integrated programming in interventions. In the medium and long-term, it is the responsibility of the broader research community to include far more detail on the operating principles of interventions. This may involve recording the ‘theory of change’ in the evaluation documents where possible, but also programme planning, M&E data and other such material. It should also be noted that while this process may begin to cater for ‘complicated’ or ‘integrated’ interventions in SRs, it does not address the issue of ‘complexity’ – a concern that is often overlooked but that has relevance beyond social science SRs and requires distinct and separate approaches that cater for the risks associated with the unforeseen impacts of interventions.

Written by David Walker, ODI Research Officer (d.walker@odi.org.uk), Gina Bergh, ODI Research Officer (g.bergh@odi.org.uk), Ella Page, Consultant (ellagrainnepage@gmail.com) and Maren Duvendack, ODI Research Fellow (m.duvendack@odi.org.uk).

References and endnotes

References

- Boaz, A., Ashby, D., & Young, K. (2002) 'Systematic reviews: what have they got to offer evidence based policy and practice?' London: ESRC UK Centre for Evidence Based Policy and Practice.
- CRD (2009) Systematic Reviews: CRD's guidance for undertaking systematic reviews in healthcare, Centre for Reviews and Dissemination. York: University of York
- DFID (2011) 'Systematic Reviews in International Development: An Initiative to Strengthen Evidence-Informed Policy Making'. London: UK Department for International Development.
- DFID (2013) 'Assessing the Strength of Evidence: A 'How to' Note/ Practice Paper'. London: UK Department for International Development.
- Gasteen, M. (2010) DFID Research and Evidence Division - Systematic Review Program. London: UK Department for International Development.
- 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.
- Hagen-Zanker, J., Mallett, R., Duvendack, M, & Slater, R. (2012) 'Making Systematic Reviews work for international development research', SLRC-ODI. London: Overseas Development Institute
- Jones, H., Jones, N. A., Shaxson, L., & Walker, D. (2012) *Knowledge, policy and power in international development: A practical guide*. Bristol: The Policy Press.
- Long, A. F., & Godfrey, M. (2004) 'An evaluation tool to assess the quality of qualitative research studies'. *International Journal of Social Research Methodology*, 7, (2), 181-196.
- Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012) 'The benefits and challenges of using systematic reviews in international development research'. *Journal of development effectiveness*, 4, (3), 445-455.
- McGlynn, E.A, Kosecoff, J., Brook, R.H. (1990) 'Format and conduct of consensus development conferences: a multi-nation comparison'. *Improving consensus development for health technology assessment: an international perspective*, 32.
- Mikton, C., and Butchart A. (2009) 'Child maltreatment prevention: A systematic review of reviews.' *Bulletin of the World Health Organization*. 87, (5), 325–404.
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., Seller, R. (2012) Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49, (1), 47-53.
- Pawson, R. (2006) *Evidence-based policy: A realist perspective*. Sage Publications Limited.
- Petrosino, A., Boruch, R. F., Soydan, H., Duggan, L., & Sanchez-Meca, J. (2001) 'Meeting the challenges of evidence-based policy: The Campbell Collaboration'. *The ANNALS of the American academy of political and social science*, 578, (1), 14-34.
- Petticrew, M. and Roberts, H. (2006) *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*. Oxford: ESRC.
- Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Boardman, F., Gagnon, M.P., & Rousseau, M.C. (2011) Proposal: A mixed methods appraisal tool for systematic mixed studies review. Montreal: McGill University.
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Duffy, S. (2006) *Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme*. Version, 1. Oxford: ESRC.
- Popay, J., Rogers, A., & Williams, G. (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative health research*, 8, (3), 341-351.
- Ramalingam, B., & Jones, H. with Reba, T. and Young, J. (2008) Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. ODI Working Paper 285. London: Overseas Development Institute.
- Snilstveit, B. (2012). 'Systematic reviews: from 'bare bones' reviews to policy relevance'. *Journal of development effectiveness*, 4, (3), 388-408.
- Spencer, L., Ritchie, J., Lewis, J., Dillon, L. (2003) 'Quality in qualitative evaluation: A framework for assessing research evidence'. Government Chief Social Researcher's Office. London: National Centre for Social Research.
- van Eeten, M. (1999) 'Dialogues of the Deaf' on Science in Policy Controversies', *Science and Public Policy* 26, (3), 185–92.
- Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., Tugwell, P. (2012) 'How to do a good systematic review of effects in international development: a tool kit'. *Journal of development effectiveness*, 4, (3), 359-387.
- Walsh, D., and Downe, S. (2006) 'Appraising the quality of qualitative research', *Midwifery* 22, 108–119.
- White and Waddington (2012) 'Why do we care about evidence synthesis?' An introduction to the special issue on systematic reviews. *Journal of development effectiveness*, 4, (3), 1-12.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., Pawson, R. (2013) 'RAMESES publication standards: realist syntheses'. *BMC medicine*, 11, (1), 21.

Endnotes

1. Assembled by the CRD (2009).



Overseas Development Institute, 203 Blackfriars Road, London SE1 8NJ, Tel: +44 (0)20 7922 0300, Email: publications@odi.org.uk. This and other ODI Background Notes are available from www.odi.org.uk.

Readers are encouraged to reproduce material from ODI Background Notes for their own publications, as long as they are not being sold commercially. As copyright holder, ODI requests due acknowledgement and a copy of the publication. For online use, we ask readers to link to the original resource on the ODI website. The views presented in this paper are those of the authors and do not necessarily represent the views of ODI. © Overseas Development Institute 2013. ISSN 1756-7610.