

Evidence Gap Maps

A Tool for Promoting Evidence-Informed Policy and Prioritizing Future Research

Birte Snilstveit
Martina Vojtkova
Ami Bhavsar
Marie Gaarder

The World Bank
Independent Evaluation Group
Public Sector Evaluation Department
December 2013



Abstract

Evidence-gap maps present a new addition to the tools available to support evidence-informed policy making. Evidence-gap maps are thematic evidence collections covering a range of issues such as maternal health, HIV/AIDS, and agriculture. They present a visual overview of existing systematic reviews or impact evaluations in a sector or subsector, schematically representing the types of interventions evaluated and outcomes reported. Gap maps enable policy makers and practitioners to explore the findings and quality of the existing evidence and

facilitate informed judgment and evidence-based decision making in international development policy and practice. The gap map also identifies key “gaps” where little or no evidence from impact evaluations and systematic reviews is available and where future research should be focused. Thus, gap maps can be a useful tool for developing a strategic approach to building the evidence base in a particular sector. This paper provides an introduction to evidence-gap maps, outlines the gap-map methodology, and presents some examples.

This paper is a product of the Public Sector Evaluation Department, Independent Evaluation Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at mgaarder@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Evidence gap maps – A tool for promoting evidence-informed policy and prioritizing future research

Birte Snilstveit, Martina Vojtkova, Ami Bhavsar, Marie Gaarder¹

Keywords: evidence-informed policy, development effectiveness, evaluation, systematic reviews, gap maps, intervention effectiveness, knowledge translation, research prioritisation

JEL classification: I15; I38; D04; I28; Q10; Q50; Z18

Sector board: Social Protection

¹ Snilstveit, Vojtkova and Bhavsar: International Initiative for Impact Evaluation, 3ie, bsnilstveit@3ieimpact.org, mvojtкова@3ieimpact.org, abhavsar@3ieimpact.org; Gaarder: Independent Evaluation Group (IEG), World Bank, mgaarder@worldbank.org. Acknowledgements: The authors thank Annette Brown, Claire Allen, Dafni Skalidou, Hugh Waddington, Howard White, Phil Davies and Peter Tugwell for comments on previous drafts of this paper. The authors are also grateful for the inputs from participants at the Symposium on Systematic Reviews in International Development, Dhaka, December 2012. The work was supported by the International Initiative for Impact Evaluation (3ie). All errors are the responsibility of the authors.

1. Introduction

Over the last decade there has been an increased focus on evidence informed policy in the field of international development. This has manifested itself in an increase in both the demand and supply of rigorous evidence on “what works.” On the demand side, many funders, implementing agencies and governments in low and middle-income countries have increased their efforts to improve the use of evidence in their programming. The Department for International Development (DfID) in the UK, for instance, now requires all program proposals to use evidence in problem description and program design. Similarly, in Latin America, countries like Mexico, Colombia and Chile have institutionalized government evaluation through national evaluation bodies (Gaarder and Briceño, 2010 a, b).

On the supply side there has been an increase in efforts to produce evidence to inform policy (White and Waddington, 2012). A range of organizations and initiatives are engaged in the production of systematic reviews and high quality impact evaluations assessing the effectiveness of social and economic development interventions such as J-PAL, DIME, the World Bank Strategic Impact Evaluation Fund, 3ie, the Campbell Collaboration, the Cochrane Collaboration, the Alliance for Health Policy and Systems Research and the EPPI Centre. Thus, to date at least 850 impact evaluations and 300 systematic reviews focused on low and middle-income countries are either completed or ongoing (3ie, n.d)². The growth in the production of evidence, however, also presents challenges. How can decision makers get an overview of the existing evidence in a certain field when it is scattered around different databases, journals, websites and the grey literature? How can we make sure research is presented in a format which is useful and accessible for a non-technical audience? And how can we best ensure limited resources are spent efficiently and important evidence gaps are prioritized?

There is a need to ensure existing research is available and accessible in a format that is useful for decision making, and to develop tools to support an efficient and strategic approach to research commissioning informed by the existing evidence base. 3ie serves some of these functions by being a “clearing house” for impact evaluations and systematic reviews that focus on interventions to improve the lives of people in low and middle-income countries. Evidence gap maps are a recent innovation which aim to make existing research available to users and ensure new research is informed by the existing evidence.

² These figures exclude efficacy studies and systematic reviews of efficacy studies.

2. What is an evidence gap map?

Evidence gap maps are evidence collections that map out existing and ongoing systematic reviews or primary studies in a sector or subsector, such as maternal health, HIV/AIDS and agriculture. They present a visual overview of existing evidence using a framework of policy relevant interventions and outcomes, and provide access to user-friendly summaries of the included studies. Box 1 provides a graphic illustration of a gap map on HIV/AIDS. The rows of the framework represent the key interventions of a particular sector, while the columns cover the most relevant outcomes structured along the causal chain, from intermediate outcomes to final outcomes and cost effectiveness. Taken together, the framework should represent the universe of important interventions and outcomes in the sector or subsector covered by the gap map.

The first 3ie gap map was produced in 2010 (Gaarder, 2010) and focused on the health and nutrition impacts of agriculture interventions. Building on this we have developed the methodology as a means of improving the presentation and usefulness of our evidence databases and to identify gaps in the existing evidence base. 3ie evidence gap maps have two main objectives:

(1) To facilitate informed judgment and evidence-based decision making in international development policy and practice by providing user-friendly tools for accessing evidence and thereby enabling policy makers and practitioners to explore the findings and quality of the existing evidence on a topic quickly and efficiently.

(2) To facilitate strategic use of scarce research funding and enhance the potential for future evidence synthesis by identifying key “gaps” in the available evidence, thus indicating where future research should be focused.

A range of different approaches to evidence scoping, mapping and synthesis have been developed to support evidence informed policy making. These range from scoping studies (Anderson et al., 2008), evidence maps (Hetrick et al., 2010), the evidence-based policing matrix (Lum et al., 2012), systematic reviews (Cooper et al., 2009), overview of reviews (Becker and Oxman, 2011) and rapid reviews/ evidence assessments (Gannann et al., 2009; The Civil Service, 2008). Table 1 provides a comparison of evidence gap maps and other

scoping and mapping methodologies according to a number of key characteristics.³As can be seen from the table, their methodologies differ across a range of dimensions, from their aim, scope, type of evidence included, as well as the comprehensiveness of search, data extraction and analysis⁴.

For instance evidence maps, scoping studies and systematic maps tend to have a broad scope, and are primarily focused on identifying and describing the characteristics of the evidence base (Arksey and O'Malley, 2005; Bragge et al., 2011; Clapton and Coren, 2007; Gough and Thomas, 2012). While in some cases they may provide summaries of the included studies, the analysis is typically descriptive and limited in depth. On the other hand, rapid reviews/evidence assessments, systematic reviews and overviews of reviews have a more narrow scope, focusing on a particular intervention or outcome, but with greater depth of analysis and knowledge translation potential (Gannann et al., 2009; Becker and Oxman, 2011; Cooper et al., 2009).

³ In developing table 1 and the more detailed overview in Appendix 1 we attempted to locate key references for the various types of evidence mapping/ synthesis products, but in practice there are overlaps and the labels included in our typology are not consistently applied by all authors.

⁴ A more detailed overview of these methodologies can be found in Appendix 1.

Box 1: Graphic illustration of a gap map



Gap Map of Systematic Reviews: HIV/AIDS programmes in low and middle income countries
Updated: October 2011

Interventions / Outcomes	Awareness, knowledge	Attitudes and Beliefs	Risk behaviour / skills	HIV transmission	Take-up, retention/ adherence	Morbidity
Behaviour change interventions	Peer Education	Prevention for heterosexual men	Condom use for HIV positive women	Challenges in HIV prevention research	ICT for youth	
	Prevention for heterosexual men	Peer interventions	Couples-focused behavioural interventions	Behaviour interventions for prevention	Prevention in occupational settings	
	Peer interventions	ICT for youth	Behaviour change interventions for women			
	Prevention in Latin America		Peer Education			
	ICT for youth		Prevention in heterosexual men	Prevention in occupational settings		
			Behaviour interventions for prevention	School based interventions for youth		
			Influence of social agents			
			Peer interventions			
			Prevention in Latin America			
			ICT for youth			
		Prevention in occupational settings				
		School based interventions for youth				
Condom promotion / distribution			Prevention for youth in Africa			
Information, education and communication	Educating traditional healers	School-based sexual health in Africa	Educating traditional healers	Prevention for youth in Africa	Increasing youth's use of health services	Peer int
	School-based sexual health in Africa	Peer-led sexual health education for youth	Prevention for youth in Africa	Prevention in African youth	Peer interventions for HIV positive women	
	Peer-led adolescent sexual health education	Life skills education for youth	School-based sexual health interventions in Africa	Girls' education		
	Life skills education for youth	Parent-child communication in Africa	Peer-led sexual health education for youth	Peer interventions for HIV positive women		
			Life skills education for youth			
			Parent-child communication in Africa			
			Prevention in African youth			
		Girls' education				
		Peer interventions for HIV positive women				
Mass communication	Mass communication programmes				Increasing youth's use of health services	
	Mass media for young people					
Community mobilisation	Community interventions for youth					
	Counselling for testing of pregnant women		Family planning	Behaviour interventions for prevention	Home-based VCT	
		Routine vs. voluntary testing	VCT	Routine vs. voluntary testing	Integrating PMTCT with other health services	

Table 1: Comparison of Evidence Gap maps and other Scoping and Mapping methodologies

Methodology	Broad thematic/sector focus	Included evidence			Systematic Search	Critical appraisal	Analysis	Visualization using intervention-outcome framework	Access to user-friendly summaries
		Systematic Reviews	Impact Evaluations	Other					
3ie gapmap⁵	Yes	Yes	Can do	No	Yes*	Yes	No	Yes	Yes
Evidence map (GEM, Headspace)⁶	Yes	Yes	Yes	Yes	Yes	No	Yes (a)	No	Yes
Scoping study⁷	Yes	Can do	Can do	Can do	May be limited	No	Yes (a,b)	No	No
Systematic map⁸	Yes	Can do	Can do	Can do	Yes	Limited	Yes (a)	No	No
Rapid evidence assessment⁹	No	Can do	Yes	Can do	Yes (but may be limited)	Limited	Yes (a,c)	No	No
Systematic reviews¹⁰	No	No	Yes	Can do	Yes	Yes	Yes (a,d)	No	Can do
Overviews of systematic reviews¹¹	No	Yes	No	No	Yes	Yes	Yes (c)	No	No
Evidence-Based policing matrix¹²	Yes	No	Yes	No	Yes	Yes	No	No ¹³	Yes

* Systematic search and critical appraisal of SRs. The search for IEs may be more restricted. The IEs may not be critically appraised.

(a) summary of characteristics of the identified studies (reporting on categories such as study design, population, context, outcomes reported etc.)

(b) descriptive summary of findings

(c) analysis of results

(d) synthesis of findings

⁵ Gaarder (2010)

⁶ Parkhill et al. (2011); Bragge et al. (2011); <http://www.evidencemap.org>

⁷ Arksey & O'Malley (2005); Anderson et al. (2008); Davis et al. (2009); Grant et al. (2009); Levac et al. (2010)

⁸ Clapton and Coren (2007); Oakley et al., (2005); Bates, Peersman (1996)

⁹ Ganann et al. (2009); Khangura et al. (2012), Government Social Research REA Toolkit (2008)

¹⁰ Cooper et al (2009); Higgins and Greene (2011); Lavis (2009); Waddington et al (2012).

¹¹ Becker et al.(2011)

¹² Lum et al. (2011)

¹³ The evidence-based policing matrix includes a visualisation of the existing evidence in a matrix based on three key dimensions of crime prevention interventions. The number of studies, as well as direction of effects is also displayed in the matrix.

How do evidence gap maps compare with other evidence mapping/ synthesis products?

The 3ie evidence gap map is a recent addition to the range of approaches available to support evidence informed policy making. The approach draws on the principles and methodologies adopted in existing evidence mapping and synthesis products. For instance, by focusing on a sector or subsector they have a broad substantive scope similar to evidence maps, scoping studies and systematic maps. However, gap maps focus on studies assessing intervention effectiveness, as well as systematic reviews of such studies, while other mapping/ scoping approaches may include a broader range of evidence to address questions other than those of intervention effectiveness (Arksey and O'Malley, 2005; Clapton and Coren, 2007).

Moreover, most other mapping/ scoping approaches are generally not intended to provide inputs to policy and practice (Grant and Booth, 2009), but rather focus on mapping the existing evidence (Arksey and O'Malley, 2005; Bragge et al., 2011; Clapton and Coren, 2007), with the purpose of identifying research gaps and specific research questions.

Evidence gap maps are developed to both identify research gaps and to provide a resource for informing policy and practice.

The gap maps also share some features with synthesis products with a more narrow scope, such as rapid reviews/ evidence assessments, systematic reviews and overviews of reviews. In particular, the evidence gap maps draw on the strategies for systematic searching and critical appraisal developed for these synthesis approaches (Becker and Oxman, 2011; Khangura et al., 2012; The UK Civil Service, 2008; Waddington et al., 2012). But the broad scope of gap maps means they are more limited in terms of the depth of analysis they provide as compared to systematic reviews and in most cases, rapid evidence assessments and overviews of reviews.

The Evidence-Based Policing Matrix developed by Cynthia Lum and colleagues (2011) is probably the methodology with most similarities to evidence gap maps. It is based on a systematic search, includes high quality impact evaluations and critical appraisal of included studies. It is the only other evidence product we are aware of that combines a graphical visualization of the evidence according to a policy-relevant framework and provides users with direct links to user-friendly summaries of included studies. However, the matrix only includes primary studies, while gap maps also include systematic reviews. The Evidence-

Based Policing Matrix also has a more narrow focus on one outcome, rather than a set of outcomes.

Thus, there are overlaps between the methodologies of existing evidence products, and evidence gap maps. Evidence gap maps are unique in that they bring together the following key features: (1) visualization of the existing evidence using a framework of interventions and outcomes, using a theory-based approach with a focus on outcomes along the causal chain; (2) critical appraisal of systematic review evidence and a clear visualization of the quality of the evidence in the map; (3) links to a database with summaries of systematic reviews and impact evaluations; and (4) a broad scope covering a range of interventions and outcomes, focusing on a sector or subsector rather than a research question or field of enquiry.

What gap maps do not do

Users should consider gap maps as one of a multitude of tools and sources of information to aid decision making for policy and practice. Evidence gap maps do not answer a specific research question, but focus on providing a broad overview of the existing evidence.

Moreover, they are restricted to studies that assess the effectiveness of interventions and therefore do not include evidence on predictive factors, implementation, barriers and facilitators to effectiveness and other types of evidence.

Due to their broad scope the evidence gap maps also do not provide details about the contextual background of the included evidence, nor do they synthesize the findings of included systematic reviews and impact evaluations. Finally, evidence gap maps are not intended to provide recommendations or guidelines for policy and practice in and of themselves, but rather to be one of the sources that inform policy development and guidelines for practice. The next section outlines the key stages in developing an evidence gap map, as summarized in Box 2 below.

3. Gap map methods

Developing scope

The first step in producing a gap map is to set the scope by developing a framework which represents the universe of important interventions and outcomes in the sector or subsector covered by the gap map. The rows of the framework should cover all relevant interventions in a particular sector, while the columns should include all policy relevant outcomes structured along the causal chain, from intermediate outcomes to final outcomes and cost effectiveness.

Identifying the relevant interventions and outcomes for the framework involves identifying policy documents from major policy making bodies in the field of interest, as well as a review of existing frameworks and other literature, such as impact evaluations and systematic reviews. If there are existing frameworks that are broadly accepted and commonly used these can usefully be adopted. To ensure the relevance and acceptability of the framework this initial step should also involve consultation with stakeholders, including academic experts, policy makers, practitioners and funders. For instance, for 3ie's gap map on climate change, we developed an initial framework based on a review of policy and academic literature, as well as one-to-one consultations with a range of stakeholders. We then shared this with a group of funders and policy makers in a priority setting workshop. Participants were invited to provide feedback and present their perspectives on interventions and outcomes of high policy relevance, and the framework was adapted accordingly.

The framework is used to set the substantive inclusion criteria for the gap map. Depending on the focus of the gap map there might be additional substantive inclusion criteria of relevance that will help delimit the scope. In systematic reviews authors define the Population, Intervention, Comparison and Outcomes (PICO) when developing their inclusion criteria (O'Connor et al., 2011). Authors of gap maps could adopt a similar approach and also define the population of interest. All 3ie gap maps have been focused on low and middle income countries, but the scope of the population covered in the gap map depends on the interest of researchers and commissioners.

Setting study inclusion criteria

The next stage involves determining the types of evidence to be included in the gap map. In most cases (3ie) evidence gap maps will primarily be concerned with intervention

effectiveness. This means they map evidence designed to attribute effects of interventions on specific outcomes (i.e. impact evaluations using counterfactual analysis) or systematic reviews of such studies. Study inclusion criteria depend on the size and characteristics of the available literature, as well as the intended use of the gap map.

To address objective one outlined above, gap maps should rely on systematic reviews that critically appraise and synthesize all the available evidence on a particular intervention (Grimshaw et al., 2012).¹⁴ While single studies can be useful for informing policy in particular contexts, they should not serve as a basis for making generalized conclusions about effectiveness without being interpreted in the context of the available global evidence (Grimshaw et al. 2012). The dangers of relying on single studies have been well documented (Eg: Ioannidis, 2006) and systematic reviews attempt to deal with these issues (Waddington et al., 2012). If the gap map aims to address objective two outlined above, it should also include primary studies. Thus, there are two main options for inclusion restrictions depending on the main purpose of the gap map:

i) Include only systematic reviews - if the evidence base is relatively extensive, with a significant number of systematic reviews, and the main purpose of the evidence gap map is to provide a resource for policy makers.

We have produced two gap maps that include only systematic reviews, focused on HIV/Aids and maternal health respectively. Both these thematic areas had a large literature of systematic reviews. We had limited resources to produce the gap maps and our main aim was to produce a knowledge translation tool and draw out the main policy relevant findings for features written for International Aids Day and International Women's Day, respectively.¹⁵

ii) Include systematic reviews and primary studies: if the aim is to identify gaps in the existing evidence base to inform the commissioning of primary research and evidence synthesis, as well as to provide resources for policy makers.

¹⁴ Nevertheless, we recognize that systematic reviews, and by extension, systematic review gap maps may not always provide evidence on the circumstances under which an intervention may work. Therefore, combining evidence from a gap map with local evidence, including impact evaluations, may be useful (Lavis, 2009).

¹⁵ Available from: <http://www.3ieimpact.org/en/blog/2009/02/01/getting-zero-hiv-aids/>;
<http://www.3ieimpact.org/en/blog/2012/03/07/making-mothers-matter/>

For instance, the 3ie gap map on climate change mitigation and adaptation was used as an input to a priority setting workshop for a proposed climate change thematic window (funding program) and included both primary studies and systematic reviews (Snilstveit et al., forthcoming). We knew the evidence base was relatively limited and we had several objectives with the gap map. Our aim was that it would inform a strategic approach to funding through a climate change research program, by highlighting areas where primary studies were needed and any areas with a sufficient numbers of primary studies for systematic reviews. Additionally, we aimed to develop a comprehensive collection of evidence on the effectiveness and cost effectiveness of mitigation and adaptation activities, continuously updated as new evidence emerges.

Another example is a gap map on agriculture. It had been established that there was a relatively large number of both primary studies and systematic reviews in this field. The objective of this gap map was to inform the commissioning of future research in a planned 3ie thematic funding program on agricultural innovation, as well as providing a resource for policy making. Therefore we included both primary studies and systematic reviews in the gap map.

Searching for relevant studies and assessing inclusion

The next step is developing a strategy for populating the gap map framework with studies meeting the study inclusion criteria. The methods for doing so should draw on methods of systematic searching commonly used for systematic reviews and overviews of reviews, although the broad scope of the gap map necessitates some adaptations. The search effort depends on the time and resources available, as well as the intended use of the gap map. When the main purpose of the gap map is to inform commissioning of research, then the search can be tailored and less extensive. But if the main purpose of the gap map is to inform policy and practice, searching needs to be comprehensive and systematic.

If the gap map includes impact evaluations, the search for relevant impact studies will necessarily be more limited than in a systematic review, as gap maps have a broader scope and different purpose from systematic reviews. Therefore, while the search should be as comprehensive as possible, the search methods may need to be adapted for the research to be

manageable. Other approaches to evidence scoping, mapping and synthesis include a number of techniques to reduce the resource intensity of the search process while maximizing yields.

Search techniques adopted in rapid reviews/ evidence assessments might be particularly relevant. This includes for instance only searching English language literature, only covering a limited time period, applying specific search strings if used¹⁶, or applying a restriction on publication status (Khangura et al., 2012; The Civil Service, 2008). Moreover, focusing predominantly on key repositories of impact evaluations (such as 3ie's Database of Impact Evaluations¹⁷, J-PAL, the Cochrane Register of Trials), snowballing and citation tracking (Waddington et al., 2012), use of list serves, key databases and websites relevant to the thematic area under investigation are other techniques which can increase the efficiency of the search. The services of an experienced Information Specialist can be helpful in developing a comprehensive and efficient search strategy.

The search for systematic reviews should be comprehensive and systematic, but it will be simpler than the search for primary studies and searches typically carried out in systematic reviews (Becker and Oxman, 2011). It should focus primarily on databases of systematic reviews appropriate to the topic, including the Campbell Library, the Cochrane Library, the Environmental Evidence Library, the Database of Abstracts of Reviews of Effects (DARE),

¹⁶ 3ie gapmaps share similarities with the evidence maps produced by the Global Evidence Mapping Initiative (Bragge et al., 2011). The focus of these evidence maps is to describe the quality, design and characteristics of research in broad topic areas. In preparing the evidence maps the authors found that using highly sensitive searches with low precision proved unmanageable. Hence, information specialists developed search strings using specific search terms and terms frequently used as keywords in relevant papers and compared this with a highly sensitive search string typically used for systematic reviews (Parkhill et al., 2011). The latter returned significantly more references, but the basic search strategy did not miss any references of significance and the authors conclude 'results indicate that the EM search method is unlikely to miss the key articles necessary for decision making' (p. 159). The time and resource savings were significant for the more basic search strategy and the authors suggest the additional resources required for the highly sensitive search strategy not justified for the evidence map.

¹⁷ The 3ie database of impact evaluations is based on regular searches of a range of resources, including repositories of impact evaluations, grey literature and academic databases. Researchers use a carefully developed search strategy available in Mishra and Cameron (2013). This strategy can be combined with subject specific keywords and applied to a broader range of databases than those included in the regular searches, including subject specific databases of particular relevance to the thematic focus of the gap map.

R4D, Healthsystemevidence.org, HealthEvidence.org and the 3ie database of systematic reviews. Most of these databases are based on regular searches of academic databases and relevant websites, and provide a comprehensive coverage of reviews in their fields.¹⁸ Nevertheless, authors should supplement this with subject-specific searches in academic databases, relevant websites and the use of list serves. Other techniques such as snowballing¹⁹ and citation tracking can also be useful in ensuring the search captures all relevant systematic reviews.

Potentially relevant studies should be downloaded into reference management software and assessed against substantive and methodological inclusion criteria. In the gap maps we have produced so far we have not consistently double screened studies for inclusion, but we sought a second opinion if in doubt, and have recorded the reasons for exclusion. All systematic reviews deemed eligible for inclusion have been assessed a second time before being included in the gap map. In line with good synthesis practice we recommend future gap maps consider double screening, at least for a sample of studies to avoid introducing any researcher biases (Chandler et al., 2011). Gap maps should be updated regularly, at least every two years, to ensure the latest evidence is included.

Coding and critical appraisal

The next step in developing the gap map is to systematically code and extract data using a structured format, and to assess the quality of included systematic reviews or impact evaluations.²⁰ Studies are coded according to relevant intervention and outcome categories

¹⁸ The 3ie database of systematic reviews is based on regular searches of a range of resources, including libraries of systematic reviews, grey literature and academic databases. Researchers use a systematic review filter developed by an information specialist. This can be combined with subject specific keywords and applied to a broader range of databases than those included in the regular searches, including subject specific databases of particular relevance to the thematic focus of the gap map.

¹⁹ Snowballing refers to tracking back and pursuing references of relevant papers from both primary studies and reviews to increase the body of evidence - see Greenhalgh and Peacock (2005).

²⁰ As noted above we recommend that only evidence from systematic reviews is presented as evidence for informing policy and suggest the inclusion of impact evaluations should primarily be for informing commissioning of future primary studies and evidence syntheses. However, in some cases, for instance if the

included in the framework, as well as status of study (completed or on-going). Depending on the purpose of the gap map and the interest of the researchers/ commissioners, other categories may also be relevant, including geographical scope of the evidence, inclusion criteria (PICOs) of systematic reviews, information about intervention effectiveness and the design of primary studies.

We assess the quality of systematic reviews using a standardized critical appraisal tool, giving reviews an overall rating of high, medium or low in terms of the confidence with which their findings can be assured. The checklist is adapted from the checklist developed by the SURE collaboration (2011) and is included in Appendix 2. The critical appraisal is conducted by two people, with a final review by a third researcher with expertise in systematic reviews.

Producing user-friendly summaries, presentations, and analysis

A key feature of the gap map is that it provides direct access to user-friendly summaries. For 3ie evidence gap maps we use the same format as the summaries we produce for our evidence databases, including information about geographical location, background, methods and main findings. The summaries are peer reviewed, edited and uploaded to the 3ie evidence database. Once studies have been uploaded on the database the cells in the gap map are populated with links to the summaries of studies covering corresponding interventions and outcomes.

Studies may be repeated in several places in the gap map if they cover several interventions or outcomes. We use a traffic light color coding to indicate the quality of the systematic review evidence in the gap map so that users can visually assess the state of the evidence in the field. Lum et al. (2011) use symbols and color coding to indicate the statistical significance and direction of effects of studies in their matrix. A similar approach can also be adopted in evidence gap maps, allowing users to easily identify which interventions work or not.

evidence base is nascent and there is a lack of systematic reviews, but there is a strong demand for evidence to inform decision making researchers may decide to also highlight lessons learnt from existing impact evaluations. In such cases researchers should assess the quality of relevant impact evaluations using a recognised tool for critical appraisal.

If time and resources permit, a summary report with implications for policy and future research can add significant value to the evidence gap maps. The report should describe the size and characteristics of the available evidence, and summarize the findings from systematic reviews considered of sufficient quality for informing policy (Becker and Oxman, 2011). Researchers may also consider developing summary of findings tables for all included systematic reviews (Schünemann et al., 2011) if these are not already available, although this will add to the time and resources required to complete the project. At a minimum, gap maps should be accompanied by a note reporting details about the sources, methods and time period adopted for the search so that users can assess the comprehensiveness and reliability of the gap map.



4. Example: HIV/ AIDS gap map

The systematic review gap map on HIV/AIDS²¹ was one of the first evidence gap maps produced by 3ie for the occasion of the 2011 World AIDS day. It was produced with the aim of providing a resource to support strategic decision making on maximizing the effectiveness and cost-effectiveness of the global AIDS response. The framework for the gap map covers key policy-relevant interventions to prevent and address HIV/AIDS in low- and middle-income countries and HIV-related outcome categories along the causal chain. The framework was drawn up based on a review of the systematic review literature, causal chain analysis and key policy documents such as the UNAIDS World AIDS reports (UNAIDS 2010, UNAIDS 2011) and the new Investment Framework for HIV/AIDS (Schwartländer et al. 2011).

Systematic reviews included in the gap map were identified through an extensive search of relevant databases, including PubMed, the Cochrane Library, the 3ie Database of Systematic Reviews, Health Systems Evidence, International Child health Review Collaboration and the International Child health Review Collaboration. The search covered the period from 1993 up to November 2011. We identified 43 systematic reviews and 11 protocols which met our inclusion criteria across 16 broad intervention categories and 13 broad outcome categories.

We produced summaries of each included review and protocol, and assessed the quality of completed reviews using the 3ie checklist adapted from the Supporting the Use of Research Evidence checklist (2011). Evidence on different outcomes (in columns) was then mapped onto different categories of interventions (in rows). Each cell represents one systematic review. The color of each cell represents the confidence rating about the validity of the systematic review findings. Each cell in the online version of the gap map is linked to a user-friendly summary of the systematic review on the 3ie Systematic review database to allow the user to explore the main findings and quality appraisal information in more detail. Below is a brief overview of findings from a subsequent analysis of this subset of reviews.

²¹ http://www.3ieimpact.org/media/filer/2013/12/02/hiv_gap_map.pdf

Behavior change programs

One systematic review synthesized evidence on behavior change programs. The authors found suggestive evidence that behavior interventions for women living with HIV may increase the frequency of condom use, although the review found no effect when measuring consistent condom use (Carvalho and colleagues 2011).

Information, education, and communication programs

Only one systematic review specifically focused on information, education and communication programs. The review finds limited evidence that educating traditional healers may increase knowledge about HIV/AIDS, although it is not clear whether this led to changes in patient management or reductions in HIV incidence (Sorsdahl et al. 2009).

Implementation of biomedical interventions

Three reviews reviewed the effectiveness of biomedical interventions when implemented in real world settings. Bateganya and colleagues (2007) find limited evidence that offering voluntary counselling and treatment at an optional location leads to greater take up than at health facilities. Tudor Car and colleagues (2011) find limited evidence that integrating rapid HIV testing and structured assessment of antiretroviral treatment adherence in labor ward clinics improves mother-infant treatment coverage. Ng and colleagues conclude that community- or population-based interventions to control sexually transmitted infections do not appear to be an effective HIV prevention strategy in most settings (Ng et al. 2011)

Treatment and care programs

Two systematic reviews set out to synthesize evidence on the effectiveness of treatment and care programs for people living with HIV/AIDS such as family support and psychosocial interventions for children affected by HIV/AIDS. However, as no studies met their inclusion criteria, the evidence on these interventions remains inconclusive (Mohanam & Kamath 2009; King et al. 2009).

The systematic review gap map on HIV/AIDS aims to enable policy makers and practitioners to explore the findings and quality of existing systematic review evidence and facilitate informed judgment and evidence-based decision making on HIV/AIDS policies and programming. It also identifies key “gaps” where little or no evidence from systematic

reviews is available, such as a lack of systematic review evidence on the effectiveness of structural interventions and voluntary male circumcision interventions in real-world settings. Of the included reviews only eight were rated at high confidence in the validity of findings, highlighting a lack of high quality systematic review evidence on the effectiveness of mass-media campaigns, community mobilization, interventions targeting key populations, economic interventions and interventions to reduce HIV/AIDS stigma.

5. How can evidence gap maps be used to improve evidence informed policy?

Informing policy

Gap maps provide a visual representation of the available evidence and access to user-friendly summaries of the included impact evaluations and systematic reviews. The links to summaries allow users to explore the findings of different studies, and to make an informed judgment about the extent to which they can rely on the systematic review evidence for decision making. The visual representation of the evidence allows users to explore the key gaps in the existing evidence base.

Evidence gap maps produced with the primary purpose of informing policy should focus on systematic reviews, be as comprehensive as possible and include a critical appraisal of the included systematic reviews. For instance, for the evidence gap map on HIV/ AIDS, the primary purpose was to provide a tool to be used to inform policy. As such it only includes systematic reviews and enables policy makers and practitioners to explore the findings and quality of existing systematic review evidence and facilitate evidence-informed decision making on HIV/AIDS policies and programming.

Gap maps may be particularly useful for informing policy if the evidence base is extensive. It gives users an overview of the evidence on the effectiveness of different interventions in a particular sector or thematic area. By bringing together the evidence from systematic reviews of different interventions they may allow users to assess the comparative effectiveness of interventions quickly and efficiently (cf: Lum et al., 2011). If little or no high quality evidence is available, this can also be helpful knowledge for decision makers.

Promoting strategic research agendas

Despite an increase in funding for research on intervention effectiveness in recent years, resources fall short of what is required to close important evidence gaps. Moreover, studies conducted in isolation or without sufficient attention to existing research agendas can lead to a fragmented evidence base lacking in relevance (Ioannidis, 2006), presenting a barrier to our ability to reach generalizable conclusions beyond individual study contexts.

Gap maps can help address these concerns. They can be a useful tool for developing a strategic approach to building the evidence base in a particular sector and ensure available resources are spent efficiently. They help identify areas of high policy relevance where evidence is lacking and can feed into setting research priorities. Furthermore, by providing easy access to and overview of existing studies, evidence gap maps can be used to identify promising interventions that warrant replication in different contexts, thereby enhancing the reliability of conclusions about intervention effectiveness (Valentine et al., 2011), highlighting sources of contextual heterogeneity and addressing concerns about external validity.

Overlaying a map of impact evaluations with a map of systematic reviews provides a visual representation of different evidence gaps. Firstly, it highlights ‘absolute gaps’ where there are few or no studies and primary studies should be targeted. Identifying these gaps might be of particular relevance for funders of impact evaluations who want to target their funding towards important evidence gaps. It can also be useful for researchers who want to identify areas where their work can add particular value.

Secondly, it identifies areas with substantial un-synthesized primary evidence. Many systematic reviews fail to provide policy relevant findings due to a lack of primary evidence. While highlighting the lack of evidence is an important finding in itself, it can be discouraging if substantial time and resources have been invested in a review. Thus, evidence gap maps can help mitigate this issue by identifying areas with sufficient primary evidence for systematic reviews to be worthwhile, and where primary studies should be prioritized over evidence synthesis. This use of gap maps might be particularly relevant for commissioners and producers of systematic reviews, such as the Campbell and Cochrane Collaborations.

Thirdly, mapping the availability and quality of existing evidence syntheses highlights areas with systematic reviews of either low quality or with sufficient new evidence to warrant an

update. Inspecting the findings of systematic reviews in the gap map will also reveal where reviews have been unable to reach overall conclusions about intervention effectiveness due to insufficient high quality evidence, indicating a need for additional primary studies.

Finally, an additional way to make use of the gap map methodology which holds much promise for research priority setting in institutions involved in project financing, such as international development organizations, is to overlay a map of the existing evidence on a map of ongoing and planned projects in the sector or subsector in question. An inventory of such projects will help determine which priority research questions can actually be investigated in the short to medium term (Tanner et al., 2013).

6. Conclusions

Evidence gap maps can be used as a tool for informing policy by making evidence more readily available to users in ready packaged collections. They can also inform future research, and allow for a systematic and strategic approach to closing the evidence gap. They enable users to explore and compare the quality, characteristics, distribution and findings of available evidence on the effectiveness of interventions in a particular sector.

Moreover, they allow bigger picture comparisons of the evidence on different interventions within a particular field, exploration of intervention effectiveness along the causal chain, and provide access to user-friendly summaries of systematic reviews and impact evaluations on the 3ie database. While gap maps do not provide context specific evidence and do not draw conclusions or recommendations for policy and practice, their interactive nature enables users to explore the evidence in more detail and to derive their own conclusions and implications for policy making and research agendas.

The methodology used for evidence gap maps can be easily adapted according to the purpose of the gap map, size of the evidence base and available resources. It is work in progress and 3ie is exploring options for improving their presentation, including ways of making it interactive, adding summary of findings tables and addressing cost-effectiveness. A series of gap maps is envisaged over time, integrating the evidence from 3ie's impact evaluation and systematic reviews databases covering all major sectors. This will allow policy makers, practitioners and researchers easy access to existing research on intervention effectiveness.

References:

- Anderson, S., Allen, P., Peckham, S. and Goodwin, N. (2008) Asking the right questions: Scoping studies in the commissioning of research on the organisation and delivery of health services. *Health Research Policy and Systems*, (6) 7: 1-12.
- Arksey, H., O'Malley, L., Baldwin, S. and Harris, J. (2002) Services to Support Carers of People with Mental Health Problems: Literature Review Report, Social Policy Research Unit, University of York, York.
- Arksey, H. and O'Malley, L. (2005) Scoping Studies: Towards a Methodological Framework. , *International Journal of Social Research Methodology*, 8, 1, 19-32.
- Bates, S. and Coren, E. (2006) Mapping the literature on the extent and impact of parental mental health problems on the family, the interventions available and the acceptability; accessibility and effectiveness of interventions, Systematic Map 1: Summary report, London: Social Care Institute for Excellence, www.scie.org.uk/publications/index.asp
- Bates, S., Clapton, J. and Coren, E. (2007) Systematic maps to support evidence base in social care, *Evidence & Policy* , Vol 3, 4, 539-51.
- Bates, S. And Rutter, D. (2008) Developing the process of systematic mapping-the social welfare literature, SCIE available from http://www.campbellcollaboration.org/artman2/uploads/1/SocialWelfare_Rutter.pdf
- Becker, L.A. and Oxman, A.D. (2011), 'Overviews of reviews', Chapter 22 in Higgins, J.P.T. and Green, S. (Eds), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
- Bhavsar, A., Hombrados, J.G., Snilstveit, B., Vojtkova, M., and Waddington, H. (2012) '3ie Maternal Health Gap Map', The International Initiative for Impact Evaluation. Available from <http://www.3ieimpact.org/en/about/what-3ie-does/systematic-reviews-programme/3ies-house-systematic-review-research/>
- Bragge, P., Clavisi, O., Turner, T., Tavender, E., Collie, A. and Gruen, R. (2011) The Global Evidence Mapping Initiative: Scoping research in broad topic areas. *BMC Medical Research Methodology* 2011, (11) 92: 1-12.
- Chandler, J., Churchill, R., Higgins, J., Lasserson, T. and Tovey, D. (2011), Methodological Expectations of Cochrane Intervention Reviews (MECIR): Methodological standards for the conduct of new Cochrane Intervention Reviews, Version 2.2, available from: http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.2%2017122012_0.pdf – accessed 17/7-2013
- Clapton, J., Rutter, D. And Sharif, N. (2009) SCIE Systematic mapping guidance, April 2009 [draft]. Using knowledge in social care, research resource 03. Social Care Institute for Excellence.
- Cooper, H., Hedges, L. V. and Valentine, J. C. (2009), *The Handbook of Research Synthesis and Meta-Analysis*, New York: Russell Sage Foundation.

- Davis, K., Drey, N. and Gould, D. (2009) What are scoping studies? A review of the nursing literature. *International Journal of Nursing Studies*, 46:1386-1400.
- Ganann, R., Ciliska, D. and Thomas, H. (2010) Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*, 5 (56): 1-10.
- Gaarder, M. (2010), 'Evidence on links in the causal chain: Agriculture, nutrition and health', Delhi: International Initiative for Impact Evaluation.
- Gaarder, M. and Briceño, B (2010a), Institutionalisation of Government Evaluation: Balancing Trade-Offs, *Journal of Development Effectiveness*, Vol. 2, Issue 3.
- Gaarder, M. and Briceño, B. (2010b), Institutionalisation of Government Evaluation: Balancing Trade-Offs, International Initiative for Impact Evaluation (3ie) Working Paper 8, New Delhi: 3ie. Available from http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_8.pdf.
- GEM (n.d), About evidence mapping, available from: <http://www.evidencemap.org/about> - accessed 9/12-2012.
- Gough, D. and Thomas, J. (2012), 'Commonality and diversity in reviews', Chapter 3 in: Gough, D., Oliver, S. and Thomas, J. (Editors), 'An introduction to Systematic Reviews', London: Sage.
- Greenhalgh, T. and Peacock, R. (2005), Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources, *BMJ*, 331 (7524), 1064-5.
- Grant, M.J. and Booth, A.. (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26: 91-108.
- Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J. and Squires, J. E. (2011), Knowledge translation of research findings, *Implementation Science*, 7:50.
- Higgins, J.P.T, and Green, S. (editors) (2011), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- Ioannidis, J. P. A. (2006), Evolution and Translation of Research Findings: From Bench to Where?, *PLOS Clinical Trials*, 1(7): e36. doi:10.1371/journal.pctr.0010036
- Jepson, R. G., Harris, F.M., Platt, S., Tannahill, C. (2010), 'The effectiveness of interventions to change six health behaviours: a review of reviews', *BMC Public Health*, Vol. 10, Issue. 538.
- Jolliffe, D. & Farrington, D.P. (2011) A rapid evidence assessment of the impact of mentoring on re-offending. Home Office Online Report 11//07, United Kingdom. Available at: http://resources.civilservice.gov.uk/wp-content/uploads/2011/09/rdsolr1107_tcm6-7376.pdf
- Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J. and Moher, D. (2012), Evidence summaries: the evolution of a rapid review approach. *Systematic reviews*, 1 (10): 1-9.
- Lavis, J. (2009), How Can We Support the Use of Systematic Reviews in Policymaking?, *PLoS Med* 6(11): e1000141. doi:10.1371/journal.pmed.1000141
- Levac, D., Colquhoun, H. and O'Brien, K.K. (2010) Scoping studies: advancing the methodology. *Implementation Science*, 5 (69): 1-9.

Lum, C., Koper, C.S. And Telep, C.W. (2011), 'The Evidence-Based Policing Matrix. *Journal of Experimental Criminology*', 7, pp. 3-26.

Oakley, A., Gough, D., Oliver, S. and James, T. (2005), 'The politics of evidence and methodology: lessons from the EPPI-Centre', *Evidence & Policy*, vol 1, no 1, pp 5-31.

O'Connor, D., Green, S. and Higgins, J.P.T. (2011), 'Defining the review question and developing criteria for including studies', Chapter 5 in: Higgins, J.P.T. and Green, S. (editors), *Cochrane Handbook of Systematic Reviews of Intervention*, Version 5.1.0 (updated March 2011), The Cochrane Collaboration, available from www.cochrane-handbook.org.

Oxman, A. D. and Fretheim, A. (2008), 'An overview of research on the effects of results-based financing'

Peersman, G. (1996), *A Descriptive Mapping of Health Promotion in Young People*, EPPICentre, Social Science Research Unit, Institute of Education, University of London, London.

Parkhill, A.F., Clavisi, O., Pattuwage, L., Chau, M., Turner, T., Bragge, P. and Gruen, R. (2011) Searches for evidence mapping: effective, shorter, cheaper. *Journal of the Medical Library Association*, 99 (2): 157-160.

Mishra, A. and Cameron, D. (2013), *Registry of Impact Evaluation Published Studies Protocol*, Washington D. C.: International Initiative for Impact Evaluation.

Schünemann, H. J., Oxman, A.D., Higgins, J.P.T., Vist, G.E., Glasziou, P. and Guyatt, G. H. (2011), Chapter 11 in Higgins, J.P.T. and Green, S. (Eds), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Schwartländer, B., Stover, J., Hallet, T., Atun, R., Avila, C., Gouws, E., Bartos, M., Ghys, P.D., Opuni, M., Barr, D., Alsallaq, R., Bollinger, L., de Freitas, M., Garnett, G., Holmes, C., Legins, K., Pillay, Y., Stanciole, A.E., McClure, C., Hirschall, G., Laga, M., Padian, N. and the Investment Framework Study Group (2011) Towards an improved investment approach for an effective response to HIV/AIDS. *Lancet*, 337: 2031-2041.

Tanner, J. et al., (2013), 'Delivering the Millennium Development Goals to Reduce Maternal and Child Mortality: A Systematic Review of Impact Evaluation Evidence', Washington D.C.: Independent Evaluation Group (IEG), World Bank.

The Global Evidence Mapping Initiative (n.d.): Putting Research into context [online]. Available at: <http://www.evidencemap.org/> [Accessed 17 July].

The UK Civil Service (2008) *Government Social Research REA Toolkit 2008* [online]. Available at: <http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is> [Accessed on 17 July 2013].

The SURE Collaboration (2011), 'SURE checklist for making judgements about how much confidence to place in a systematic review'. In: *SURE guides for preparing and using policy briefs*. Version 2.1 [updated November 2011]. The SURE Collaboration, 2011. Available from www.evipnet.org/sure [Accessed on 17 July 2013].

UNAIDS (2010) *Getting to Zero: 2011-2015 Strategy*. Joint United Nations Programme on HIV/AIDS (UNAIDS). Available at:

http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2010/JC2034_UNAIDS_Strategy_en.pdf

UNAIDS (2011) World AIDS Day report 2011: How to get to zero: Faster, Smarter, Better. Available on: http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/JC2216_WorldAIDSday_report_2011_en.pdf

UNAIDS (2012) World AIDS Day report 2012: Results 2012. Available at: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/JC2434_WorldAIDSday_results_en.pdf

Underwood, L., Thomas, J., Williams, T. and Thieba, A. (2007) The effectiveness of interventions for people with common mental health problems on employment outcomes: a systematic rapid evidence assessment. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Available at: <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2315>

Valentine, J. C, Biglan, A., Boruch, R. F Castro, F. G., Collins, L. M., Flay, B. R., Kellam, S., Mościcki, E. K., Schinke, S. P. (2011), Replication in Prevention Science, *Prevention Science*, 12:103–117, DOI 10.1007/s11121-011-0217-6

Vojtkova, M., Snilstveit, B., Bhavsar, A., Hobarados, J.G., Thomae, M., Thomas, S. and Waddington, H.W., (2011) Gap Map of Systematic Reviews: HIV/AIDS programmes in low and middle income countries. International Initiative for Impact Evaluation. Available at: http://www.3ieimpact.org/media/filer/2012/06/12/hiv_gap_map.pdf

Waddington, H., White, H., Snilstveit, B., Hombrados, J.G., Vojtkova, M., Davies, P., Bhavsar, A., Evers, J., Koehlmoos, T.P., Petticrew, M., Valentine, J.C. and 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.

White, H. and Waddington, H., (2012), Why do we care about evidence synthesis? An introduction to the special issue, *Journal of Development Effectiveness*, 4 (3): 359-387.

Appendix 1: Overview of different types of mapping, scoping and synthesis

Methodology	Aim/ purpose	Inclusion	Search	Critical appraisal	Data extraction	Analysis	Presentation	Timeline	References
3ie gapmap	To map and visualize the existing an ongoing SR (and IE) evidence in a particular topic area. To inform policy and practice; to inform research planning.	SRs and IEs, some gap maps may only include systematic reviews	Comprehensive and systematic for SRs. More purposive for IEs.	Rigorous Critical appraisal (SRs, may also include critical appraisal of IEs)	Intervention, outcome (outcome measure), summary of findings, critical appraisal categories. May be limited to a single person screening and extracting data.	Summary of the quality and quantity of available evidence. Descriptive summary of key findings of systematic reviews.	Visual two-dimensional framework. Links to a database with summaries of the included studies.	2-3 months	Gaarder (2010)
Evidence map (GEM, Headspace)	To describe the nature, characteristics and volume of research in a particular area; and to identify evidence gaps by comparing the key research questions identified by stakeholders with the available literature.	SRs and primary studies	Comprehensive and systematic. More specific and less sensitive than SR searches.	None	Intervention, study design and detailed study characteristics.*	Summary of identified studies by study design, context, population, condition, and outcomes for each intervention. Often includes a commentary on the evidence base.	GEM: Report summarizing yield (no of studies), with references to relevant studies. Headspace: Database searchable by condition and outcomes.	Not clear (up to 2 years)	Parkhill et al. (2011); Bragge et al. (2011); http://www.evidencemap.org
Scoping study	To examine the extent, range and nature of research activity, identify gaps in the evidence base, determine the value of undertaking a full SR or summarize and disseminate research findings in a particular field of enquiry. To inform policy and practice; to inform research planning.	Determined by research question. All relevant literature includable.	Determined by time and resources available. May be more limited than a SR searches.	None	Determined by research question. Typically study population, intervention, outcome measures, study design; may include findings relating to effectiveness and gaps in research.	An analytic framework, or thematic construction to present a narrative account of existing literature, using descriptive thematic narratives or descriptive summaries of statistical data.	Report. May include tables.	6 months to 1 year	Arksey & O'Malley (2005); Anderson et al. (2008); Davis et al. (2009); Grant et al. (2009); Levac et al. (2010)
Systematic map	To provide a methodical overview of the quantity and quality of evidence in relation to a broad (open) question of policy or management relevance and provide a searchable database of references to be used by various stakeholders Including policy makers, service users, practitioners and researchers.	Determined by research question. All relevant literature includable.	Comprehensive search similar to that conducted for SRs. The choice of sources depends on the topic.	Limited descriptive appraisal. May be limited to a subset or sample of papers when evidence base large.	Determined by research question. For intervention related questions, inclusion criteria focus on PICO's.*	Maps out literature available in various databases. Analysis is limited to showing what literature is available and what the key gaps are. Can be translated into reports, summaries or a database interface.	Report. Can be made available as an online database. Brief summary reports can be produced.	6 months to 1 year	Clapton and Coren (2007); Oakley et al., (2005); Bates et al (2006; 2008); Peersman (1996)

Methodology	Aim/ purpose	Inclusion	Search	Critical appraisal	Data extraction	Analysis	Presentation	Timeline	References
Rapid evidence assessment	To provide a quick review and synthesis of the available evidence to facilitate informed decision-making about the effectiveness of an intervention or policy under time and/or resource constraints; provide a map of evidence in a topic area to identify where there is existing evidence and direct future research; or serve as interim evidence assessment until a more systematic review can be conducted.	IEs and/or SRs (may include other primary studies).	May be more limited than a full systematic search with limitations adopted for years, languages, publication status, search strings, and sources searched.	Limited quality appraisal. The rigor and detail of the quality appraisal may vary.	Might use PICOs. May be limited to a single person screening and extracting data. Data collection may be limited to key results and key data for simple quality assessment.	Simple narrative, descriptive or tabular analysis reporting quantities of literature and overall quality/direction of effect reported in the literature with limited interpretation of the findings	Report: may contain tables.	3 weeks to 6 months	Ganann et al. (2009); Khangura et al. (2012), Government Social Research REA Toolkit (2008)
Systematic reviews	To provide a comprehensive, unbiased assessment and synthesis of the available evidence to answer a specific research question.	IEs. For non-effectiveness questions may include other types of evidence.	Comprehensive and systematic search	Rigorous critical appraisal with a comprehensive risk of bias assessment of effectiveness studies.	PICOs, study characteristics, risk of bias/critical appraisal categories and study findings / information necessary to calculate effect sizes. Two people independently screen studies and extract data.	Meta-analysis or narrative / thematic synthesis of findings from all included studies. Additional analysis such as moderator analysis may be conducted.	Report outlining the methodology and findings, using forest plots if possible.	1 to 2 years	Cooper et al (2009); Higgins and Greene (2011); Lavis (2009); Waddington et al (2012).
Overviews of systematic reviews	To provide users with an accessible overview of SRs available in a particular area summarizing systematic review findings of effects of two or more interventions or systematic review findings addressing the effectiveness of the same intervention on different outcomes	SRs (of intervention effectiveness)	Comprehensive search for systematic reviews, focusing on databases of systematic reviews	Critical appraisal of SRs.	Data from included systematic reviews, in some cases additional data from included studies.*	Summarize results from all included reviews. Additional analysis may be conducted.	Report outlining the methodology and findings.	Not clear	Becker et al.(2011)
Evidence policing matrix	To guide the formulation and selection of strategies in policing, develop an agenda for future policing research and serve as a practice-oriented research translation tool that may better facilitate the adoption of evidence-based policing and evidence-based funding.	IEs	Search of key reports, SRs, library databases and websites of professional and government organizations.	Quality appraisal using the Maryland Scientific Methods Scale (SMS).	Key intervention dimensions (specificity of strategy, level of pro-activity of intervention, nature of target), evaluation design and outcome (direction and statistical significance of effect)*	Studies presented visually in a three dimensional matrix based on intervention characteristics. Each study is represented by a dot or triangle, with the shape and color indicating direction of effect and statistical significance.	Visual: a three-dimensional matrix Links to a database with summaries of the included studies.	Not clear	Lum et al. (2011)

This table captures typical features of the included methodologies. Considerable variation and overlap may exist within and between these methodologies. Timeline information is approximate. SR stands for systematic review. IE stands for impact evaluation. * indicated when it is not clear whether two people independently screen studies for inclusion and extract data.

Appendix 2: Checklist for making judgments about how much confidence to place in a systematic review of effects (adapted version of SURE checklist)ⁱ

Assessed by:
Date:

Section A: Methods used to identify, include and critically appraise studies

<p>A.1 Were the criteria used for deciding which studies to include in the review reported?</p> <p>Did the authors specify:</p> <p><input type="checkbox"/> Types of studies</p> <p><input type="checkbox"/> Participants/ settings/ population</p> <p><input type="checkbox"/> Intervention(s)</p> <p><input type="checkbox"/> Outcome(s)</p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Partially</p> <p><input type="checkbox"/> No</p> <p><i>Coding guide - check the answers above</i></p> <p><i>YES: All four should be yes</i></p> <p><i>NO: All four should be no</i></p> <p><i>PARTIALLY: Any other</i></p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>A.2 Was the search for evidence reasonably comprehensive?</p> <p>Were the following done:</p> <p><input type="checkbox"/> Language bias avoided (no restriction of inclusion based on language)</p> <p><input type="checkbox"/> No restriction of inclusion based on publication status</p> <p><input type="checkbox"/> Relevant databases searched (<u>Minimum criteria</u>: All reviews should search at least one source of grey literature such as Google; for health: Medline/ Pubmed + Cochrane Library; for social sciences IDEAS + at least one database of general social science literature and one subject specific database)</p> <p><input type="checkbox"/> Reference lists in included articles checked</p> <p><input type="checkbox"/> Authors/experts contacted</p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Partially</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Can't tell</p> <p><i>Coding guide - check the answers above:</i></p> <p><i>YES: All five should be yes</i></p> <p><i>PARTIALLY: Relevant databases and reference lists are both reported</i></p> <p><i>NO: Any other</i></p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>A.3 Does the review cover an appropriate time period?</p> <p><i>Is the search period comprehensive enough that relevant literature is unlikely to be omitted?</i></p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Can't tell (only use if no information about time period for search)</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Unsure</p> <p><i>Coding guide:</i></p> <p><i>YES: Generally this means searching the literature at least back to 1990</i></p> <p><i>NO: Generally if the search does not go back to 1990</i></p> <p><i>CAN'T TELL: No information about time period for search</i></p> <p><i>Note: With reference to the above – there may be important reasons for adopting different dates for the search, e.g. depending on the intervention. If you think there are limitations with the timeframe adopted for the search which have not been noted and justified by the authors, you should code this item as a NO and specify your reason for doing so in the comment box below. Older reviews should not be downgraded, but the fact that the search was conducted some time ago should be noted in the quality assessment. Report the time period for the search in the comment box.</i></p>
<p><i>Comments (note search period, any justification provided for the search period, or uncertainty):</i></p>	

<p>A.4 Was bias in the selection of articles avoided?</p> <p>Did the authors specify:</p> <p><input type="checkbox"/> Independent screening of full text by at least 2 reviewers</p> <p><input type="checkbox"/> List of included studies provided</p> <p><input type="checkbox"/> List of excluded studies provided</p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Partially</p> <p><input type="checkbox"/> No</p> <p><i>Coding guide:</i> YES: All three should be yes, although reviews published in journals are unlikely to have a list of excluded studies (due to limits on word count) and the review should not be penalised for this. PARTIALLY: Independent screening and list of included studies provided are both reported NO: All other. If list of included studies provided, but the authors do not report whether or not the screening has been done by 2 reviewers review is downgraded to NO.</p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>A.5 Did the authors use appropriate criteria to assess the quality and risk of bias in analysing the studies that are included?ⁱⁱ</p> <p><input type="checkbox"/> The criteria used for assessing the quality/ risk of bias were reported</p> <p><input type="checkbox"/> A table or summary of the assessment of each included study for each criterion was reported</p> <p><input type="checkbox"/> Sensible criteria were used that focus on the quality/ risk of bias (and not other qualities of the studies, such as precision or applicability/external validity). “Sensible” is defined as a recognised quality appraisal tool/ checklist, or similar tool which assesses bias in included studies. Please see footnotes for details of the main types of bias such a tool should assess.</p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Partially</p> <p><input type="checkbox"/> No</p> <p><i>Coding guide:</i> YES: All three should be yes PARTIALLY: The first and third criteria should be reported. If the authors report the criteria for assessing risk of bias and report a summary of this assessment for each criterion, but the criteria may be only partially sensible (e.g. do not address all possible risks of bias, but do address some), we downgrade to PARTIALLY. NO: Any other</p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>A.6 Overall – how much confidence do you have in the methods used to identify, include and critically appraise studies?</p> <p><i>Summary assessment score A relates to the 5 questions above.</i></p> <p><i>High confidence applicable when the answers to the questions in section A are all assessed as ‘yes’</i></p> <p><i>Low confidence applicable when any of the following are assessed as ‘NO’ above: not reporting explicit selection criteria (A1), not conducting reasonably comprehensive search (A2), not avoiding bias in selection of articles (A4), not assessing the risk of bias in included studies (A5)</i></p> <p><i>Medium confidence applicable for any other – i.e. section A3 is assessed as ‘NO’ or ‘can’t tell’ and remaining sections are assessed as ‘partially’ or ‘can’t tell’</i></p>	<p><input type="checkbox"/> Low confidence (limitations are important enough that the results of the review are not reliable)</p> <p><input type="checkbox"/> Medium confidence (limitations are important enough that it would be worthwhile to search for another systematic review and to interpret the results of this review cautiously, if a better review cannot be found)</p> <p><input type="checkbox"/> High confidence (only minor limitations)</p>
<p><i>Comments (note important limitations):</i></p>	

Section B: Methods used to analyze the findings

<p>B.1 Were the characteristics and results of the included studies reliably reported?</p> <p>Was there:</p> <p><input type="checkbox"/> Independent data extraction by at least 2 reviewers</p> <p><input type="checkbox"/> A table or summary of the characteristics of the participants, interventions and outcomes for the included studies</p> <p><input type="checkbox"/> A table or summary of the results of all the included studies</p>	<p><input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Partially <input type="checkbox"/> Not applicable (e.g. no included studies)</p> <p><i>Coding guide:</i> <i>YES: All three should be yes</i> <i>PARTIALLY: Criteria one and three are yes, but some information is lacking on second criteria.</i> <i>No: None of these are reported. If the review does not report whether data was independently extracted by 2 reviewers (possibly a reporting error), we downgrade to NO.</i> <i>NOT APPLICABLE: if no studies/no data</i></p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>B.2 Are the methods used by the review authors to analyse the findings of the included studies clear, including methods for calculating effect sizes if applicable?</p>	<p><input type="checkbox"/> Yes <input type="checkbox"/> Partially <input type="checkbox"/> No <input type="checkbox"/> Not applicable (e.g. no studies or no data)</p> <p><i>Coding guide:</i> <i>YES: Methods used clearly reported. If it is clear that the authors use narrative synthesis, they don't need to say this explicitly.</i> <i>PARTIALLY: Some reporting on methods but lack of clarity</i> <i>NO: Nothing reported on methods</i> <i>NOT APPLICABLE: if no studies/no data</i></p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>B.3 Did the review describe the extent of heterogeneity?</p> <p><input type="checkbox"/> Did the review ensure that included studies were similar enough that it made sense to combine them, sensibly divide the included studies into homogeneous groups, or sensibly conclude that it did not make sense to combine or group the included studies?</p> <p><input type="checkbox"/> Did the review discuss the extent to which there were important differences in the results of the included studies?</p> <p><input type="checkbox"/> If a meta-analysis was done, was the I², chi square test for heterogeneity or other appropriate statistic reported? If no statistical test was reported, is a qualitative justification made for the use of random effects?</p>	<p><input type="checkbox"/> Yes <input type="checkbox"/> Partially <input type="checkbox"/> No <input type="checkbox"/> Not applicable (e.g. no studies or no data)</p> <p><i>Coding guide:</i> <i>YES: First two should be yes, and third category should be yes if applicable should be yes</i> <i>PARTIALLY: The first category is yes</i> <i>NO: Any other</i> <i>NOT APPLICABLE: if no studies/no data</i></p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	

<p>B.4 Were the findings of the relevant studies combined (or not combined) appropriately relative to the primary question the review addresses and the available data?</p> <p>How was the data analysis done?</p> <ul style="list-style-type: none"> <input type="checkbox"/> Descriptive only <input type="checkbox"/> Vote counting based on direction of effect <input type="checkbox"/> Vote counting based on statistical significance <input type="checkbox"/> Description of range of effect sizes <input type="checkbox"/> Meta-analysis <input type="checkbox"/> Meta-regression <input type="checkbox"/> Other: specify <input type="checkbox"/> Not applicable (e.g. no studies or no data) <p>How were the studies weighted in the analysis?</p> <ul style="list-style-type: none"> <input type="checkbox"/> Equal weights (this is what is done when vote counting is used) <input type="checkbox"/> By quality or study design (this is rarely done) <input type="checkbox"/> Inverse variance (this is what is typically done in a meta-analysis) <input type="checkbox"/> Number of participants (sample size) <input type="checkbox"/> Other: specify <input type="checkbox"/> Not clear <input type="checkbox"/> Not applicable (e.g. no studies or no data) <p>Did the review address unit of analysis errors?</p> <ul style="list-style-type: none"> <input type="checkbox"/> Yes - took clustering into account in the analysis (e.g. used intra-cluster correlation coefficient) <input type="checkbox"/> No, but acknowledged problem of unit of analysis errors <input type="checkbox"/> No mention of issue <input type="checkbox"/> Not applicable - no clustered trials or studies included 	<ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> Partially <input type="checkbox"/> No <input type="checkbox"/> Not applicable (e.g. no studies or no data) <input type="checkbox"/> Can't tell <p><i>Coding guide:</i> YES: If appropriate table, graph or meta-analysis AND appropriate weights AND unit of analysis errors addressed (if appropriate). PARTIALLY: If appropriate table, graph or meta-analysis AND appropriate weights AND unit of analysis errors not addressed (and should have been). NO: If narrative OR vote counting (where quantitative analyses would have been possible) OR inappropriate reporting of table, graph or meta-analyses. NOT APPLICABLE: if no studies/no data CAN'T TELL: if unsure (note reasons in comments below)</p>
<p><i>Comments (note important limitations or uncertainty):</i></p>	
<p>B. 5 Does the review report evidence appropriately?</p> <ul style="list-style-type: none"> <input type="checkbox"/> The review makes clear which evidence is subject to low risk of bias in assessing causality (attribution of outcomes to intervention), and which is likely to be biased, and does so appropriately <input type="checkbox"/> Where studies of differing risk of bias are included, results are reported and analyzed separately by risk of bias status 	<ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Partially <input type="checkbox"/> Not applicable <p><i>Coding guide:</i> YES: Both criteria should be fulfilled (where applicable) NO: Criteria not fulfilled PARTIALLY: Only one criteria fulfilled, or when there is limited reporting of quality appraisal (the latter applies only when inclusion criteria for study design are appropriate) NOT APPLICABLE: No included studies</p> <p><i>Note on reporting evidence and risk of bias:</i> For reviews of effects of 'large n' interventions, experimental and quasi-experimental designs should be included (if available). For reviews of effects of 'small n' interventions, designs appropriate to attribute changes to the intervention should be included (e.g. pre-post with assessment of confounders)</p>

<i>Please specify included study designs and any other comments (note important limitations or uncertainty):</i>	
<p>B.6 Did the review examine the extent to which specific factors might explain differences in the results of the included studies?</p> <p><input type="checkbox"/> Were factors that the review authors considered as likely explanatory factors clearly described?</p> <p><input type="checkbox"/> Was a sensible method used to explore the extent to which key factors explained heterogeneity?</p> <p style="margin-left: 20px;"><input type="checkbox"/> Descriptive/textual</p> <p style="margin-left: 20px;"><input type="checkbox"/> Graphical</p> <p style="margin-left: 20px;"><input type="checkbox"/> Meta-analysis by sub-groups</p> <p style="margin-left: 20px;"><input type="checkbox"/> Meta-regression</p> <p style="margin-left: 20px;"><input type="checkbox"/> Other</p>	<p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Partially</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Not applicable</p> <p><i>Coding guide:</i> YES: Explanatory factors clearly described and appropriate methods used to explore heterogeneity PARTIALLY: Explanatory factors described but for meta-analyses, sub-group analysis or meta-regression not reported (when they should have been) NO: No description or analysis of likely explanatory factors NOT APPLICABLE: e.g. too few studies, no important differences in the results of the included studies, or the included studies were so dissimilar that it would not make sense to explore heterogeneity of the results</p>
<i>Comments (note important limitations or uncertainty):</i>	
<p>B.7 Overall - how much confidence do you have in the methods used to analyse the findings relative to the primary question addressed in the review?</p> <p><i>Summary assessment score B relates to the 5 questions in this section, regarding the analysis.</i></p> <p><i>High confidence applicable when all the answers to the questions in section B are assessed as 'yes'.</i></p> <p><i>Low confidence applicable when any of the following are assessed as 'NO' above: critical characteristics of the included studies not reported (B1), not describing the extent of heterogeneity (B3), combining results inappropriately (B4), reporting evidence inappropriately (B5).</i></p> <p><i>Medium confidence applicable for any other: i.e. the "Partial" option is used for any of the 6 preceding questions or questions and/or B.2 and/ or B.6 are assessed as 'no'.</i></p>	<p><input type="checkbox"/> Low confidence (limitations are important enough that the results of the review are not reliable)</p> <p><input type="checkbox"/> Medium confidence (limitations are important enough that it would be worthwhile to search for another systematic review and to interpret the results of this review cautiously, if a better review cannot be found)</p> <p><input type="checkbox"/> High confidence (only minor limitations)</p>
<i>Use comments to specify if relevant, to flag uncertainty or need for discussion:</i>	

Section C: Overall assessment of the reliability of the review

<p>C.1 Are there any other aspects of the review not mentioned before which lead you to question the results?</p>	<p><input type="checkbox"/> Additional methodological concerns – only one person reviewing</p> <p><input type="checkbox"/> Robustness</p> <p><input type="checkbox"/> Interpretation</p> <p><input type="checkbox"/> Conflicts of interest (of the review authors or for included studies)</p> <p><input type="checkbox"/> Other</p> <p><input type="checkbox"/> No other quality issues identified</p>
<p>C.2 Are there any mitigating factors which should be taken into account in determining the reviews reliability?</p>	<p><input type="checkbox"/> Limitations acknowledged</p> <p><input type="checkbox"/> No strong policy conclusions drawn (including in abstract/ summary)</p> <p><input type="checkbox"/> Any other factors</p>
<p><i>Use comments to specify if relevant, to flag uncertainty or need for discussion:</i></p>	
<p>C.3 Based on the above assessments of the methods how would you rate the reliability of the review?</p> <p><input type="checkbox"/> Low confidence in conclusions about effects: The systematic review has the following major limitations...</p> <p><input type="checkbox"/> Medium confidence in conclusions about effects: The systematic review has the following limitations...</p> <p><input type="checkbox"/> High confidence in conclusions about effects : If applicable: The review has the following minor limitations...</p> <p><i>Coding guide:</i> High confidence in conclusions about effects: high confidence noted overall for sections A and B, unless moderated by answer to C1. Medium confidence in conclusions about effects: medium confidence noted overall for sections A or B, unless moderated by answer to C1 or C2. Low confidence in conclusions about effects: low confidence noted overall for sections A or B, unless moderated by answer to C1 or C2.</p> <p>Limitations should be summarized above, based on what was noted in Sections A, B and C.</p>	

ⁱ Adapted from the SURE Collaboration (2011), 'SURE checklist for making judgements about how much confidence to place in a systematic review'. In: SURE guides for preparing and using policy briefs. Version 2.1 [updated November 2011]. The SURE Collaboration, 2011. Available from www.evipnet.org/sure.

ⁱⁱ **Risk of bias** is the extent to which bias may be responsible for the findings of a study. **Bias** is a systematic error or deviation from the truth in results or inferences. In studies of the effects of social, economic and health care interventions, the main types of bias arise from systematic differences in the groups that are compared (selection bias), the intervention that is provided, or exposure to other factors apart from the intervention of interest (performance bias/contamination), withdrawals or exclusions of people entered into a study (attrition bias) or how outcomes are assessed (detection bias) and reported (reporting bias). Reviews of social science studies may be particularly affected by reporting bias, where a biased subset of all the relevant data and analyses is presented.

Assessments of the risk of bias are sometimes also referred to as assessments of the **validity** or **quality** of a study.

Validity is the extent to which a result (of a measurement or study) is likely to be true.

Quality is a vague notion of the strength or validity of a study, often indicating the extent of control over bias.