

MINISTRY OF FOREIGN AFFAIRS OF DENMARK

DANIDA | INTERNATIONAL
DEVELOPMENT COOPERATION



Systematic Reviews Questions, Methods and Usage



EVALUATION STUDY

2013/1

Evaluation Study

Systematic Reviews Questions, Methods and Usage

May 2013

Henrik Hansen, Department of Economics, University of Copenhagen

Neda Trifkovic, Department of Food and Resource Economics, University of Copenhagen

The views expressed are those of the authors and do not necessarily represent the views of the Ministry of Foreign Affairs of Denmark. Errors and omissions are the responsibility of the authors.

Table of Contents

Executive summary	1
Abbreviations	3
1. Introduction	4
2. Background: history and expansion	6
3. Defining systematic reviews	8
4. Methods and data	12
5. Results	14
5.1. The protocol	18
5.2. Primary studies and primary research methods.....	18
5.3. Methodology of conducting systematic reviews	21
5.4. Comprehensiveness of the review	24
5.5. Questions addressed in systematic reviews.....	25
5.6. Usage of systematic reviews	25
6. Typologies of systematic reviews.....	27
7. Systematic reviews in international development.....	32
8. Conclusion.....	38
References	40
Appendix	48
A1. A brief history of systematic reviews	48
A2. Databases used for the study	49
A3. Glossary	52
A4. Specific tools for quality assurance of systematic reviews	56

Executive summary

Policymakers and researchers all value systematized research evidence. Particularly relevant research synthesis products are systematic reviews that present a well-established way of mapping all relevant evidence, assessing its quality and synthesizing it. Systematic reviews have a long history in medicine, from where they have diffused into many arts and science disciplines – including, recently, international development.

In this evaluation study we provide a mapping of the use of systematic reviews across the arts and sciences. Based on the mapping exercise, we assess the extent to which the practice of systematic reviewing in the context of development interventions corresponds to practices in other related fields of research.

By analyzing 49 different information sources, including academic and institutional databases, journals and books, we identify several patterns in systematic reviewing practice. The structure and methods for conducting the review, the criteria for including primary studies, the comprehensiveness of the search, the criteria used in assessing the quality of the individual studies, the questions addressed and the usage of the review are important, if not essential, components of a systematic review. We use these six main features of systematic reviews to organize our findings around several themes that enable the appraisal of systematic reviewing practice.

Many practitioners agree that the main defining characteristic of *systematic* reviews is the existence of a protocol that ensures reproducibility and provides the essential procedures for conducting the review. A systematic review usually proceeds in several stages: formulating the review question(s), planning the review, locating studies (literature search), appraising contributions, analyzing and synthesizing information and reporting the best available evidence.

We stress the importance of the question addressed in the systematic review, because all characteristics of systematic reviews can be traced to the research question. The question determines not only the objectives, but also the stages of conducting the review: selection and appraisal of primary studies, methods in which these studies are synthesized and the comprehensiveness, i.e. the coverage of the final review.

We have identified three broad groups of review questions: *focused questions* that aim to determine ‘what works’ for various interventions; *complex questions* that aim to address context-sensitive issues such as people’s attitudes and experiences, environmental or practical concerns and *hybrid questions* that tackle both focused and complex issues. From the three broad categories, we further define five specific subgroups of questions: effect-driven, explanatory, economic, hermeneutic and mixed. We transpose each review question type into a specific type of systematic review.

The typology of the systematic reviews we propose may prove useful in dealing with complexities caused by different types of review products that are conducted with different users and purposes in mind. It can also serve as a tool for guiding the review process and identifying the most suitable evidence synthesis methods. For example, although systematic reviews have a quantitative tradition, we are witnessing an increased recognition of qualitative studies included as primary types of evidence in systematic reviews and qualitative methods that are used to synthesize available evidence. This is particularly true for more recent systematic reviews in international development and other social sciences that assess diverse intervention programs and face a wide range of outcome indicators while addressing complex research questions.

Systematic reviews in international development mostly address health-related interventions, but the donor involvement since the mid-2000s has resulted in a broader range of topics in which a specific intervention is assessed over a range of variables and intervention groups. Still, the focused questions dominate in estimating direct, easily measurable effects or the intervention impacts even though the development reviews operate in a complex multidisciplinary environment, which requires acknowledging the influence of institutions and social interaction. In addition, the scarcity of comparable evidence about the effects of development interventions necessitates that authors change their strategies when assessing the strength of evidence or synthesizing data. For example, narrative synthesis is used to systematize empirical evidence when data quality and quantity do not allow meta-analysis, which is the preferred method for calculating the effect size in more traditional systematic reviews.

Due to the inherent differences in value judgments, different ways of reviewing and interpreting the same data (evidence) can lead to conflicting conclusions. The focus on asking the ‘right’ questions in international development reviews is important precisely because no review process is immune to bias. We emphasize in this study that systematic reviews in international development may be vulnerable to a range of biases and warn that these reviews should not aim, at all cost, at pursuing the classical approach suitable for traditional, ‘easy-to-measure’ situations. Instead, the development reviews should adjust the review process so it caters for the type of question they are trying to address. In this way, the differences in type and quality of the included primary studies, methodological approach and the study comprehensiveness will not be a source of bias, but will add to the overall success of the review.

Abbreviations

3ie – International Initiative for Impact Evaluation
AMSTAR – Assessment of Multiple Systematic Reviews
CRD – Centre for Reviews Dissemination, University of York
DFID – Department for International Development, UK
DIME – Development Impact Evaluation, World Bank
EPPI-Centre – Evidence for Policy and Practice Information and Coordinating Centre
ERIC – Education Resources Information Center
ICT – Information and communications technology
IPA – Innovations for Poverty Action
JPAL – Abdul Lateef Jameel Poverty Action Lab
NHS – Evidence in Health and Social Care
OQAQ – Overview Quality Assessment Questionnaire
PICOS – Participants, Interventions, Comparators, Outcomes, and Study design
PRISMA – Preferred Reporting Items for Systematic reviews and Meta-Analyses
QUORUM – Quality Of Reporting Of Meta-analyses
RCT – Randomized Controlled Trial
SIEF – Strategic Impact Evaluation Fund, World Bank
SIGLE – System for Information Retrieval on Gray Literature
SURE – Supporting the Use of Research Evidence

1. Introduction

Systematic reviews are nowadays an indispensable component of both scientific and policy life. Through systematic reviews huge numbers of scientific studies and analyses are condensed and transformed to knowledge accessible to policy-makers and the broader scientific community. Some consider the use of systematic reviews to be increasing to an extent where it is starting to replace primary research, in particular considering health care decisions (Evans & Pearson, 2001).

A systematic review is a well-established way of impartially mapping the relevant evidence, assessing the quality of the evidence and synthesizing it. Systematic reviews are focused on reporting what is known and what is not known about a specific question that usually has high policy relevance. But there is no consensus about the application of systematic reviews. While some investigate treatment effectiveness, diagnoses or epidemiology, other reviews may focus on measurement or the methodological rigor of primary studies. The defining attribute is that a *systematic* review offers objective inference based on available evidence and not a description of everything on the subject.

Systematic reviews have a long history in health-related disciplines, while the tradition is much shorter in most scientific and social science disciplines. Within medicine, systematic reviewing was introduced in the 1970s and further developed in the 1980s, leading to the establishment of the Cochrane Collaboration in 1992. Further on, the EPPI-Centre has been undertaking systematic reviews in education since the early 1990s and the Campbell Collaboration in the context of social policy since 2000. International development is one field in which the attention to systematizing the many individual pieces of empirical evidence has only recently come to focus, particularly because of the donor interest in evidence-based policies. The lack of systematic reviews in development research is sometimes cited as a reason for pursuing RCTs in developing countries, for example in the case of the Abdul Lateef Jameel Poverty Action Lab projects (see e.g. Duflo et al., 2008; Baird et al. 2012).

Although evidence gathering and synthesis in the context of development interventions inarguably face challenges distinct from other fields, it is highly likely that it can benefit from the experiences and advances of the systematic review methodology in other contexts. Thus, even though the interest of Danida is primarily within systematic reviews of development interventions, it is equally useful to have a broader understanding of the practice of systematic reviewing across all art and science disciplines. The aim of this study is, therefore, twofold. First, we provide a mapping of the use of systematic reviews across the arts and sciences. Second, based on the mapping exercise, we assess the extent to which the practice of systematic reviewing in the context of development interventions corresponds to practices in other, related, scientific fields.

We examine the occurrence of systematic reviews within various disciplines as well as the types of systematic reviews that occur in different research areas, within which topics and with what research objectives. We also develop an operational definition of ‘systematic review’, which is sufficiently broad to encompass both traditional systematic reviews, such as those undertaken under the auspices of the Cochrane and Campbell Collaborations, and more unconventional ones, such as realist reviews, employed within the social sciences. Our definition is, at the same time, suitably restricted to allow for establishment of clear criteria for distinguishing between systematic reviews and other forms of literature reviews. The definition is based on examining the methodological literature and statements from major institutions working with systematic reviews. This process has enabled us to describe and identify criteria for classification of systematic reviews by type.

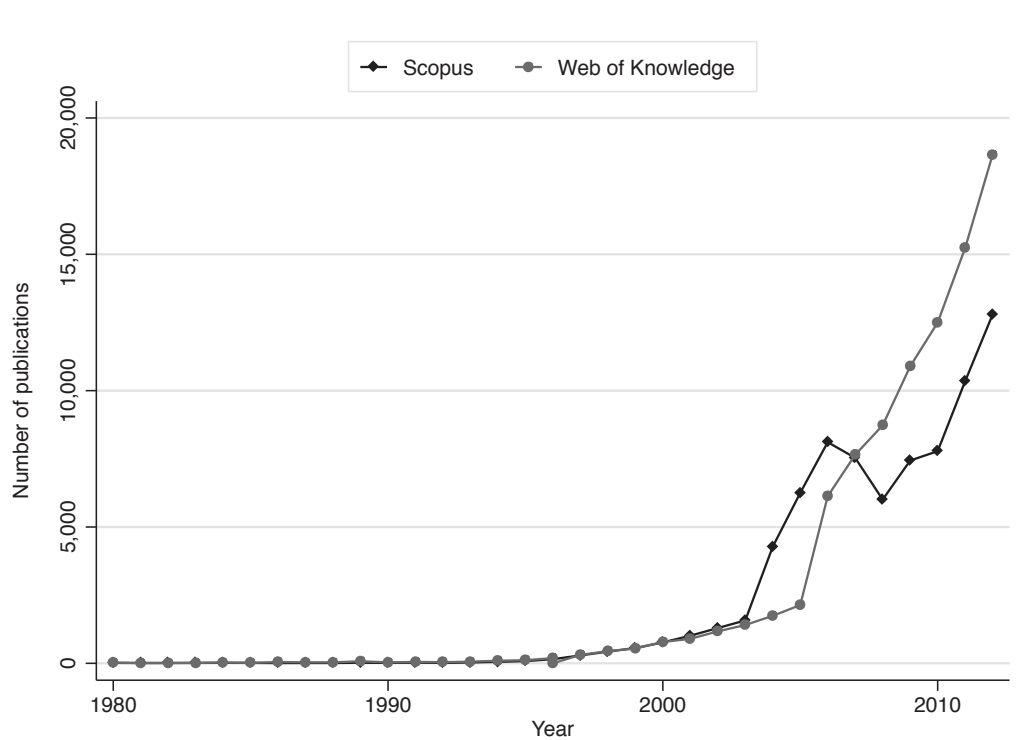
In this way, we envisage that the present study will provide a comprehensive background for a more profound discussion of the appropriateness of the current practice of systematic reviewing in international development and the role that such reviews can, and should, play in evidence-based knowledge creation and dissemination, informing the design of future development interventions.

2. Background: history and expansion

Historical accounts are not consistent neither when it comes to determining when the first systematic review was conducted, nor which review it was. It is apparent that systematic reviews have a long record in medicine and health care (Smith et al., 1980), but some consider that systematic reviews originate from educational research (Smith & Glass, 1980). Several sources trace the beginning to the first meta-analysis by the statistician Gene Glass in 1976 (Davies, 2000). Others attribute the invention of systematic reviews to Scottish naval surgeon James Lind, who is also considered to be the inventor of RCTs (Chalmers et al., 2002; Dunn, 1997; Lind, 1753). Still other researchers describe the production of systematic reviews since the foundation of the *Oxford Database of Perinatal Trials* project in 1985 (Chalmers et al., 1986).

Browsing scientific databases also results in ambiguities. The Web of Knowledge shows that the first use of the phrase ‘systematic review’ occurred in 1916, Scopus has 1945 as the earliest year, while ProQuest reports 1905. However, it is unlikely that the early papers encompass all the traits of a structured process present in modern-day systematic reviews. From the 1930s onwards, the term ‘systematic review’ was used to refer to literature reviews and the early examples of literature reviews often described themselves as ‘systematic literature surveys’ (Petticrew & Roberts, 2008). The use of the term ‘systematic review’ spread rapidly in the 1970s, but it was not until the 1990s that the term became used extensively, as shown in Figure 1.

Figure 1. Number of publications in Web of Science and Scopus using the term ‘systematic review’



Source: Authors' elaboration.

A range of conceptual initiatives preceded the introduction of systematic reviews, primarily the methods for analyzing groups of experiments, or combining data and results from independent studies (Table A1, Appendix A1). Advances in statistics, such as the least squares method, correlation coefficient and p -value were necessary for a successful development of systematic review practice. Statistical pooling of findings from primary studies that grew from Glass and Smith's evaluation of outcomes in psychotherapy and counseling (Smith & Glass, 1977; Glass, 1976) – a practice known as meta-analysis – is a common component of most systematic reviews. Later on, the recognition that the medical profession needs 'critical summaries' of RCTs led to the establishment of a collaborative database of perinatal trials (Cochrane Collaboration, 2012a), which is considered a cornerstone of modern systematic review practice. However, the medical profession was not alone in making explicit efforts to limit bias in the review of literature. Similar efforts have been reported by social scientists at least since the 1960s (Chalmers et al., 2002). Petticrew and Roberts (2008, p. 19) conclude that 'contrary to what is commonly supposed, neither the term "systematic review" nor the general approach of systematic literature reviewing are particularly new, nor particularly biomedical'.

3. Defining systematic reviews

There is no shortage of definitions of systematic reviews. For some, it is simply a process of offering accountable, replicable and updateable piece of research to the involved users (EPPI-Centre, 2009a). For others, it is more narrowly defined. The Campbell Collaboration defines a systematic review as *‘a transparent procedure to find, evaluate and synthesize the results of relevant research’*. According to the Cochrane Collaboration, a systematic review is a *‘high-level overview of primary research on a particular research question that tries to identify, select, synthesize and appraise all high quality research evidence relevant to that question in order to answer it’*. As such, systematic reviews seek, summarize and interpret primary studies while attempting to provide unbiased research evidence on a given topic. They need to be rigorous in their approach to summarizing and interpreting the evidence. If not, they are *‘little more than ... subjective commentaries on the state of the science’* (Weed, 2013, p. 280).

The aim of a systematic review, in its original form, is to produce results that are generalizable to other contexts such that it can be used to make reasonable predictions of future events (Briner & Denyer, 2012). And, as a latecomer in the field, DFID states that *‘Systematic reviews ... make it easier for policy makers and practitioners to rapidly understand the body of evidence and use this as a strong foundation on which to base policy and practice decisions’* (DFID, 2012).

Several organizations involved in producing systematic reviews have issued guidelines on how to plan and structure a systematic review in order to help minimize bias and enhance transparency and objectivity. Guidelines have been elaborated in medicine (Higgins & Green, 2011), social sciences (Petticrew & Roberts, 2008) and computer science (Kitchenham & Charters, 2007). In Table 1 we give the key elements of systematic reviews from the Cochrane and Campbell Collaborations.

Table 1. Key elements of a systematic review

Cochrane Collaboration	Campbell Collaboration
1. Identification of relevant studies from a number of different sources (including unpublished sources)	1. A systematic search for unpublished reports (to avoid publication bias)
2. Selection of studies for inclusion and evaluation of their strengths and limitations on the basis of clear, predefined criteria	2. International scope
3. Systematic collection of data	3. A protocol (project plan) for the review is developed in advance and undergoes peer review
4. Appropriate synthesis of data	4. Study inclusion and coding decisions are accomplished by at least two reviewers who work independently and compare results
	5. Peer review and editorial review

Source: Authors' elaboration, Cochrane Collaboration (2012a) and Campbell Collaboration (2012).

In analyzing the different guidelines, we identified several overlapping traits regarding the stages and information any systematic review needs to contain. Broadly speaking, a systematic review involves the following steps:

1. Formulating the review question(s),
2. Planning the review,
3. Locating studies (literature search),
4. Appraising contributions,
5. Analyzing and synthesizing information and
6. Reporting the best available evidence.

Unlike other forms of research synthesis, systematic reviews are based on a rigorous protocol (a standard set of stages) for organization and systematization of research results. Therefore, the *first criteria* by which to identify a systematic review is the existence of a research protocol, which guarantees that the review process can be replicated.

Second, the search and selection of primary research to be included in a systematic review needs to be comprehensive enough to include both electronic and print sources, as well as unpublished material and gray literature. The purpose is to avoid reliance on anecdotal evidence and 'cherry picking' of favorable cases.

Third, clear criteria for inclusion of primary research must be established for a systematic review. The quality of primary research plays a decisive role. Primary research needs to be appropriate for review not only in terms of topic, but also in terms of the rigor and success with which the research was conducted.

Fourth, a systematic review needs to include an analysis of the primary research. The analysis can be quantitative (most commonly meta-analysis) or qualitative (e.g., thematic synthesis). Purely quantitative findings of primary research can be analyzed qualitatively through methods such as narrative synthesis. The opposite case can be found as well – qualitative research can be quantitatively incorporated into systematic reviews through vote counting.

Fifth, a systematic review must include a synthesis of the information contained in the primary research. Hence, a systematic review must go beyond the simple summary of primary research findings. Examples of synthesis activities include: assessing the size of the effect of a given treatment, assessing the causes of a given outcome, assessing the consistency across different studies and assessing the quality of primary data. In addition to presenting key findings, a systematic review should identify reasons for differing results across studies and state limitations of current knowledge.

Sometimes the term ‘systematic review’ is used loosely, without particular reference to the rigorous approach to literature synthesis. In these instances, the authors neither aim to assess available evidence nor to provide an answer to a specific, policy-related question, but to provide an overview or to describe practices of, say, cases, research methods and measurements. It is usually assumed in such articles that if one conducts a broad review of the literature then one is conducting a systematic review. However, such articles cannot be classified as true systematic reviews because they lack a replicable search protocol, methods and criteria for selecting relevant literature.

We also find that some expressions are used synonymously with the term ‘systematic review’. For example, structured review, scoping review and systematic mapping appear under the ‘systematic review’ heading in EPPI-Centre’s library. Moreover, terms such as: systematic synthesis of research, systematic literature review, in-depth review and narrative review are often used to signify ‘systematic review’. Also, a meta-analysis is commonly considered as a form of systematic review, not only a form of statistical technique for secondary data analysis (see Appendix A3). While these different forms of reviews offer relevant information about empirical knowledge from experienced authors, they typically use an implicit – and not always replicable – process in assimilating evidence to support the statements being made. In case of narrative reviews, it may be difficult for the reader to determine if the statements are based on the author's experience or the range of available literature. It may be equally difficult to identify the reasons why some studies were given more emphasis than others and whether some reports were selectively cited to reinforce preconceived views of a topic (Garg et al., 2008).

The quality of systematic reviews is often disputed because the review process may suffer from several forms of bias. The review process inevitably includes studies that are diverse in their design, methods and types of data used, so a reviewer’s decision about which studies to include, how to assess and combine them, and how to conclude remains sensitive to subjective judgment. While evaluating the quality of systematic reviews in the emergency medicine literature, Kelly et al. (2001) concluded that the overall scientific quality of the reviews is low and that only 10% of reviews had minimal flaws. To increase the reliability of systematic reviews, various quality assessment tools have appeared over the past two decades. We briefly describe the most relevant ones in the Appendix A4. It appears, however, that these tools are not frequently used, which undermines the efforts of the whole systematic review process.

Lavis (2009) explains that policymakers and stakeholders have access to at least three types of review-derived products: summaries of systematic reviews highlighting decision-relevant information; overviews of systematic reviews providing a ‘map’ of what policy questions have been addressed by systematic reviews and where additional reviews are needed; and policy briefs drawing on many systematic reviews to better understand a problem and possible implementation strategies. In addition, rapid evidence synthesis reports (differently called rapid reviews, defined in Appendix A3) are competing with systematic reviews for the attention of users. Neither the Campbell nor the Cochrane Collaboration offer rapid evidence syntheses. Analyzing the MEDLINE database between 1950 and 2007, Bastian et al. (2010) have identified a rise in non-systematic reviews, case reports and trials, which surpasses the rise in systematic reviews. They conclude that ‘the staple of medical literature synthesis remains the non-systematic narrative review’ (Bastian et al., 2010, p.1).

4. Methods and data

The data for the present study are from meticulous database searches that took place between 1 December 2012 and 28 February 2013. The literature search included databases dedicated to systematic reviews; several medical and scientific databases; larger subject and multidisciplinary databases; specific journals and databases maintained by research institutions and international organizations. We searched 49 different sources of information about systematic reviews as shown in Table 2. More detailed information about these sources is in Appendix A2.

Table 2. Sources of information about systematic reviews

Source type	Prevalence	Source names
Academic databases	53%	Web of Knowledge*, ProQuest, WorldCat, Scopus, Taylor and Francis Online, Wiley Online Library, Science direct, Unbound Medline*, MEDLINE, BioMed Central, Ebsco: Academic Search Complete, SAGE journals, Emerald, Ingentaconnect, Annual Reviews, SSRN, Sociological Abstracts, International Political Science Abstracts, PAIS, CAB Abstracts, Communication and Mass Media Complete, EconLit, JSTOR, International Bibliography of the Social Sciences, PsycINFO, SocINDEX
Specialized databases	14%	Cochrane Collaboration*, Campbell Collaboration, Centre for Reviews Dissemination, Evidence in Health and Social Care, Health-Evidence.ca*, Open Grey*, ERIC
Institutions	19%	EVIPNet, World Health Organization*, 3ie, World Bank*, ODI*, EPPI*, Collaboration for Environmental Evidence*, DFID, AusAID*
Journals	14%	Systematic Reviews Journal, Journal of Development Effectiveness, Research Synthesis Methods, The Lancet, Trials, PLOS ONE, Evidence-based medicine

Note: * indicates that the database search could not be restricted to Title, Abstract and Keywords, which was the preferred search option; the search was performed on full text instead. Source: Authors' elaboration.

The search term 'systematic review' is restricted to scientific disciplines as each specific database allows. In some databases, the range of scientific areas is pre-determined and sub-divisions are not possible, while others allow a finer, more customized search. We give an overview of the search stages across scientific disciplines, topics and relevant international development phrases in Table 3. Applying this set of search criteria has enabled positioning the systematic reviews on international development within the broader frame of the arts and sciences. When the database search system allowed, the search was restricted to abstracts, titles and keywords. When this option was not available, we searched for the terms of interest in the abstract only or, as the least preferred option, in the full text. We avoided searching for the keywords in full text because some keywords may appear in the text without capturing the essence of the article. As expected, the outcome is that databases, which do not allow restricted search, show more articles that are irrelevant for our purposes. We mark these by asterisks in Table 2. For this study, more than 50% of the information came from academic databases, around 20% from different institutional databases and some 15% from topical journals and other databases.

Table 3. Scientific areas, topics and methods relevant for systematic reviews

a) We searched for 'systematic review' within following scientific fields:			
Agriculture	Earth and planetary science	Health, medicine,	Politics, political sciences
Anthropology	Economics, finance, business,	dentistry, nutrition and	and international
Area studies	management, marketing, accounting	nursing	relations Social care,
Arts	Education	Humanities	social work and social
Behavioral sciences	Energy (bio-energy, fossil fuels, bio-	Information science	services
Bioscience (biology, life	fuels)	Language, linguistics and	Sociology
sciences, plant science)	Engineering and technology	literature	Social sciences
Built environment	Environmental studies and	Law	Sports and leisure
Communication studies	management	Mathematics	Statistics
Computer science	Food science and technology	Meteorology (weather)	Travel and tourism
Cultural studies	Geography	Physical sciences	Urban studies
Development studies		(physics)	Zoology, animal health
			and veterinary medicine
b) We searched for 'systematic review' and following topics of interest:			
Africa	Development assistance	Development work	International
			development
Asia	Development cooperation	Economic development	Latin (South) America
Developing countries	Development intervention	Economic growth	Low-income countries
Development aid	Development studies	Global development	Transition economies
c) We searched for following research synthesis methods:			
Aggregated analysis	Integrated review	Qualitative meta-analysis	Scoping review
Aggregated synthesis	Meta-analysis	Qualitative synthesis	Supplementary analysis
Content analysis	Meta-ethnography	Research synthesis	Structured review
Critical appraisal	Meta-narrative	Rapid evidence	Synthesis of qualitative
Critical evaluation	Meta-regression	assessment (rapid review)	research
Critical synthesis	Meta-study	Realist review	Systematic mapping
Evidence synthesis	Meta-synthesis	Re-analysis	Narrative synthesis
Evidence-based review	Mixed-methods synthesis	Review of reviews	Thematic synthesis
Information synthesis			

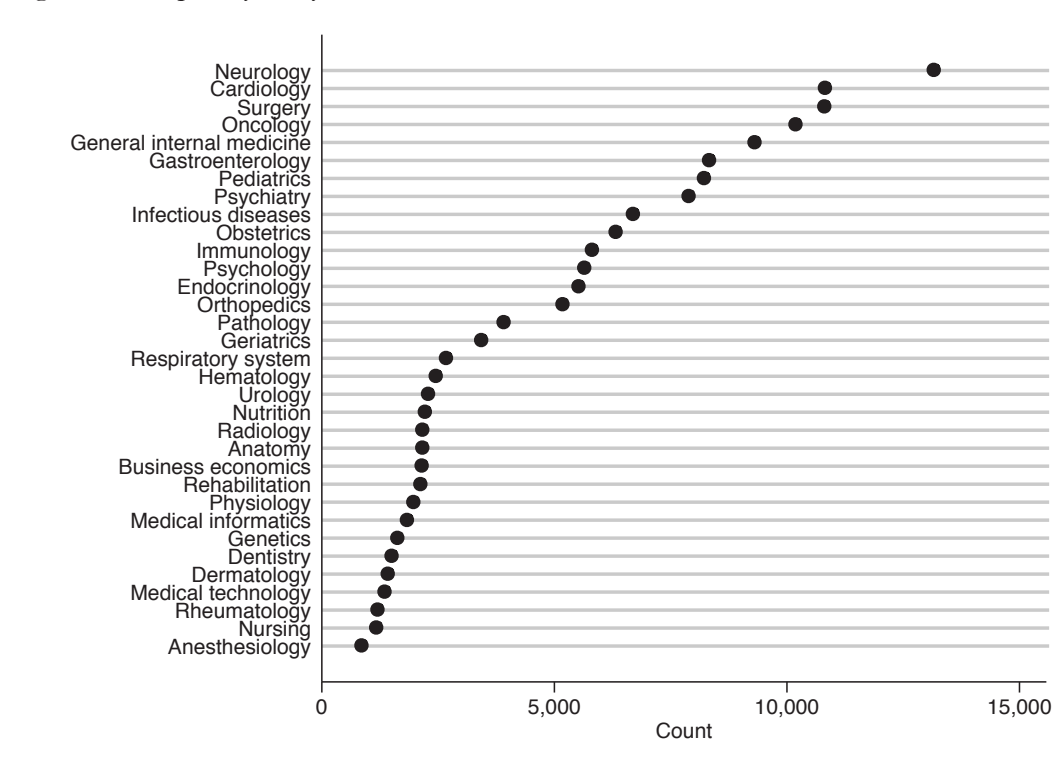
Source: Authors' elaboration.

5. Results

The popularity of systematic reviews stretches across all fields of research. Our search returned articles containing the phrase ‘systematic review’ in practically all arts and sciences. However, the search outcomes were disproportionally in medicine and health-related fields. Still, the systematic reviewing practice is spreading rapidly within the broadly defined fields of life sciences and education. The least common was to find systematic reviews in philosophy, arts and humanities.

Systematic reviews are conducted in almost any branch of medicine, as Figure 2 shows. We find the highest occurrence in neurology, followed by cardiology and surgery, where systematic reviews are ten times more frequent than in the smaller areas of medicine, such as nursing or anesthesiology. These differences are linked to the tradition of systematic reviews, which in general focus on the effect size and give preference to data obtained through experimental designs (e.g., RCTs). The observational and interpretative research methods are either excluded from the reviews or classified as lower-level forms of evidence (Evans & Pearson, 2001).

Figure 2. Frequency of systematic reviews in different branches of medicine



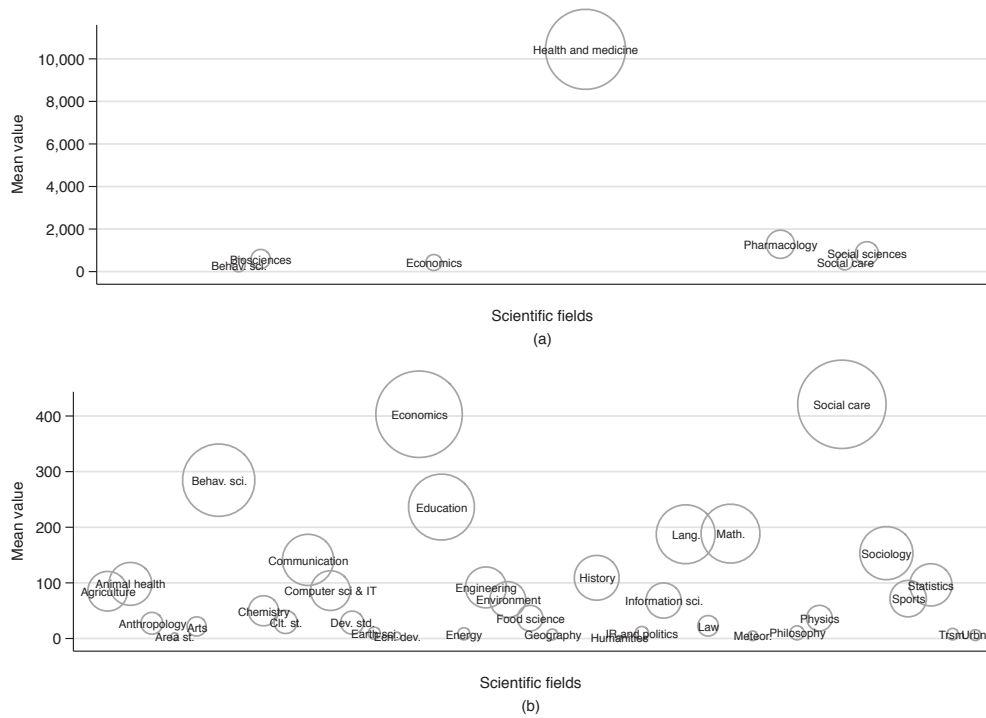
Source: Authors' elaboration based on Web of Knowledge (2013).

In Figure 3 the average number of systematic reviews across different databases are given. The mean value represents the average number of systematic reviews in 26 databases.

To separate large and small values, we split the figure in two panels so that panel (a) shows scientific fields in which systematic reviews appear the most, while panel (b) shows scientific fields in which systematic reviews occur less frequently. For example, systematic reviews are 25 times more prevalent in medicine (around 10,000 reviews) than in economics and social care (around 400 reviews). The early topics of systematic reviews in non-medical fields are often related to various aspects of health. For example, in the social sciences, such as anthropology and sociology, the earliest topics of systematic reviews were measures of social psychological attitudes (Robinson & Shaver, 1973), abortion support (Rosoff, 1975) and immigrant mortality (Marmot et al., 1984). In economics, the interests revolved around costs of specific treatments or health services, such as costs of drug prescription (Vuturo et al., 1980) and cost-effective choices of anti-microbial therapy (Weinstein et al., 1986). Interestingly, the early mentions of systematic reviews in political science were not related to health, but to e.g. urban political processes (Schnore & Fagin, 1967) and the Taiwan issue in Peking's foreign relations (Shen, 1981).

The closer inspection shows how articles appearing in non-medical fields are in fact often applications of medicine to other disciplines. This is also the case for the arts as articles classified under the subject heading 'Art' are mostly concerned with medical uses of art, *i.e.*, art as a form of therapy for different medical conditions. Likewise, for language and literature, the topics covered in systematic reviews are mostly related to language-learning abilities and treating speech impairments. The high numbers of systematic reviews in mathematics and engineering visible in Figure 3 are also driven by medical research.

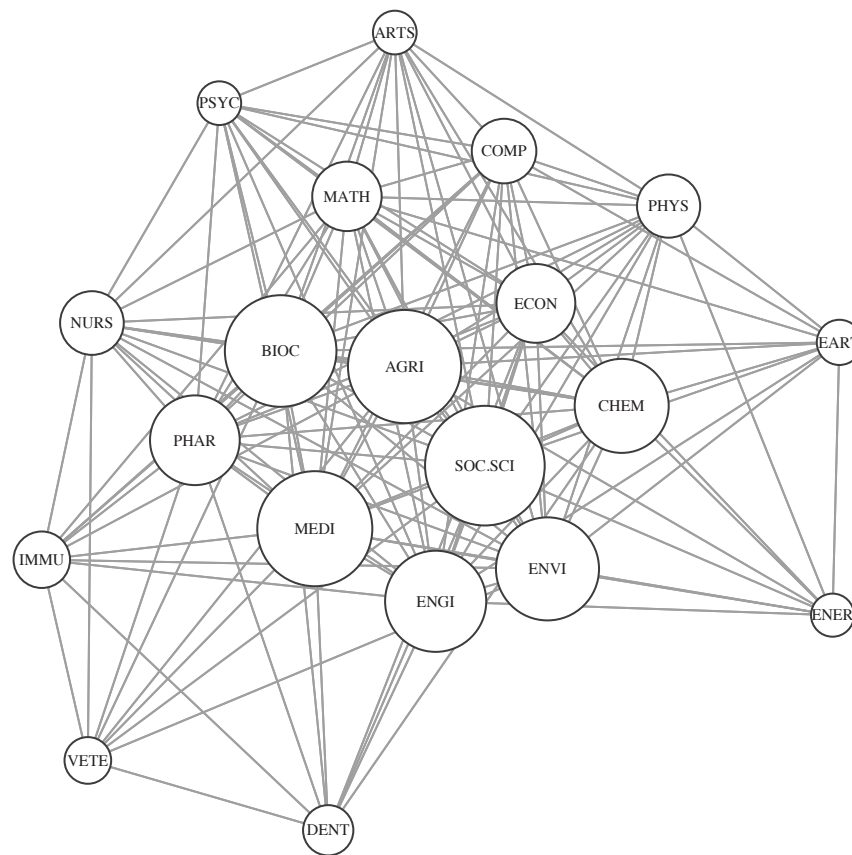
Figure 3. Prevalence of systematic reviews across different scientific disciplines



Source: Authors' elaboration.

We investigate the transfer of influence from medicine to other research areas through the degree of betweenness centrality (Freeman, 1977), which we show in Figure 4. The figure measures both the frequency of systematic reviews in different disciplines, as Figure 3, and also the importance of different disciplines for the spread of information about systematic review practice. In Figure 4, the size of the circle is proportional to the frequency of systematic reviews in each scientific field while the lines between circles show which scientific fields are connected in the systematic review practice. The lines illustrate the co-occurrence of systematic reviews in different scientific fields and they are constructed by searching for 'systematic reviews' in 'one scientific field' and 'other scientific field', e.g. 'systematic reviews' in fields 'medicine' AND 'dentistry'. The figure shows that, for example, dentistry is connected with medicine, but not with energy research, nursing or earth sciences.

Figure 4. The importance of medicine for systematic review practice relative to other research areas



Note: The following scientific fields are included: AGRI: Agricultural and Biological Sciences; ARTS: Arts and Humanities; BIOC: Biochemistry, Genetics and Molecular Biology; CHEM: Chemistry and Chemical Engineering; COMP: Computer Science; DENT: Dentistry; EART: Earth and Planetary Sciences; ECON: Economics, Econometrics, Finance, Business, Management and Accounting; ENER: Energy; ENGI: Engineering; ENVI: Environmental Science; IMMU: Immunology and Microbiology; MATH: Mathematics; MEDI: Medicine and Health; NEUR: Neuroscience; NURS: Nursing; PHAR: Pharmacology, Toxicology and Pharmaceuticals; PHYS: Physics and Astronomy; PSYC: Psychology; SOC. SCI: Social Sciences; and VETE: Veterinary medicine. Source: Authors' elaboration.

While medicine has a strong influence on the increase in systematic review practice in biosciences, agriculture, environmental and social sciences, other health-related fields, such as pharmacology, psychology, nursing, immunology, dentistry and veterinary medicine do not. There is also a strong link between medicine and engineering, due to the importance of medical engineering. Finally, energy, earth sciences, physics and arts are on the periphery of systematic review practice, while computer sciences, chemistry, mathematics and economics appear closer to the main information sources.

We were able to identify several patterns in systematic reviewing practice. It is apparent that the protocol, the criteria for including the primary studies to be summarized and interpreted, the comprehensiveness of the search, the criteria to be used in assessing the quality of the individual studies, the questions addressed and the usage of the review

are important, if not essential, components of a systematic review. We use these features of systematic reviews to organize our findings around several topics that enable the appraisal of systematic reviewing practice.

5.1. The protocol

As stated in Section 3, the main defining characteristic of systematic reviews is the existence of a protocol describing the essential procedures for conducting the review. Cook et al. (1995) suggest that a protocol should include: a question that specifies the population, intervention and outcomes of interest; specification of the methods used to retrieve, select, assess and analyze relevant data; specification of the hypothesis-testing analyses and disclosure of any changes in the protocol. The protocol format of the Cochrane Collaboration requires certain information to be presented: funding sources, the text of the review (background, objectives, criteria for selecting studies for this review, search strategies for identification of studies, methods of the review, acknowledgements, conflicts of interests), references, tables and figures, and comments and criticism (Higgins & Green, 2011). Interestingly, though, Moher et al. (2007) revealed that less than half of the systematic reviews in their sample were working from such a protocol.

5.2. Primary studies and primary research methods

The objective to consider a wide range of primary literature sources while looking for policy-relevant evidence is also well reflected in our results. Literature sources typically include research and institutional databases, journals, book chapters, trial registries and gray literature. It is actually often required that a search of gray literature is performed so the working and conference papers, pre-prints and similar materials are not omitted from the search simply because they may be difficult to find. These are important requirements since it has been documented that published studies in medicine tend to have more positive results, whereas unpublished studies tend to show smaller effects or even insignificant findings (Schlosser, 2007). However, as gray literature is not subject to peer-review, it must be considered accordingly (Schlosser, 2007). The good practice for conducting systematic reviews proposes that reviewers examine differences between outcomes of published and unpublished studies. In this way, it is possible to eliminate doubt that the actual size of the estimated effect may in fact be lower due to publication bias.

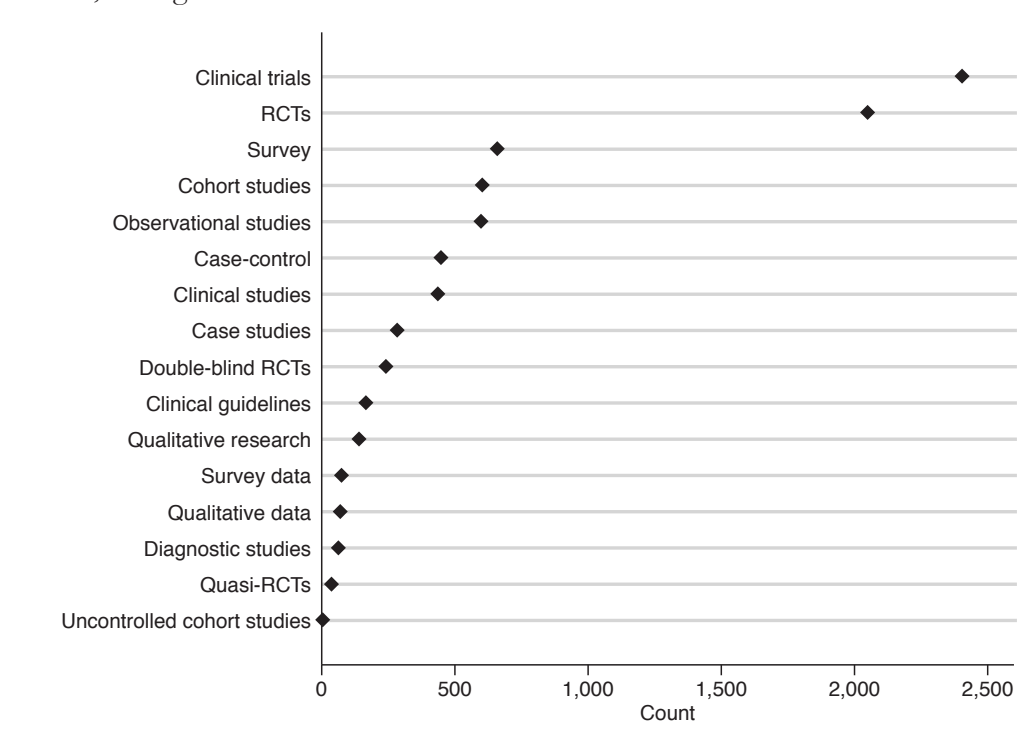
Systematically reviewing evidence often means considering various forms of empirical data. The differences in quality of evidence require rigor in data selection. Even though the concept of quality in research is elusive, several guidelines on how to judge research quality have been developed. The quality assessments are usually made against checklists or validated scales that consider suitability of study design to the research objective, risk of bias, choice of outcome measures, statistical issues, quality of reporting, quality of the intervention and generalizability. Several explicit quality assessment guidelines issued by various institutions are in use (Atkins et al., 2004). RCTs, cohort, diagnostic and economic

studies present the most reliable forms of evidence, whereas diagnostic and case-control studies present the least reliable evidence that can be incorporated in systematic reviews (CEBM, 2013). There are also quality scales devised specifically for primary methods. For example, a so-called Jadad Scale is used to assess the quality of RCTs (Jadad et al., 1996). Frequently used is the Checklist for Measuring Quality, which is applicable to both randomized and non-randomized studies of health care interventions (Downs & Black, 1998).

Unfortunately, this plurality of systems for grading the quality of evidence is a source of frustration, if not confusion, for many. The same evidence and recommendation could be evaluated as ‘the best’ or only as ‘good’ depending on the system used. An evidence-based medicine workgroup has therefore developed the GRADE (Grades of Recommendation, Assessment, Development and Evaluation Working Group) system that addresses the major limitations of its predecessors. Cochrane Collaboration relies exclusively on this approach for evaluating the quality of evidence (Higgins & Green, 2011) and the *British Medical Journal* states that authors preferably should use the GRADE system (BMJ, 2013).

We conducted an extensive search in 42 academic and institutional databases to uncover the preferred primary research methods. We show definitions of these research methods in Appendix A3. The search results are presented in Figure 5 from which it is clear that systematic reviewers rely heavily on clinical trials and the empirical evidence obtained through RCTs. Drastically less common sources of evidence incorporated in systematic reviews are surveys, cohort and observational studies. Only very few systematic reviews rely on diagnostic studies, quasi-RCTs and so-called uncontrolled cohort studies. The most frequent method in Cochrane and Campbell Collaborations, the Centre for Review Dissemination (CRD) at the University of York, UK and the Evidence in Health and Social Care (NHS) databases is the double-blind RCT. In other databases, whose primary focus is not medicine, we observe preferences for other, non-experimental methods. A more detailed overview of a database-specific methodological focus is available in Table A2 in Appendix A2.

Figure 5. Most frequent methods in primary studies that are incorporated in systematic reviews, averaged across various databases.



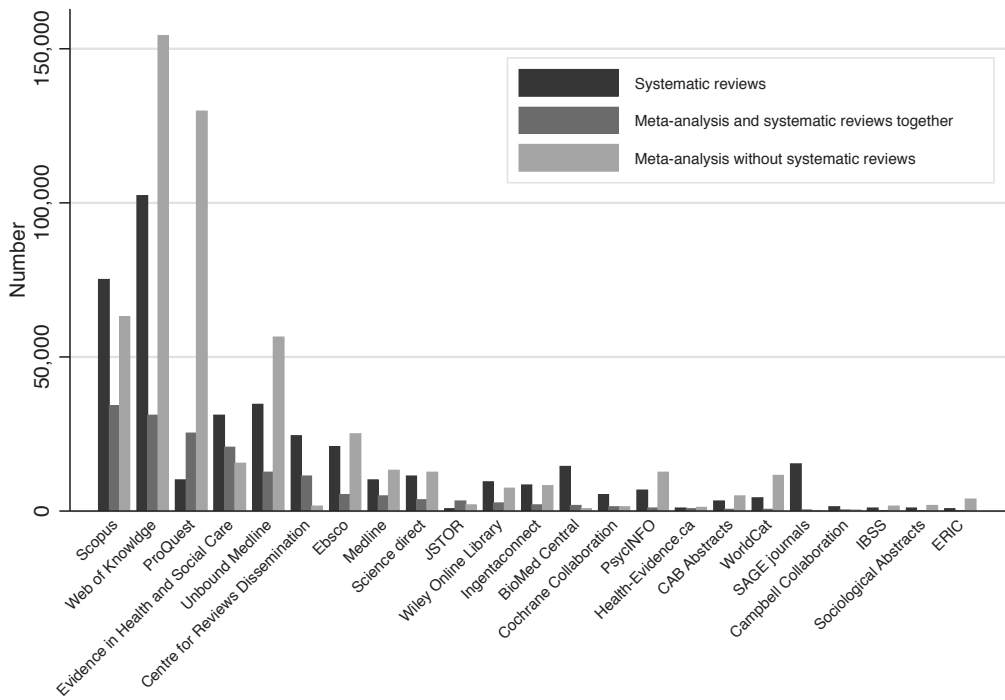
Source: Authors' elaboration.

Currently we see an increasing demand for inclusion of qualitative information in systematic reviews. Some institutions, such as the EPPI-Centre insist on including both quantitative and qualitative original research in their systematic reviews. In as much as the early systematic reviews at the EPPI-Centre addressed the 'what works?' situations and tested the effects of interventions, more recent EPPI-Centre reviews aim to address a broader range of questions, i.e. the acceptability of an intervention, or the factors influencing implementation of interventions, for which qualitative research is indispensable. The focus of the reviews has thus evolved to incorporate the understanding of a specific health issue from the experiences and points of view of people targeted by different interventions (Thomas & Harden, 2008). This is where we see an important contribution from social sciences as several articles instruct about the best way of choosing qualitative data sources, appraising the data quality, assessing the explanatory power of qualitative evidence and combining qualitative with quantitative data (see, e.g., Barnett-Page & Thomas, 2009; Petticrew & Roberts, 2008; Paterson et al., 2001; Noblit & Hare, 1988).

5.3. Methodology of conducting systematic reviews

The methods section is a third crucial component of any systematic review. It provides assurance that the review has been designed rigorously. The content of the methods section could include: (1) a description of the literature search terms, (2) inclusion and exclusion criteria, (3) an assessment of publication bias, (4) quality criteria for each study included, (5) study validity considerations of bias and confounding, and (6) descriptions of ‘weight of evidence’ or research synthesis used in the review. A wide range of methods is available for the analysis of secondary data. These methods, usually termed meta-methods, can be both qualitative and quantitative. Some examples of these methods and their definitions are given in the Glossary (Table A3, Appendix A3). The most common synthesis method is meta-analysis, followed by meta-regression and narrative synthesis, as Figures 6 and 7 illustrate.

Figure 6. Numbers of systematic reviews and meta-analyses across various databases

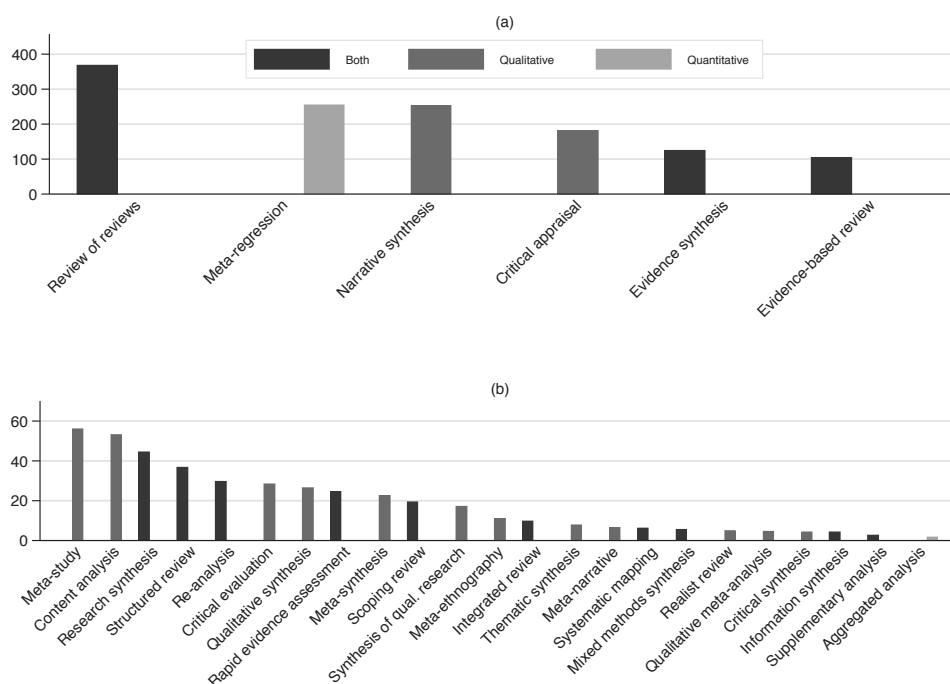


Source: Authors' elaboration.

It is not surprising that the most common synthesis methods are quantitative, as the tradition of systematic reviews rests on summarizing effects of various medical treatments. It appears particularly popular to conduct a systematic review and a meta-analysis at the same time. This is true for 30,926 articles available in Web of Knowledge and 25,253 articles available in ProQuest. Scopus reports that around 35 thousand meta-analyses are incorporated in some 75 thousand systematic reviews, which means that a meta-study is performed in 45% of all systematic reviews. Depending on the focus of a database and the precision of the search process, the share of meta-analyses in systematic reviews differs. MEDLINE shows that 49% of systematic reviews include meta-analysis, while the CRD shows 47%. However, combining systematic reviews and meta-analysis in one project has not decreased the popularity of conducting meta-analyses independently; they are three times more frequent than systematic reviews, as shown in Figure 6. The knowledge from systematic reviews is often integrated further through a 'review of reviews' method (see Appendix A3), which is designed to systematize evidence from already available systematic reviews. Figure 7 shows that meta-regression, narrative synthesis and critical appraisal are also frequently used for analyzing secondary data.

Quantitative systematic reviews and meta-analyses can achieve increased power and precision that come from pooling primary data. Mulrow (1994) has compiled several examples of the advantages of systematic reviews in various quantitative effect assessments. When data from different primary studies are pooled, the sample size and thus the power of estimating the combined effect size is increased. Increasing the statistical power is particularly relevant when assessing small effects or events with low incidence rates. In addition, meta-analysis can provide answers that no single study can, or settle arising controversies. Of equal importance, meta-analysis can quantify the between-study heterogeneity (Lau et al., 1998).

Figure 7. Average number of different research synthesis methods in combination with systematic reviews across various databases



Note: For clearer illustration, we present the less frequent research methods in panel (b). Source: Authors' elaboration.

The appropriateness of meta-analyses depends on the review question and the criteria developed for determining which primary studies should be included, simply because the degree of heterogeneity introduced in the meta-analysis can affect the result. Moreover, the interpretations of the meta-analysis results depend on the way data are synthesized, through weighted average, regression analysis or individual data modeling (Lau et al., 1998).

Warnings that systematic reviews can underestimate the magnitude of evidence in relevant literature have been voiced since the late 1990s. The problem arises when the reviews only include primary studies of a certain 'methodological quality' (Edwards et al., 1998). For example, including only evidence from RCTs and clinical trials may lead to a distorted evidence synthesis if the reviewers automatically, and without reflection, exclude results obtained through other, weaker, study designs that contain equally relevant information. Conversely, including such weak study designs can misleadingly amplify the strength of an estimated effect if the weaker designs have biased effect estimates. It is thus common to trade off the loss of additional perspectives for the improved precision of findings (Booth, 2001). Another approach to minimizing heterogeneity includes assessing the 'message' within each individual piece of research. Edwards et al. (1998) propose to assess both the methodological quality and the weight of its message, rather than rejecting studies that fall below a certain quality threshold.

Narrative synthesis is frequently used to summarize quantitative data in systematic reviews. Different systematic review traditions, either Cochrane, Campbell or peer-reviewed articles, have incorporated this method into their reviews. Moreover, there is a general ‘softening’ towards qualitative, theoretical and interpretative methods and appraisal techniques, which are required when answering ‘messier’ questions. Qualitative data are delivered in the form of narratives, where themes and concepts function as the analytical device (Dixon-Woods et al., 2001, p. 126). By supporting a narrative review, ‘the toolkit of the evidence-based policy movement is expanded, enhanced and enriched’ (Jones, 2004).

We show in Figure 7 which qualitative methods are often incorporated in systematic reviews. Barnett-Page and Thomas (2009) have identified nine distinct approaches to qualitative synthesis: meta-ethnography, grounded theory, thematic synthesis, textual narrative synthesis, meta-study, meta-narrative, critical interpretive synthesis, ecological triangulation and framework synthesis. The first qualitative synthesis methods appeared late in the 1980s (Noblit & Hare, 1988) and they dealt with synthesis of ethnographic research. The method proposed by Noblit & Hare (1988) was termed ‘meta-ethnography’, but it found applications beyond ethnographic studies (Campbell et al., 2003) and extensions in methods termed ‘meta-study’ (Paterson et al., 2001), ‘critical interpretive synthesis’ (Dixon-Woods et al., 2006), ‘meta-synthesis’ (Sandelowski & Barroso, 2002) and ‘thematic synthesis’ (Thomas & Harden, 2008).

Some organizations, like the EPPI-Centre and the UK National Health Service's research promote the inclusion of non-quantitative methods, whereas systematic reviews published by Cochrane Collaboration rarely rely on qualitative data and methods. This indicates certain disagreements over the use of qualitative studies in systematic reviews. These disagreements stem from the nature of qualitative data as it is not clear how this kind of data should be synthesized using the traditional systematic review methodology. One of the problems comes from the need to apply quantitative methods of synthesis that are reductionist to data from a study genre that is intended to be explorative. Further, extracting qualitative data is susceptible to imprecision due to the subjective nature of identifying themes from text descriptions, in contrast to the practice of extracting data from tables in quantitative systematic reviews. But a specific meta-triangulation is possible when the same theme is identified in different studies conducted among different populations. In sum, bringing together qualitative findings requires consistent synthesis methods that preserve the essential context and complexity of qualitative research (Thomas & Harden, 2008).

5.4. Comprehensiveness of the review

The comprehensiveness or the coverage of a review depends on the availability of literature and on how easy it is to retrieve all relevant studies. While some types of questions require literature that can be easily located in electronic databases, other questions may require

higher reliance on non-journal sources, gray literature, reports from government and research institutes, or websites.

Only a small fraction of trial reports is incorporated in up-to-date systematic reviews (Bastian et al., 2010). A typical Cochrane systematic review contains about six trials, in which the median number of participants is 945 per review (Mallett & M. Clarke, 2002). Reviews of social interventions count usually tens or fewer, rather than hundreds of studies (Petticrew & Roberts, 2008). This is due to a relatively weak evidence base characterized by few replication studies and scarce outcome evaluations of social interventions (Oakley, 2002).

The language in which the primary studies are written is highly relevant for inclusion in the review. As there is a well-documented bias towards English language articles (Bronson & Davis, 2011), systematic reviews of good quality should aim at including non-English language publications. This is advisable as reviews that include only studies reported in English may yield biased results and inferences (Grégoire et al., 1995). Even if language bias does not influence estimates, it may affect precision simply because the analysis will be based on fewer data (Moher et al., 2000).

5.5. Questions addressed in systematic reviews

Systematic reviews are generally motivated by the need to answer a pressing question deemed important for either policy or practice. Thus, it is important to look into both the existence of a well-defined question and the type of question asked in a systematic review. Counsell (1997) states that ‘A good systematic review is based on a well-formulated, answerable question. The question guides the review by defining which studies will be included, what the search strategy to identify the relevant primary studies should be, and which data need to be extracted from each study’.

Best practices of conducting systematic reviews propose that the questions be formulated using the PICOS approach, which includes several components: the patient population or the disease being addressed (P), the intervention or exposure (I), the comparator group (C), the outcome or endpoint (O), and the study design chosen (S) (Oxman & G. H. Guyatt, 1993).

5.6. Usage of systematic reviews

Systematic reviews are used much broader than in the purely academic circles. They are very often undertaken to inform decision-making by non-academic users of research such as policymakers and practitioners (Rees & Oliver, 2012). The Cochrane Collaboration has emphasis on user involvement and encourages authors to incorporate the views of the users, such as consumers and clinicians (Higgins & Green, 2011). The Campbell Collaboration (2008) considers that the user group can comprise either i) people who

receive a service, intervention or program, ii) practitioners (*i.e.* social workers, teachers, doctors), iii) policy-makers, iv) researchers or v) funders. The CRD considers their users to be any person or group who might potentially use the findings of a review (CRD, 2009). All systematic reviews produced by CRD have an advisory group, comprising a range of users who provide input at various stages of the review process. Similarly, the EPPI-Centre stresses the importance of user involvement in obtaining a wide range of viewpoints (EPPI-Centre, 2009b).

As different organizations commission and conduct systematic reviews, the questions about reliability of different reviews often arise. For example, industry funded reviews of drug trials appear limited in their value to guide decisions as they are less transparent, have few reservations about limitations and have more favorable conclusions than corresponding Cochrane reviews (Jorgensen et al., 2006). Not only reviews funded by industry, but also reviews published in peer-reviewed journals suffer from serious methodological flaws (Jadad, 2000). Similarly, Cochrane reviews are favored for their greater methodological rigor and more frequent updates than systematic reviews published in paper-based journals (Jadad et al., 1998).

6. Typologies of systematic reviews

It is evident that the systematic review practice experiences various complexities. We have therefore attempted to offer a way in which to organize thinking and practices around systematic reviews, which are recognizable by a question of specific type, comprehensive search and retrieval of the relevant research, explicit inclusion criteria, critical analysis and synthesis of the primary studies, methodological rigor and user involvement. These characteristics of systematic reviews are further analyzed and applied in the context of systematic review types.

Presently there are four different typologies of systematic reviews. The typologies are listed in Table 4. Starting with the Cochrane Collaboration (Cochrane Collaboration, 2012a), we observe a distinction between three types of systematic reviews: intervention reviews, diagnostic test accuracy reviews and methodology reviews. Lavis (2009) focuses on the type of data and evaluation method and defines four review types: reviews of observational studies, reviews of qualitative studies, reviews of effectiveness studies and reviews of economic evaluations. The Dutch Knowledgecenter Measurement Instruments of the VU University Medical Center also focuses on data and methods, but emphasizes the extensiveness of the studies in their typology, *i.e.*, if the object of investigation is a single instrument or all available instruments for measurement performed against a defined or unspecified construct (KMIN, 2012). Finally, based on the review objectives and synthesis methods, Gough et al. (2012) identify two main types of systematic reviews – aggregative and configurative to which they add the in-between type with varying degrees of both aggregation and configuration.

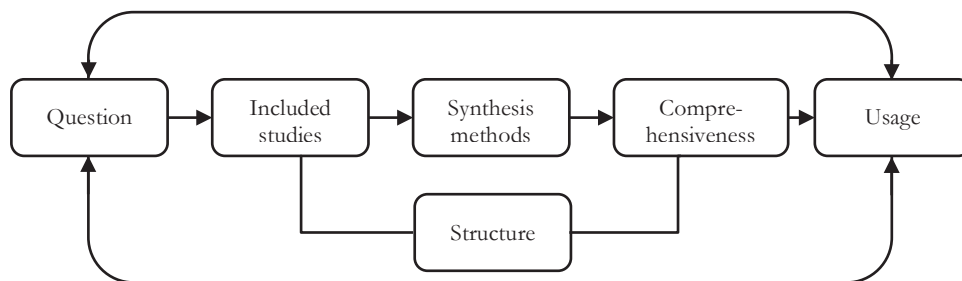
Table 4. Typologies of systematic reviews

Cochrane Systematic Reviews (Cochrane Collaboration, 2012a)	Policymaking systematic reviews (Lavis, 2009)	Systematic reviews of measurement instruments (KMIN, 2012)	Examples of review types from Gough et al. (2012)
1. Intervention reviews assess the benefits and harms of interventions used in healthcare and health policy. 2. Diagnostic test accuracy reviews assess how well a diagnostic test performs in diagnosing and detecting a particular disease. 3. Methodology reviews address issues relevant to how systematic reviews and clinical trials are conducted and reported.	Reviews of 1. Observational studies 2. Qualitative studies 3. Effectiveness studies 4. Economic evaluations	A systematic review of 1. One measurement instrument 2. A selection of measurement instruments that aim to measure a particular construct in a particular population 3. All available measurement instruments that aim to measure a particular construct in a particular population 4. All available measurement instruments (without specifying the construct to be measured) in a particular population	1. Aggregative a) ‘What works?’ reviews b) Diagnostic test c) Cost benefit d) Prevalence 2. Configurative a) Meta-ethnography b) Critical interpretative synthesis c) Meta-narrative review 3. Configuring and aggregative a) Realist synthesis b) Framework synthesis

Source: Author’s elaboration based on the sources cited in the table.

Any categorization of systematic review types should, in our view, include the dimensions of systematic review practice we have mapped out in Section 5. In Figure 8 we illustrate how all characteristics of systematic reviews can be traced to the research question, which informs the structure of the review, included studies, synthesis methods, the overall comprehensiveness of the review and its usage. At the same time, the research question is informed by the usage of the review as it reflects needs and desires of the users, i.e. commissioning bodies.

Figure 8. The relationship between the main criteria for appraising systematic review practice



Source: Authors’ elaboration.

Every review question contains implicit ideological and theoretical assumptions that determine specific choices researchers make during the review process. Depending on the question, a systematic review will include both qualitative and quantitative research, be limited to only experimental evidence, or allow all types of research evidence, even the

non-empirical one. In practical terms, this means that questions determine not only the underlying conceptual framework used to interpret and understand the research evidence, but also the way in which the review is undertaken, including the decision about the primary studies included, the way they are analyzed and synthesized into the final review.

It is often emphasized that the topic of a review should be based on a concise question (Schlosser, 2007). As many types of review products may be taken for systematic reviews, it is important that the true systematic review clearly indicates the question it tries to answer. Gough et al. (2012) have compiled a list of most frequent questions in systematic reviews:

- What is the effect of this intervention?
- What is the accuracy of this diagnostic tool?
- What is the cost of this intervention?
- What is the meaning of a process or a phenomenon?
- What is the effect of this complex intervention?
- What is the effect of this approach to social policy in this context?
- What are the attributes of this intervention or activity?

Starting from this list, we identify three broad groups of questions (Table 5). We term the first group as ‘traditional’ or *focused questions*. Such questions are usually direct and aim to determine ‘what works’ for various interventions, healthcare procedures or drug treatments. We term the second group *complex questions* (in some contexts recognized as ‘messy’) because they go beyond ‘what works’ and aim to address context-sensitive issues such as people’s attitudes and experiences, environmental or practical concerns. Where the focused questions offer straightforward ways of understanding if and how relatively simple interventions work, the complex questions provide more useful insights when investigating various organizational and policy interventions. The complexity comes at a price of increased methodological challenges and reliance on evidence forms that are deemed less desirable by some authors. We term the third group *hybrid questions* because systematic reviews based on such questions aim to tackle both focused and complex issues. From the three broad categories, we identify five specific groups of questions, which form the basis of our typology. The five categories of questions are:

- Effect-driven,
- Explanatory,
- Economic,
- Hermeneutic,
- Mixed.

Effect-driven, explanatory and economic questions are all variations of focused questions. The hermeneutic category is a natural way of thinking about complex questions, as hermeneutics is the study of the interpretation of written texts. Finally, we would argue that the terms ‘hybrid’ and mixed can both be used when the questions are aimed both at determining effect size and describing and interpreting the causal factors. Further,

the term mixed is also used as a description of the approach applying both quantitative and qualitative assessments (*e.g.*, a mixed methods approach). An overview of the most common systematic review questions is given in Table 5 together with the corresponding category. Notice, how each type of research question results in a specific type of systematic review.

As noted, Lavis (2009) classifies systematic reviews based on the methods used to synthesize the research evidence. Further, Gough et al. (2012) state that the synthesis methods reflect many of the ‘approaches, assumptions, and methodological challenges of the primary research that they include’. But Lavis (2009) and Gough et al. (2012) overlook the fact that the choice of synthesis methods can sometimes be conditioned by other factors, such as the type and quality of primary study designs. One could say that using the methods to classify systematic reviews would be appropriate if the best methods to answer a specific question were always possible and always chosen. But in this case categorizing by method would be equivalent to categorizing by question whereby our suggestion for typology would coincide with Lavis’ classification. In the same vein, we consider the classifications based on the research objectives by the Cochrane Collaboration and Gough et al., (2012) to be directly related to a typology using the type of question because we would argue that the research objective is in an expression of the research question.

Table 5. Classification of types and questions commonly addressed by systematic reviews

	Classification of questions and systematic review types	Characteristic questions	Examples of possible synthesis methods
Focused	Effect-driven (impact) question	What is the effect of this treatment or intervention?	Statistical meta-analysis of experimental trials
	Questions concerned with impact of different interventions	What is the difference in the effect of two or more competing interventions?	Statistical meta-analysis of experimental trials
		What is the effect of this complex intervention?	Multi-component mixed methods reviews
		What is the effect of this approach to social policy in this context?	Realist synthesis of evidence of efficacy and relevance across different policy areas
		What is the accuracy of this diagnostic tool?	Meta-analysis of evaluations of diagnostic tests
	Explanatory (causal)	What is the cause of a specific phenomenon?	Statistical meta-analysis
	Questions concerned with why and how events occur	What are the sources of variation between two groups, conditions or states?	Statistical meta-analysis
		How and why the outcomes between groups differ?	Statistical meta-analysis or qualitative meta-synthesis
	Economic	What is the cost of this intervention?	Synthesis of cost-benefit analyses
	Questions about economic impact		

Complex	Hermeneutic (interpretative and/or descriptive) Questions aiming to interpret, describe, summarize and present data, events and observations	What is the meaning of a specific phenomenon? What is the process of a specific action, phenomenon or event? What are the attributes of this intervention or activity? Is this complex intervention appropriate?	Critical interpretative synthesis of ethnographic studies Conceptual synthesis such as meta-ethnography Framework synthesis framed by dimensions explicitly linked to particular perspectives Multi-component mixed methods reviews
Hybrid	Mixed Questions aiming to determine effect and interpret phenomena	What is the impact of an intervention and what are the participants' views and experiences of participating in the intervention?	Multi-component mixed methods reviews such as meta-analysis alongside the thematic narrative synthesis

Source: Authors' elaboration and Gough et al. (2012).

However, compared to the four typologies in Table 4, the typology we propose clearly links the users of the reviews and the research questions. One should not overlook the importance of 'whose questions are being asked' (Gough et al., 2012) because users of the reviews can bring a wide range of different perspectives into the review process. Users contribute to the review production in more than one way: for example, by 'identifying and prioritizing review topics, defining review questions and important outcomes, conducting reviews, editing review protocols and reports, and disseminating and implementing review findings in practice' (Campbell Collaboration, 2008). The EPPI-Centre argues for considering users' perspectives in order 'to make a considered decision about the question that the review is attempting to answer' (EPPI-Centre, 2009b) and users are appointed to comment on the external validity of reviews (Higgins & Green, 2011).

The typology of the systematic reviews we propose may prove useful in dealing with complexities caused by different types of review products that are conducted with different user and purposes in mind. It can also serve as a tool for guiding the review process and identifying the most suitable evidence synthesis methods.

7. Systematic reviews in international development

Policy briefs – especially summing up different donor-led impact evaluations – can be considered as predecessors of systematic reviews in international development. Several development research groups and organizations, notably the J-PAL, Innovations for Poverty Action (IPA) and International Initiative for Impact Evaluation (3ie) strongly promote evidence from studies with rigorous protocols (White & Waddington, 2012). This demand for rigorous evidence has led to an increased interest in systematic reviews within international development since the mid-2000s. A recent estimate indicates that at least 100 new reviews, covering topics of international development, are on-going or completed, many of which are funded by donors (White & Waddington, 2012).

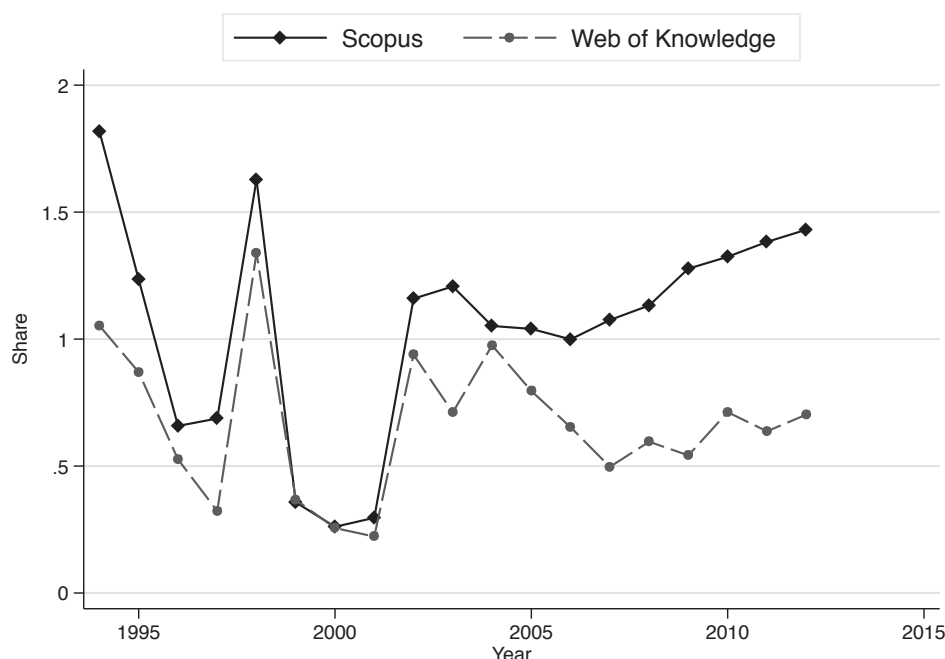
Experiences from conducting systematic reviews in international development were described in a recent special issue of the *Journal of Development Effectiveness*, emphasizing the importance of systematic reviews for better-informed development policies, as well as the conceptual and methodological challenges specific for international development. This issue includes a ‘good practice guide’ on how to do systematic reviews in international development (Waddington et al., 2012), and special guidance on how to synthesize qualitative information (Snilstveit, 2012; Snilstveit et al., 2012). The authors also explain how to assess the risk of bias in development reviews through *e.g.* methodological pragmatism (Duvendack et al., 2012; Stewart et al., 2012).

Even though the interest in systematic reviews in international development is high, it presents only a small fraction of the total number of reviews. Figure 9 shows that development reviews comprise 1% of the total number of systematic reviews based on the information in the databases Scopus and Web of Knowledge (The total number of systematic reviews is shown in Figure 1). Our interest was identifying as many systematic reviews in development research as possible using several keywords related to development assistance. We show the prevalence of relevant terms in Figure 10. In addition to specifying that a review concerns developing countries, the most common is to use ‘international development’ and ‘development work’. Similar numbers of systematic reviews have been conducted in Africa and Asia; there are only 10% more reviews in Africa than in Asia, but 50% more in Africa than in Latin America. In contrast, there are only few systematic reviews in transition economies.

There are two approaches to systematic reviews in international development; the early reviews were conducted without much donor participation while later reviews are increasingly conducted with donors’ interests in mind. Interestingly, donor involvement coincides with the topics addressed in the reviews. While the early reviews almost exclusively assessed health-related issues, the later reviews are more within the social sciences, assessing the impact of development intervention programs such as credit

provision, anti-corruption policies and infrastructure investments, which, as a norm, depart from easily measurable indicators and need to account for complex social interactions.

Figure 9. Systematic reviews in international development as a share of total number of systematic reviews in databases Scopus and Web of Knowledge

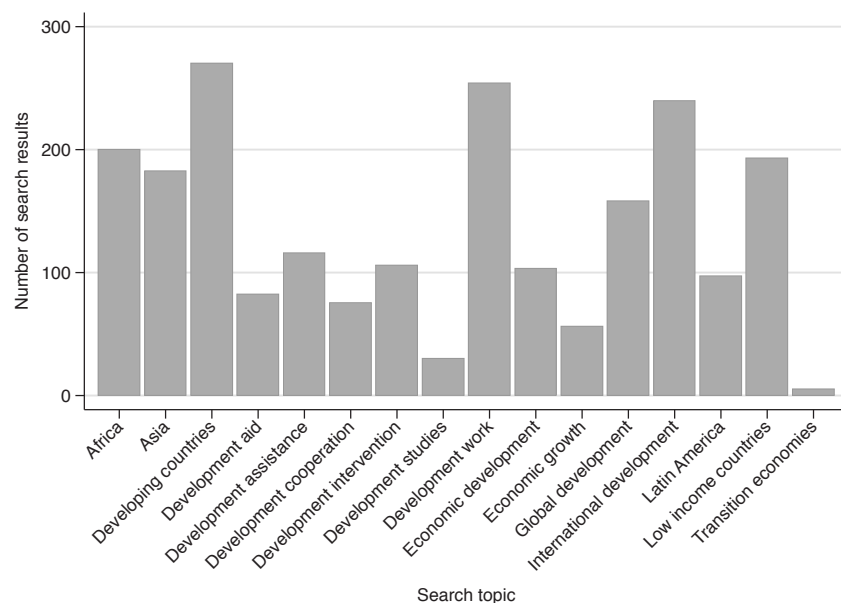


Source: Authors' elaboration.

The 3ie commissions finding and using evidence on what works, when, why and for how much. Moreover, the 3ie is in charge of peer reviewing systematic reviews in international development (3ie, 2012a). Further, some of the first systematic reviews performed by the EPPI-Centre were funded by DFID and in 2007, the Alliance for Health Policy and Systems Research awarded grants to four institutions in low- and middle-income countries to establish centers for systematic review of health policy and systems research, which are supported by the EPPI-Centre (EPPI-Centre, 2009c).

Systematic reviews commissioned by development agencies are different from traditional, Cochrane and Campbell systematic reviews in several ways. Most of the 'developmental reviews' do not state in the title that the material is in fact a systematic review, despite the explicit request from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to indicate in the title that the review is systematic (PRISMA Statement, 2009). Also, systematic reviews in international development often serve as scoping reviews, in the sense that they state in the objectives section that they aim to establish how much methodologically reliable and comprehensive literature exists for a specific topic. A fraction of reviews also aims to identify gaps in research and to recommend further research paths.

Figure 10. Number of systematic reviews in the international development context obtained by searching for a range of keywords and averaged across various databases



Source: Authors' elaboration.

The topics in international development cover a range of issues, such as: health and nutrition, education, social protection and social inclusion, governance and fragile states, environment, infrastructure and technology, agriculture and rural development, aid delivery and effectiveness, economic development and gender (DFID, 2013). The earliest record of a systematic review in 3ie's database is from 2002 and the frequencies of different topics are presented in Table 6 and Figure 11.

Clearly, systematic reviews in development research deal mostly with health-related topics, as this is where it all started (see Figure 11). The second most frequent group of issues includes different forms of human development and economic interventions, such as interventions in schooling, employment and finance. Several systematic reviews are multidisciplinary as they analyze several types of interventions at the same time. In such reviews, health-related topics are combined with economic or social interventions. The topics of systematic reviews dealing with developing countries simply reflect the areas of specialization of individual organizations conducting the reviews. For example, the EPPI-Centre systematic reviews related to developing countries deal mostly with various aspects of public and social policy, such as education.

Table 6. Systematic review topics in international development in EPPI-Centre publications

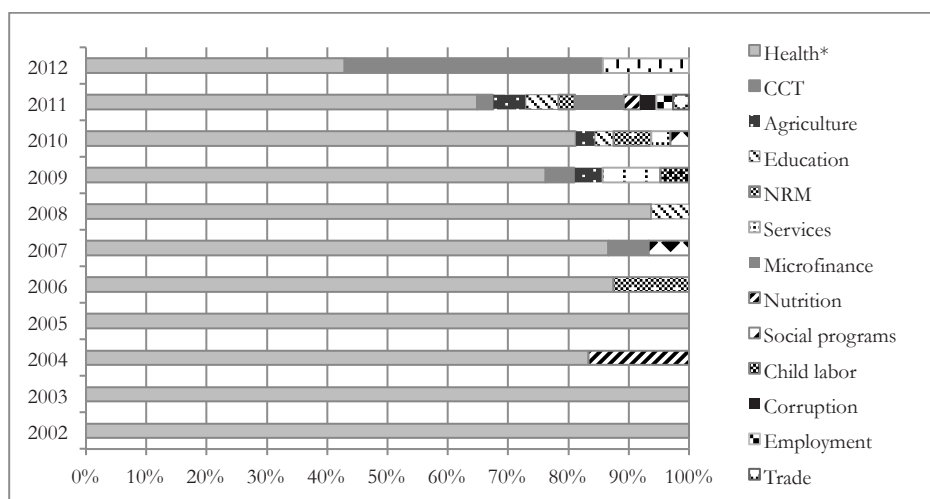
Topic	Number	Percent
Education	79	56.03
Health	29	20.57
Health services	9	6.38
Public services	4	2.84
Micro-finance	3	2.13
HIV/AIDS	2	1.42
Criminal justice	2	1.42
Economic growth	2	1.42
Poverty	2	1.42
Nutrition	2	1.42
Social interventions	2	1.42
Conditional cash transfers	1	0.71
RCTs	1	0.71
Childhood interventions	1	0.71
Tariff reductions	1	0.71
Trade	1	0.71
Total	141	100

Source: Authors' elaboration.

The 3ie's database shows that it was not until 2006 that non-health systematic reviews started appearing. These reviews investigated interventions related to natural resources, social programs and conditional cash transfers. The real diversification of topics started in 2009, when the interest spread to child labor, service provision and agriculture. Non-health topics have remained relevant until present but still do not surpass the amount of health research.

The typology set out in Section 6 implies that evidence synthesis can be achieved in a focused, complex or a hybrid way. Examples of question categories within international development are illustrated in panel (a) in Figure 12. Around 85% of the reviews are motivated by focused questions as they attempt to evaluate effects or the impact of various interventions. Usually, the impact of a given intervention is assessed over a range of variables, with attention paid to both direct and indirect effects. Much less common are complex (10%) and hybrid (5%) questions that are formulated more broadly and attempt to determine whether interventions work for different groups identified by race, ethnic origin, occupation, education, gender or socioeconomic status.

Figure 11. Main topics in international development reviews in 3ie publications over time

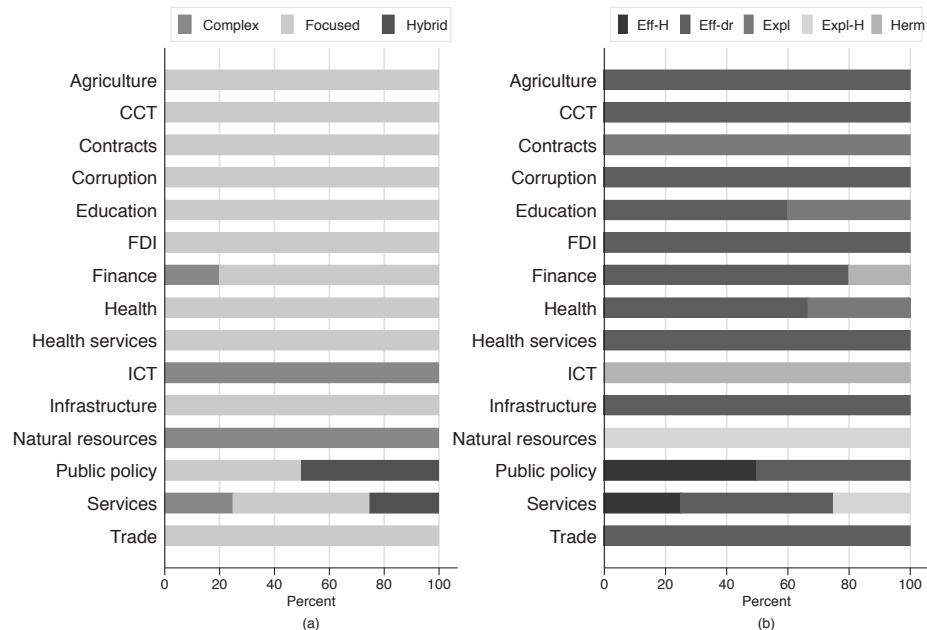


Note: *For clearer presentation, the Health category contains research on health, health services and HIV/AIDS (taking up 32%, 30% and 38% of all health research, respectively); CCT is “conditional cash transfer” and NRM is “natural resource management”. The topics are sorted in a descending order of frequency. Source: Authors’ elaboration.

In terms of the main question subcategories, international development reviews express greatest interest in answering the effect-driven questions, which are present in 74% of reviews. This is visible in panel (b) in Figure 12, which also shows that explanatory and hermeneutic questions are considerably less common (11% and 5% respectively), as are the combinations of effect-driven and explanatory questions with hermeneutic questions (both appearing in 5% of cases). This simple analysis of questions addressed in development reviews illustrates that although operating in a multidisciplinary environment, which requires acknowledging the influence of institutions and social interaction, the systematic reviews tend to focus on systematizing the ‘easy-to-measure’ knowledge.

Most of the reviews in international development, however, struggle with finding acceptable evidence. All the studies that enter the analysis need to be evaluated, which leaves authors either devising the appraisal criteria themselves or using some of the standardized checklists. For example, 3ie recommends that systematic reviews be assessed based on a checklist, which is an adapted version of the SURE Guide designed by Supporting the Use of Research Evidence (SURE) Collaboration (3ie, 2013; WHO, 2013). The checklist is focused on three main areas: methods used to identify, include and critically appraise studies; methods used to analyze the findings; and the overall assessment of the reliability of the review.

Figure 12. Categories of research questions and topics occurring in systematic reviews on international development in DFID's database



Note: Panel (a) shows topics of systematic reviews and the main categories of questions that are identified for each topic. Panel (b) shows the main topics of reviews and subcategories of questions. 'Eff-H' is effect-driven and hermeneutic question; 'Eff-dr' is effect-driven question; 'Expl' is explanatory question; 'Expl-H' is explanatory and hermeneutic question and finally, 'Herm' is hermeneutic question. Source: Authors' elaboration.

A lack of studies of appropriate quality has consequences for the choice of synthesis methods. Looking into DFID's systematic reviews database, we have found that narrative synthesis is slightly more common than meta-analysis (30% and 25% of reviews, respectively). We believe that narrative synthesis is more prevalent because of the data scarcity and incompatibility of different outcomes measures. Systematic mapping and realist review approaches appear sparsely, as well as the combinations of narrative synthesis with other methods, such as meta-analysis, vote counting or causal chain analysis.

The large number of indicators used for impact assessment of just a single development program poses problems for systematic review practice. The wide range of conceptually similar indicators that can be measured in different ways means that the choice of indicators matters greatly for comparability of primary studies. For example, in assessing the effectiveness of an intervention on poverty reduction, researchers find it difficult to compare outcomes between studies because different poverty measures that have been used in primary studies: poverty indices, income and expenditure indicators provide different evidence. This is a special problem and it is distinct from the many evidence-based healthcare projects, which tend to focus on variables that can be easily measured. Therefore, it is not uncommon to discover that a systematic review within development cannot make conclusions and recommendations. This issue highlights the importance of having standardized data available and harmonizing the way in which the impact evaluation studies are performed.

8. Conclusion

While the boundaries between different scientific fields have become increasingly fuzzy, the relevance of systematic reviews remains strong, not the least because they reinforce the debate about how we make discoveries and learn about the world. In analyzing several definitions and characteristics of systematic reviews, we conclude that the best approach to research synthesis that can be understood as a systematic review is the one containing a predefined research protocol, thorough literature search, criteria for inclusion of primary studies, and finally, both analysis and synthesis of primary data. We have shown several patterns in systematic reviewing practice. Although systematic reviews have a quantitative tradition, we are witnessing an increased recognition of non-quantitative types of evidence in systematic reviews. This is particularly true for more recent systematic reviews in international development that deal with diverse intervention programs and a plethora of outcome measures while attempting to answer complex research questions.

The question addressed in a systematic review is the most important segment of the review practice as it both reflects the interests of a wide range of users and also influences the way in which the systematic review is operationalized. It affects the inclusion criteria for the primary studies, the synthesis methods employed, the comprehensiveness of the review and the key steps of the review process. Based on the characteristics of the question posed by a review, we have categorized questions first into three broad groups: focused, complex and hybrid; and subsequently into five narrower subgroups. Using these five types of questions, we propose classifying systematic reviews into the five categories: effect-driven, explanatory, economic, hermeneutic and mixed.

We apply the typology to systematic reviews in international development. We observe a tendency to investigate focused questions, which as a rule tackle the more easily measurable intervention effects. But there is an increasing interest in also looking at complex questions concerning a multitude of aspects of specific donor interventions. Thus, as development reviews are moving away from the initial form they should not aim at pursuing the classical review approach suitable for traditional, ‘easy-to-measure’, focused questions that rest on experimental designs and concentrate on determining the effectiveness of healthcare. Instead, international development systematic reviews should adjust the review process to cater for the type of question they are addressing. In this way, the differences in type and quality of the included primary studies, methodological approach and the study comprehensiveness will not be a source of bias, but will add to the overall success of the review.

Empirically driven, traditional systematic reviews (e.g. Cochrane or Campbell-style reviews) are suitable for assessing focused interventions such as drug trials, but perform worse when it comes to assessing complex social interventions. When the purpose is to simultaneously

analyze outcomes from multiple studies, it becomes almost impossible to account for the effects of culture, community history, geo-political contexts, study design and theories, which characterize complex social interventions. A step toward correcting for these difficulties could be a wider use of realist reviews that do not ask ‘Does it work or not?’ but rather, ‘What works, for whom, and in what circumstances?’ as proposed by Pawson et al. (2005). Understanding *why* a specific intervention has or has not worked is of equal relevance for policy as whether or not an intervention has worked.

References

- 3ie. (2012a). Quality Assurance Services. *3ie: International Initiative for Impact Evaluation*. Retrieved February 24, 2013, from <http://www.3ieimpact.org/en/engage/quality-assurance-services/>
- 3ie. (2012b). *3ie Impact Evaluation Glossary*. New Delhi: International Initiative for Impact Evaluation.
- 3ie. (2013). Systematic Reviews. *3ie: International Initiative for Impact Evaluation*. Retrieved February 10, 2013, from <http://www.3ieimpact.org/en/evidence/systematic-reviews/>
- Ashrafian, H., Darzi, A. & Athanasiou, T. (2011). Evidence Synthesis: Evolving Methodologies to Optimise Patient Care and Enhance Policy Decisions, in: Athanasiou, T. and Darzi, A. (Eds.), *Evidence Synthesis In Healthcare: A Practical Handbook for Clinicians*, (pp. 1–46). London: Springer.
- Atkins, D., Eccles, M., Flottorp, S., Guyatt, G. H., Henry, D., Hill, S., et al. (2004). Systems for Grading the Quality of Evidence and the Strength of Recommendations I: Critical Appraisal of Existing Approaches The GRADE Working Group. *BMC Health Services Research*, 4(1), 38.
- Baird, S. Garfein, R., McIntosh, C., Özler, B., (2012). Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet*, 379(9823), pp.1320–1329.
- Barnett-Page, E. & Thomas, J. (2009). Methods for the Synthesis of Qualitative Research: a Critical Review. *BMC Medical Research Methodology*, 9(1), 59.
- Bastian, H., Glasziou, P. & Chalmers, I. (2010). Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? *PLoS Med*, 7(9).
- Bates, S., Clapton, J. & Coren, E. (2007). Systematic Maps to Support the Evidence Base in Social Care. *Evidence & Policy: A Journal of Research, Debate & Practice*, 3(4), 539–551.
- Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, IL: Free Press.
- Birge, R. T. (1932). The Calculation of Errors by the Method of Least Squares. *Physical Review*, 40(2), 207–227.
- BMJ. (2013). Practice. *The British Medical Journal*. Retrieved February 14, 2013, from <http://www.bmj.com/about-bmj/resources-authors/article-types/practice>
- Boeije, H. R., Van Wesel, F. & Alisic, E. (2011). Making a Difference: Towards a Method for Weighing the Evidence in a Qualitative Synthesis. *Journal of Evaluation in Clinical Practice*, 17(4), 657–663.
- Booth, A. (2001). Cochrane or Cock-eyed? How Should We Conduct Systematic Reviews of Qualitative Research? In *Qualitative Evidence-based Practice Conference, Taking a Critical Stance*. May 14-16 2001. Coventry: Coventry University.
- Braun, V. & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Briner, R. B. & Denyer, D. (2012). Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool, in: Rousseau, D. M. (Ed.), *The Oxford Handbook of Evidence-Based Management*. Oxford: Oxford University Press.
- Bronson, D. E. & Davis, T. S. (2011). *Finding and Evaluating Evidence: Systematic Reviews and Evidence-Based Practice: Systematic Reviews and Evidence-Based Practice*. Oxford: Oxford University Press.

- Broome, M. E. (1993). Integrative Literature Reviews for the Development of Concepts, in: Rodgers, B. L. and Knafl, K. A. (Eds.), *Concept development in nursing: foundations, techniques, and applications*. Philadelphia, PA: Saunders Co.
- Brunt, L. (2001). The Advent of the Sample Survey in the Social Sciences. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 50(2), 179–189.
- Campbell Collaboration. (2008). *User Involvement in the Systematic Review Process*. Oslo: The Campbell Collaboration. Retrieved February 27, 2013, from http://www.campbellcollaboration.org/resources/research/user_involvement_in_the_systematic_review_process.php
- Campbell Collaboration. (2012). What Is a Systematic Review? *The Campbell Collaboration*. Retrieved December 12, 2012, from http://www.campbellcollaboration.org/what_is_a_systematic_review/index.php
- Campbell Collaboration. (2013). Background: The Campbell Collaboration. *The Campbell Collaboration*. Retrieved February 23, 2013, from <http://www.campbellcollaboration.org/background/index.php>
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M., et al. (2003). Evaluating Meta-ethnography: a Synthesis of Qualitative Research on Lay Experiences of Diabetes and Diabetes Care. *Social Science & Medicine*, 56(4), 671–684.
- CEBM. (2013). Levels of Evidence. *Oxford Centre for Evidence-based Medicine*. Retrieved February 4, 2013, from <http://www.cebm.net/index.aspx?o=1025>
- Chalmers, I, Hetherington, J., Newdick, M., Mutch, L., Grant, A., Enkin, M., et al. (1986). The Oxford Database of Perinatal Trials: Developing a Register of Published Reports of Controlled Trials. *Controlled clinical trials*, 7(4), 306–324.
- Chalmers, I, Hedges, L. V. & Cooper, H. (2002). A Brief History of Research Synthesis. *Evaluation & the health professions*, 25(1), 12–37.
- Chambers, R. (1998). *Clinical Effectiveness Made Easy: First Thoughts on Clinical Governance*. Oxford: Radcliffe Med. Press.
- Clapton, J., Rutter, D. & Sharif, N. (2009). *SCIE Systematic Mapping Guidance*. Social Care Institute for Excellence.
- Cochrane, A. L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.
- Cochrane, A. L. (1979). 1931-1971: A Critical Review, with Particular Reference to the Medical Profession, in: Teeling-Smith, G. and Wells, N. (Eds.), *Medicines for the year 2000*, (pp. 1–11). London: Office of Health Economics.
- Cochrane Collaboration. (2012a). Cochrane Reviews. *The Cochrane Collaboration*. Retrieved December 12, 2012, from <http://www.cochrane.org/cochrane-reviews>
- Cochrane Collaboration. (2012b). History. *The Cochrane Collaboration*. Retrieved February 25, 2013, from <http://www.cochrane.org/about-us/history>
- Cochrane Collaboration. (2013a). About The Cochrane Library. *The Cochrane Library*. Retrieved February 23, 2013, from <http://www.thecochranelibrary.com/view/0/AboutTheCochraneLibrary.html>
- Cochrane Collaboration. (2013b). Glossary. *The Cochrane Collaboration*. Retrieved February 26, 2013, from <http://www.cochrane.org/glossary/5#term371>
- Cook, D J, Sackett, D. L. & Spitzer, W. O. (1995). Methodologic Guidelines for Systematic Reviews of Randomized Control Trials in Health Care from the Potsdam Consultation on Meta-Analysis. *Journal of clinical epidemiology*, 48(1), 167–171.
- Cooper, H. M. & Hedges, L. V. (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

- Counsell, C. (1997). Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Annals of internal medicine*, 127(5), 380–387.
- CRD. (2009). *Systematic Reviews. CRD's Guidance for Undertaking Reviews in Health Care*. York: Centre for Reviews and Dissemination, University of York.
- Davies, P. (2000). The Relevance of Systematic Reviews to Educational Policy and Practice. *Oxford Review of Education*, 26(3-4), 365–378.
- DFID. (2012). Systematic Reviews in International Development : An Initiative to Strengthen Evidence-Informed Policy Making. *Department for International Development*. Retrieved February 14, 2013, from <http://www.dfid.gov.uk/What-we-do/Research-and-evidence/case-studies/research-case-studies/2011/Systematic-Reviews-Background/>
- DFID. (2013). Systematic Reviews. *Department for International Development*. Retrieved February 7, 2013, from <http://www.dfid.gov.uk/r4d/systematicreviews.aspx>
- Dixon, E., Hameed, M., Sutherland, F., Cook, Deborah J. & Doig, C. (2005). Evaluating Meta-analyses in the General Surgical Literature. *Annals of Surgery*, 241(3), 450–459.
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., et al. (2006). Conducting a Critical Interpretive Synthesis of the Literature on Access to Healthcare by Vulnerable Groups. *BMC Medical Research Methodology*, 6(1), 35.
- Downs, S. H. & Black, N. (1998). The Feasibility of Creating a Checklist for the Assessment of the Methodological Quality Both of Randomised and Non-randomised Studies of Health Care Interventions. *Journal of epidemiology and community health*, 52(6), 377–384.
- Dufo, E., Glennerster, R. & Kremer, M. (2008). Using Randomization in Development Economics Research: A Toolkit, in Schultz, T. P. and Strauss, J. (Eds.) *Handbook of Development Economics*, (pp. 3895-3962). Amsterdam: Elsevier.
- Dunn, P. M. (1997). James Lind (1716-94) of Edinburgh and the Treatment of Scurvy. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 76(1), F64–F65.
- Duvendack, M., Hombrados, J. G., Palmer-Jones, R. & Waddington, H. (2012). Assessing “what Works” in International Development: Meta-analysis for Sophisticated Dummies. *Journal of Development Effectiveness*, 4(3), 456–471.
- Edwards, A. G., Russell, I. T. & Stott, N. C. (1998). Signal Versus Noise in the Evidence Base for Medicine: An Alternative to Hierarchies of Evidence? *Family practice*, 15(4), 319–322.
- Egger, M. & Smith, G. D. (1998). Meta-analysis Bias in Location and Selection of Studies. *British Medical Journal*, 316(7124), 61–66.
- EPPI-Centre. (2009a). What Is a Systematic Review? *EPPI-Centre*. Retrieved February 14, 2013, from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=67>
- EPPI-Centre. (2009b). User Involvement. *EPPI-Centre*. Retrieved February 27, 2013, from <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=169>
- EPPI-Centre. (2009c). International Development Review Group. *EPPI-Centre*. Retrieved February 24, 2013, from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2958>
- EPPI-Centre. (2009d). History of Systematic Reviews. *EPPI-Centre*. Retrieved February 25, 2013, from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=68>
- Evans, D. & Pearson, A. (2001). Systematic Reviews: Gatekeepers of Nursing Knowledge. *Journal of Clinical Nursing*, 10(5), 593–599.
- Finfgeld, D. L. (2003). Metasynthesis: The State of the Art—So Far. *Qualitative Health Research*, 13(7), 893–904.
- Fisher, R. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.
- Freeman, L. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35–41.

- Garg, A. X., Hackam, D. & Tonelli, M. (2008). Systematic Review and Meta-analysis: When One Study Is Just Not Enough. *Clinical Journal of the American Society of Nephrology*, 3(1), 253–260.
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3–8.
- Goldschmidt, P. G. (1986). Information Synthesis: a Practical Guide. *Health Services Research*, 21(2 Pt 1), 215–237.
- Gough, D., Thomas, J. & Oliver, S. (2012). Clarifying Differences Between Review Designs and Methods. *Systematic reviews*, 1(1), 28.
- Greenhalgh, Trisha, Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. & Peacock, R. (2005). Storylines of Research in Diffusion of Innovation: a Meta-narrative Approach to Systematic Review. *Social science & medicine* (1982), 61(2), 417–430.
- Grégoire, G., Derderian, F. & Le Lorier, J. (1995). Selecting the Language of the Publications Included in a Meta-analysis: Is There a Tower of Babel Bias? *Journal of clinical epidemiology*, 48(1), 159–163.
- Guyatt, G. (1991). Evidence-based Medicine. *ACP J Club*, A16–114.
- Hampton, J. R. (1998). The End of Medical History? *Journal of the Royal College of Physicians of London*, 32(4), 366–375.
- Harden, A. (2008). Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. *BMC Medical Research Methodology*, 8(1), 45.
- Heaton, J. (2004). *Reworking Qualitative Data*. SAGE.
- Heyvaert, M., Maes, B. & Onghena, P. (2013). Mixed Methods Research Synthesis: Definition, Framework, and Potential. *Quality & Quantity*, 47(2), 659–676.
- Higgins, J. P. & Green, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration. Retrieved February 4, 2013, from <http://handbook.cochrane.org/>
- Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley Pub. Co.
- Jadad, A. R. (2000). Systematic Reviews and Meta-analyses on Treatment of Asthma: Critical Evaluation. *British Medical Journal*, 320(7234), 537–540.
- Jadad, A. R., Cook, D J, Jones, A., Klassen, T P, Tugwell, P, Moher, M., et al. (1998). Methodology and Reports of Systematic Reviews and Meta-analyses: a Comparison of Cochrane Reviews with Articles Published in Paper-based Journals. *JAMA: the Journal of the American Medical Association*, 280(3), 278–280.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Controlled clinical trials*, 17(1), 1–12.
- Jones, K. (2004). Mission Drift in Qualitative Research, or Moving Toward a Systematic Review of Qualitative Studies, Moving Back to a More Systematic Narrative Review. *Qualitative Report*, 9(1), 95–112.
- Jorgensen, A. W., Hilden, J. & Gotzsche, P. C. (2006). Cochrane Reviews Compared with Industry Supported Meta-analyses and Other Meta-analyses of the Same Drugs: Systematic Review. *BMJ*, 333(7572), 782–0.
- Kavanagh, J., Campbell, F., Harden, A. & Thomas, J. (2011). Mixed Methods Synthesis: A Worked Example, in: Hannes, K. and Lockwood, C. (Eds.), *Synthesizing Qualitative Research: Choosing the Right Approach*, (pp. 113–136). Chichester, UK: John Wiley & Sons, Ltd.
<http://onlinelibrary.wiley.com/doi/10.1002/9781119959847.ch6/summary>

- Kelly, K. D., Travers, A., Dorgan, M., Slater, L. & Rowe, B. H. (2001). Evaluating the Quality of Systematic Reviews in the Emergency Medicine Literature. *Annals of Emergency Medicine*, 38(5), 518–526.
- Kitchenham, B. & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Keele University and Durham University Joint Report.
- KMIN. (2012). Systematic Reviews of Measurement Instruments. *Knowledgecenter measurement instruments*. Retrieved February 15, 2013, from <http://kmin-vumc.nl/index.php?change=1&id=16&lang=en>
- Krippendorff, K. (2003). *Content Analysis: An Introduction to Its Methodology*. London: SAGE Publications Ltd.
- Kung, J., Chiappelli, F., Cajulis, O. O., Avezova, R., Kossan, G., Chew, L., et al. (2010). From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *The Open Dentistry Journal*, 4, 84–91.
- Lau, J., Ioannidis, J. P. & Schmid, C. H. (1998). Summing up Evidence: One Answer Is Not Always Enough. *Lancet*, 351(9096), 123–127.
- Lavis, J. N. (2009). How Can We Support the Use of Systematic Reviews in Policymaking? *PLoS Med*, 6(11), e1000141.
- Lind, J. (1753). *A Treatise of the Scurvy. In Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of That Disease. Together with a Critical and Chronological View of What Has Been Published on the Subject*. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson. Retrieved February 14, 2013, from http://www.jameslindlibrary.org/illustrating/records/a-treatise-of-the-scurvy-in-three-parts-containing-an-inquiry/title_pages
- Mallett, S. & Clarke, M. (2002). The Typical Cochrane Review. How Many Trials? How Many Participants? *International Journal of Technology Assessment in Health Care*, 18(04), 820–823.
- Marmot, M., Adelstein, A. & Bulusu, L. (1984). Lessons from the Study of Immigrant Mortality. *Lancet*, 1(8392), 1455–1457.
- Mathison, S. (2005). *Encyclopedia of Evaluation*. Thousand Oaks, CA: SAGE Publications, Inc.
- Moher, D., Pham, B., Klassen, Terry P, Schulz, K. F., Berlin, J. A., Jadad, A. R., et al. (2000). What Contributions Do Languages Other Than English Make on the Results of Meta-analyses? *Journal of Clinical Epidemiology*, 53(9), 964–972.
- Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M. & Altman, D. G. (2007). Epidemiology and Reporting Characteristics of Systematic Reviews. *PLoS Medicine*, 4(3), e78.
- Mulrow, C. D. (1994). Rationale for Systematic Reviews. *BMJ (Clinical research ed.)*, 309(6954), 597–599.
- NHS. (2011). Glossary. *National Institute for Health and Clinical Excellence*. Retrieved February 26, 2013, from <http://www.nice.org.uk/website/glossary/glossary.jsp?alpha=D>
- Noblit, G. W. & Hare, R. D. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: SAGE Publications, Inc.
- Oakley, A. (2002). Social Science and Evidence-based Everything: The Case of Education. *Educational Review*, 54(3), 277–286.
- Oxman, A. D. & Guyatt, G. H. (1991). Validation of an Index of the Quality of Review Articles. *Journal of clinical epidemiology*, 44(11), 1271–1278.
- Oxman, A. D. & Guyatt, G. H. (1993). The Science of Reviewing Research. *Annals of the New York Academy of Sciences*, 703(1), 125–134.

- Paterson, B. L., Canam, C., Jillings, C. & Thorne, S. E. (2001). *Meta-Study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis*. Thousand Oaks, CA: SAGE Publications Ltd.
- Pawson, R., Greenhalgh, T., Harvey, G. & Walshe, K. (2004). Realist Synthesis - an Introduction. Manchester: ESRC Research Methods Programme. Retrieved January 25, 2013, from <http://discovery.ucl.ac.uk/180102/>
- Pawson, R., Greenhalgh, Trisha, Harvey, Gill & Walshe, Kieran. (2005). Realist Review-a New Method of Systematic Review Designed for Complex Policy Interventions. *Journal of health services research & policy*, 10 Suppl 1, 21–34.
- Peters, C. C. (1933). Summary of the Penn State Experiments on the Influence of Instruction in Character Education. *Journal of Educational Sociology*, 7(4), 269–272.
- Petticrew, M. & Roberts, H. (2008). *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell Publishing.
- Pieper, D., Buechter, R., Jerinic, P. & Eikermann, M. (2012). Overviews of Reviews Often Have Limited Rigor: a Systematic Review. *Journal of Clinical Epidemiology*, 65(12), 1267–1273.
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., et al. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. Lancaster: Lancaster University.
- PRISMA Statement. (2009). The PRISMA Statement. *PRISMA - Transparent Reporting of Systematic Reviews and Meta-Analyses*. Retrieved February 15, 2013, from <http://www.prisma-statement.org/statement.htm>
- Rees, R. & Oliver, S. (2012). Stakeholder Perspectives and Participation in Reviews, in: Gough, D., Oliver, S., and Thomas, J. (Eds.), *An Introduction to Systematic Reviews*, (pp. 17–34). London: SAGE Publications Ltd.
- Robinson, J. P. & Shaver, P. R. (1973). Measures of Social Psychological Attitudes. Revised. Ann Arbor, MI: University of Michigan, pp. 750.
- Rosoff, J. I. (1975). Is Support of Abortion Political Suicide? *Family Planning Perspectives*, 7(1), 13.
- Russell, R., Chung, M., Balk, E. M., Atkinson, S., Giovannucci, E. L., Ip, S., et al. (2009). Issues and Challenges in Conducting Systematic Reviews to Support Development of Nutrient Reference Values: Workshop Summary: Nutrition Research Series, Vol. 2. Rockville (MD): Agency for Healthcare Research and Quality (US).
- Sackett, D., Rosenberg, W. M., Gray, J. A., Haynes, R. B. & Richardson, W. S. (1996). Evidence Based Medicine: What It Is and What It Isn't. *BMJ (Clinical research ed.)*, 312(7023), 71–72.
- Sandelowski, M. & Barroso, J. (2002). Finding the Findings in Qualitative Studies. *Journal of Nursing Scholarship*, 34(3), 213–219.
- Schlosser, D. (2007). Appraising the Quality of Systematic Reviews. *Focus Technical Brief*, 17, 1–8.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., Hirsh, H. R., Sackett, P. R., Schmitt, N., et al. (1985). Forty Questions About Validity Generalization and Meta-Analysis. *Personnel Psychology*, 38(4), 697–798.
- Schnore, L. F. & Fagin, H. (1967). *Urban Research and Policy Planning*. Beverly Hills: Sage Publications.
- Schreiber, R., Crooks, D. & Stern, P. N. (1997). Qualitative Meta-analysis, in: Morse, J. M. (Ed.), *Completing a Qualitative Project: Details and Dialogue*, (pp. 311–326). Thousand Oaks, CA: Sage Publications, Inc.
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., et al. (2007). Development of AMSTAR: a Measurement Tool to Assess the

- Methodological Quality of Systematic Reviews. *BMC Medical Research Methodology*, 7(1), 10.
- Shen, L. (1981). The Taiwan issue in Peking's foreign relations in the 1970s: a systematic review. *Chinese Yearbook of International Law and Affairs*, 1, 74–96.
- Skinner, C. J., Holt, D. & Smith, T. M. F. (1989). *Analysis of Complex Surveys*. Chichester: Wiley.
- Smith, M. L. & Glass, G. V. (1977). Meta-analysis of Psychotherapy Outcome Studies. *The American Psychologist*, 32(9), 752–760.
- Smith, M. L. & Glass, G. V. (1980). Meta-Analysis of Research on Class Size and Its Relationship to Attitudes and Instruction. *American Educational Research Journal*, 17(4), 419.
- Smith, M. L., Glass, G. V. & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Snijlsteit, B. (2012). Systematic Reviews: From “bare Bones” Reviews to Policy Relevance. *Journal of Development Effectiveness*, 4(3), 388–408.
- Snijlsteit, B., Oliver, S. & Vojtkova, M. (2012). Narrative Approaches to Systematic Review and Synthesis of Evidence for International Development Policy and Practice. *Journal of Development Effectiveness*, 4(3), 409–429.
- Stewart, R., Van Rooyen, C. & De Wet, T. (2012). Purity or Pragmatism? Reflecting on the Use of Systematic Review Methodology in Development. *Journal of Development Effectiveness*, 4(3), 430–444.
- Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: Harvard University Press.
- Stigler, S. M. (1989). Francis Galton's Account of the Invention of Correlation. *Statistical Science*, 4(2), 73–79.
- Stone, P. J., Dunphy, D. C., Smith, M. S. & Ogilvie, D. M. (1968). *The General Inquirer: a Computer Approach to Content Analysis. User's Manual*. Cambridge, MA: MIT Press.
- Tashakkori, A. & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications, Inc.
- Thomas, J. & Harden, A. (2008). Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. *BMC Medical Research Methodology*, 8(1), 45.
- Thompson, S. G. & Higgins, J. P. (2002). How Should Meta-regression Analyses Be Undertaken and Interpreted? *Statistics in medicine*, 21(11), 1559–1573.
- Vuturo, G. J., McCormick, W. C. & Krischer, J. P. (1980). Potential and Actual Drug Product Selection Rates in Florida. *Contemporary pharmacy practice*, 3(3), 163–170.
- Waddington, H., White, H., Snijlsteit, B., Hombrados, J. G., Vojtkova, M., Davies, P., et al. (2012). How to Do a Good Systematic Review of Effects in International Development: a Tool Kit. *Journal of Development Effectiveness*, 4(3), 359–387.
- Weber, R. P. (1990). *Basic Content Analysis*. London: SAGE Publications Ltd.
- Weed, D. L. (2013). The Quality of Nutrition and Cancer Reviews: a Systematic Assessment. *Critical reviews in food science and nutrition*, 53(3), 276–286.
- Weinstein, M. C., Read, J. L., MacKay, D. N., Kresel, J. J., Ashley, H., Halvorsen, K. T., et al. (1986). Cost-effective Choice of Antimicrobial Therapy for Serious Infections. *Journal of general internal medicine*, 1(6), 351–363.
- White, H. & Waddington, H. (2012). Why Do We Care About Evidence Synthesis? An Introduction to the Special Issue on Systematic Reviews. *Journal of Development Effectiveness*, 4(3), 351–358.
- WHO. (2013). SURE Guides for Preparing and Using Evidence-Based Policy Briefs. *World Health Organization*. Retrieved February 14, 2013, from <http://www.who.int/evidence/sure/guides/en/>

- Winkelstein, W., Jr. (1998). The First Use of Meta-analysis? *American journal of epidemiology*, 147(8), 717.
- Yates, F. & Cochran, W. G. (1938). The Analysis of Groups of Experiments. *The Journal of Agricultural Science*, 28(04), 556–580.

Appendix

A1. A brief history of systematic reviews

Table A1. Timeline of systematic review development

What	Who	Year	Field	Reference
Statement: "As it is no easy matter to root out prejudices . . . it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy, and that in a chronological order, by which the sources of these mistakes may be detected. Indeed, before the subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish."	James Lind	1753	Medicine (scurvy)	Dunn (1997), Hampton (1998)
Statement: "It is impossible from single experiments, or from a great number, in different lands, separately considered, to deduce a satisfactory proof of the superiority of any method."	Arthur Young	1770	Sample surveys	(Brunt, 2001, p. 181)
The method of least squares to solve the problem of combining data from different astronomical observatories where the errors were known to be different	Adrien-Marie Legendre	1805	Statistics	(Stigler, 1986)
Concept of correlation	Francis Galton	1888	Statistics	Stigler (1989)
Review of theories and experiments on the psychology of time	Herbert Nichols	1891	Psychology experiments	Chalmers et al. (2002)
Calculation of the correlation coefficient for inoculation efficiency	Karl Pearson	1904	Statistics	Chalmers et al. (2002)
Writing a study protocol, criteria used to select studies for analysis, abstracting data and calculating average of pooled data	Joseph Goldberger	1906	Medicine	Winkelstein (1998)
Derived average results from two experiments	Edward L. Thorndike and Georgie J. Ruger	1916	Education experiments	Chalmers et al. (2002)
Combining the p-values that came from independent tests of the same hypothesis	Ronald A. Fisher	1925	Statistics	Fisher (1925)
The calculation of errors by the method of least squares	Raymond T. Birge	1932	Physics	Birge, (1932)
Summary of more than 180 experiments on the effects of education	Charles C. Peters	1933	Education	Peters (1933)
Analysis of groups of experiments in agriculture	F. Yates and William G. Cochran	1938	Agriculture	Yates and Cochran (1938)
Advocating the use of RCTs as the most reliable source of evidence in the book 'Effectiveness and efficiency'	Archie Cochrane	1972	Medicine	Cochrane (1972)
Meta-analysis	Gene Glass	1976	Education Psychotherapy	Glass (1976), Smith and Glass (1977), Smith and Glass (1980) Smith et al. (1980)
A lack of critical summary in medical research	Archie Cochrane	1979	Medicine	Cochrane (1979)
Oxford Database of Perinatal Trials	US Public Health Service and English Department of Health	1985	Medicine	Cochrane Collaboration (2012b)
Evidence-based medicine	Gordon Guyatt	1992	Medicine	Guyatt (1991)
First Cochrane Centre opened in the UK	The Cochrane Collaboration	1992	Medicine	Cochrane Collaboration (2012b)
Evidence-informed policy and practice	EPPI-Centre	1992	Non-clinical health issues: social, behavioral and educational interventions	EPPI-Centre (2009d)
The Cochrane Database of Systematic Reviews launched	The Cochrane Collaboration	1995	Medicine	Cochrane Collaboration (2013a)
Evidence-based practice	David Sackett	1996	Medicine	Sackett et al. (1996)
Evidence based public policy	The Campbell Collaboration	2000	Public policy	Campbell Collaboration (2013)
Grants for centers to conduct systematic reviews in low- and middle-income countries	Alliance for Health Policy and Systems Research, hosted by the WHO	2007	Health policy and systems research	EPPI-Centre (2009c)

A2. Databases used for the study

Table A2. A list of databases used for the study

Involved in SR	Link	Based in	Scientific disciplines (thematic fields)	Most frequent methods in primary studies	Date search performed
<u>Databases</u>					
Cochrane Collaboration	http://summaries.cochrane.org/	UK/global	Medicine	Double-blind RCTs, RCTs, clinical trials and observational studies	13/12/2012
Campbell Collaboration	http://www.campbellcollaboration.org/	Norway/global	Crime and justice Education International Development Methods Social Welfare Medicine	Double-blind RCTs, quasi-RCTs and qualitative research	04/12/2012
Centre for Reviews Dissemination University of York	http://www.crd.york.ac.uk/crdweb	UK	Medicine	Double-blind RCTs, RCTs, clinical trials and observational studies	04/12/2012
Evidence in Health and Social Care	http://www.evidence.nhs.uk/search?om=%5B%7B%22i%22%3A%5B%22%20systematic%20reviews%20%22%5D%7D%5D	UK (linked with CRD and EPPI)	Medicine	Double-blind RCTs, survey data, quasi-RCTs and qualitative research	03/12/2012
Health-Evidence.ca	http://health-evidence.ca/articles/search	Canada	Medicine	Survey data, double-blind RCTs, qualitative research and diagnostic studies	12/12/2012
SIGLE System for Information Retrieval on Gray Literature	http://www.opengrey.eu/	EU (France)	Medicine, pharmacy, psychology, management, social sciences, education, law, urban/rural, IT	Quasi-RCTs and survey data	19/12/2012
ERIC	http://www.eric.ed.gov/	US	Education	Case-control, uncontrolled cohort studies, cohort studies and survey data	18/12/2012
PsycINFO	http://www.apa.org/pubs/databases/psycinfo/index.aspx	US	Behavioral sciences and mental health	Registry reports, case-control and clinical guidelines	07/12/2012
ProQuest	http://www.proquest.co.uk/en-UK/	US	Multidisciplinary	Registry reports, clinical studies, uncontrolled cohort studies and case-control	05/12/2012
SocINDEX	http://www.ebscohost.com/academic/socindex-with-full-text	US/global	Sociology	RCTs, registry reports, case studies and qualitative research	14/12/2012
Web of Knowledge	http://wokinfo.com/	US/global	Multidisciplinary	RCTs, observational studies, quasi-RCTs and survey data	12/12/2012
Libraries WorldCat	http://www.worldcat.org/	US/global	Multidisciplinary	Cohort studies, survey and RCTs	12/12/2012
<u>Academic databases</u>					
Scopus	http://www.scopus.com/	Netherlands/global	Multidisciplinary (medicine, life sciences, physical sciences, social sciences and humanities)	Clinical trials, clinical studies, RCTs and survey	11/12/2012
Taylor and Francis Online	http://www.tandfonline.com/	UK	Multidisciplinary	Survey, case studies, qualitative research	18/12/2012

Wiley Online Library	http://onlinelibrary.wiley.com/advanced/search	US	Multidisciplinary	and clinical trials Clinical studies, RCTs, clinical trials and case studies	17/12/2012
Science direct	http://www.sciencedirect.com/	Netherlands/global	Multidisciplinary (medicine, life sciences, physical sciences, social sciences and humanities)	RCTs, clinical trials, observational studies and cohort studies	12/12/2012
Unbound Medline	http://www.unboundmedicine.com/medline/ebm	US/global	Medicine	RCTs, survey, clinical trials and observational studies	11/12/2012
Medline	http://www.thelancet.com/search	US/global	Medicine	RCTs, clinical trials, cohort studies and observational studies	11/12/2012
BioMed Central	http://www.biomedcentral.com/search	UK	Medicine	Double blind RCTs, RCTs, clinical trials and observational studies	10/12/2012
Ebsco Academic Search Complete	http://web.ebscohost.com/ehost/	US/global	Multidisciplinary	RCTs, clinical trials, observational studies and cohort studies	19/12/2012
SAGE journals	http://online.sagepub.com/	US/global	Multidisciplinary (medicine, life sciences, engineering, social sciences and humanities)	RCTs, clinical trials, survey and cohort studies	18/01/2013
Emerald	http://www.emeraldinsight.com/	UK	Multidisciplinary	Survey, case studies, RCTs and observational studies	11/12/2012
Ingentaconnect	http://www.ingentaconnect.com/	UK	Multidisciplinary	RCTs, clinical trials, cohort studies and observational studies	12/12/2012
Annual Reviews	http://www.annualreviews.org/	US	Multidisciplinary	Survey, clinical trials, cohort studies and case studies	17/01/2013
SSRN	http://papers.ssrn.com/	US/global	Social sciences	Survey, clinical trials, case studies and clinical guidelines	18/01/2013
Sociological Abstracts	http://www.csa.com/factsheets/socioabs-set-c.php	US	Sociology	RCTs, survey, clinical trials and case studies	18/01/2013
International Political Science Abstracts	http://iab.sagepub.com/	US	Political science	Survey, survey data, qualitative data and qualitative research	12/12/2012
PAIS	http://www.csa.com/factsheets/pais-set-c.php	US	Political science	RCTs, survey, observational studies and cohort studies	18/12/2012
CAB Abstracts	http://www.cabi.org/default.aspx?site=170&page=1016&pid=125	Global	Life sciences	RCTs	18/01/2013
Communication and Mass Media Complete	http://www.ebscohost.com/academic/communication-mass-media-complete	US	Communication	RCTs, case studies and survey	18/01/2013
EconLit	http://search.proquest.com/econlit/	US	Economics	Survey, clinical trials, observational studies and case studies	18/01/2013
JSTOR	http://www.jstor.org	US	Multidisciplinary	Case control, RCTs, survey and double blind RCTs	18/01/2013
International Bibliography of the Social Sciences (IBSS)	http://www.proquest.co.uk/en-UK/catalogs/databases/detail/ibss-set-c.shtml	US	Social sciences	Survey, clinical trials, cohort studies and case studies	18/01/2013
<u>Journals</u>					
Systematic Reviews Journal	www.systematicreviewsjournal.com	UK	Systematic Reviews	RCTs, clinical trials, observational studies and survey	22/01/2013
Journal of Development Effectiveness	http://www.tandfonline.com/toc/rjde20/current	US	International Development	Case-control, RCTs, clinical trials and cohort studies	22/01/2013
Research Synthesis Methods	http://onlinelibrary.wiley.com/advanced/search/results	US	Methodology	Uncontrolled cohort studies, diagnostic studies, cohort studies	22/01/2013

The Lancet	http://www.thelancet.com/search	UK	Health	and case studies Observational studies, qualitative data, case-control and RCTs	22/01/2013
Trials	http://www.trialsjournal.com/search	UK	Health	Survey data, cohort studies, clinical trials and qualitative research	22/01/2013
PLOS ONE	http://www.plosone.org/search	US	Multidisciplinary	Qualitative data, clinical trials, diagnostic studies and RCTs	22/01/2013
Evidence-based medicine	http://ebm.bmj.com/	UK	Health	Survey data, clinical studies, observational studies and cohort studies	22/01/2013
<u>Institutions</u>					
EVIPNet	http://search.bvsalud.org/evipnet/index.php	Global	Health, public policies, international development	RCTs, clinical trials, cohort studies and double blind RCTs	23/01/2013
Evidence Informed Policy Network	http://search.who.int/	Global	Health, international development	RCTs, clinical trials, case studies and case control	23/01/2013
World Health Organization	http://search.who.int/	Global	Health, international development	RCTs, clinical trials, case studies and case control	23/01/2013
World Bank	http://search.worldbank.org/	Global	Economics, international Development	Survey, survey data, case studies and qualitative data	23/01/2013
International Initiative for Impact Evaluation (3ie)	http://www.3ieimpact.org/en/evidence/systematic-reviews/?q=	US	International Development	Case studies, clinical studies, qualitative data and clinical trials	21/01/2013
Department for International Development, UK	http://www.dfid.gov.uk/R4D/Search/BrowseByTheme.aspx	UK	International Development	Case studies, survey, qualitative research and RCTs	21/01/2013
Overseas Development Institute	http://www.odi.org.uk/search	UK	International Development	Case studies, survey, qualitative data and qualitative research	21/01/2013
EPPI-Centre	http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=62	UK	Public policies	Mixed methods	25/01/2013
Collaboration for Environmental Evidence	http://www.environmentalevidence.org/Reviews.htm	UK, Australia and South Africa	Environment	Survey, experiment, case-control	25/01/2013
AusAID	http://www.ausaid.gov.au/research/Pages/Selection.aspx	Australia	International Development	Case studies, clinical studies, qualitative data and clinical trials	25/01/2013

A3. Glossary

Table A3. Definition of key terms

Term	Definition	Author
Aggregated (disaggregated) analysis	Analysis intended to follow the population structure in detail, estimating all level variances, for example, is referred to as a disaggregated analysis. Analysis focused on the relations between just a portion of the variables, for example, ignoring cluster structure, is aggregated analysis.	Skinner et al. (1989)
Blinding (single-blind, double-blind, triple-blind)	The process of preventing those involved in a trial from knowing to which comparison group a particular participant belongs. The risk of bias is minimized when as few people as possible know who is receiving the experimental intervention and who the control intervention. Participants, caregivers, outcome assessors, and analysts are all candidates for being blinded. Blinding of certain groups is not always possible, for example surgeons in surgical trials. The terms single blind, double blind and triple blind are in common use, but are not used consistently and so are ambiguous unless the specific people who are blinded are listed.	Cochrane Collaboration (2013b)
Case-control study	A study that compares people with a specific disease or outcome of interest (cases) to people from the same population without that disease or outcome (controls), and which seeks to find associations between the outcome and prior exposure to particular risk factors. This design is particularly useful where the outcome is rare and past exposure can be reliably measured. Case-control studies are usually retrospective, but not always.	Cochrane Collaboration (2013b)
Case study	In medicine, a case study is a study reporting observations on a single individual. It is also called an anecdote, a case history or a single case report.	Cochrane Collaboration (2013b)
Clinical guideline	A systematically developed statement for practitioners and participants about appropriate health care for specific clinical circumstances.	Cochrane Collaboration (2013b)
Clinical trial	An experiment to compare the effects of two or more healthcare interventions. Clinical trial is an umbrella term for a variety of designs of healthcare trials, including uncontrolled trials, controlled trials, and randomized controlled trials. It is also called an intervention study.	Cochrane Collaboration (2013b)
Cohort study	An observational study in which a defined group of people (the cohort) is followed over time. The outcomes of people in subsets of this cohort are compared, to examine people who were exposed or not exposed (or exposed at different levels) to a particular intervention or other factor of interest. A prospective cohort study assembles participants and follows them into the future. A retrospective (or historical) cohort study identifies subjects from past records and follows them from the time of those records to the present. Because subjects are not allocated by the investigator to different interventions or other exposures, adjusted analysis is usually required to minimize the influence of other factors (confounders).	Cochrane Collaboration (2013b)
Content analysis	The study of the content with reference to the meanings, contexts and intentions contained in messages. “Any technique for making inferences by objectively and systematically identifying specified characteristics of messages.” “Research technique for making replicable and valid inferences from data to their context.” “Research methodology that utilizes a set of procedures to make valid inferences from text. These inferences are about sender(s) of message, the message itself, or the audience of message.” “Any procedure for assessing the relative extent to which specified references, attitudes, or themes permeate a given message or document.”	Berelson (1952) Holsti (1969) Krippendorff (2003) Weber (1990)
Critical appraisal	The assessment of evidence by systematically reviewing its relevance, validity and results to specific situations.	Stone et al. (1968) Chambers (1998)
Critical interpretive synthesis (critical synthesis)	An approach to reviewing multi-disciplinary and multi-method evidence. It is an iterative approach: stages such as defining research question, searching and selecting from literature, defining and applying codes and categories are not exclusively fixed. It determines the quality of reviewed research in terms of their theoretical contribution. The product of the synthesis is not aggregations of data, but theory grounded in the studies included in the review.	Dixon-Woods et al. (2006)
Cross-case analysis	An analysis that examines themes, similarities, and differences across cases is referred to as a cross-case analysis. Cross-case analysis is used when the unit of analysis is a case, which is any bounded unit, such as an individual, group, artifact, place, organization, or interaction. It is used in quantitative, statistical analysis, such as hierarchical modeling and in qualitative analysis, such as in grounded theory approach. The focus is on a particular common outcome for a number of cases.	Mathison (2005, p. 95)
Diagnostic study	A study to assess the effectiveness of a test or measurement in detecting	NHS (2011)

Evidence synthesis	whether someone has (or does not have) a specific disease. Evidence synthesis is a synthesis (or integration) of variable data to produce information in the form of best evidence. It provides a set of methodologies to identify areas of agreement and disagreement of qualitative and quantitative datasets. By integrating datasets it enables calculating the concordance and magnitude of effects from multiple studies. The aim of evidence synthesis is to address questions by providing the best evidence derived through the integration of data and knowledge to present information of factual integrity and least uncertainty.	Ashrafiyan et al. (2011)
Information Synthesis	It involves systematically gathering, evaluating, and presenting information in a form useful to the intended audience.	Goldschmidt (1986)
Integrated reviews	An integrative review is a specific review method that summarizes past empirical or theoretical literature to provide a more comprehensive understanding of a particular phenomenon or problem. It allows for the inclusion of diverse methodologies (i.e. experimental and non-experimental research) and combining data from the theoretical as well as empirical literature.	Broome (1993)
Meta-analysis (Bayesian synthesis)	The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings	Smith and Glass (1977)
Meta-ethnography	An approach to synthesize understanding from ethnographic accounts. The function is more interpretative than aggregative. Meta-ethnography requires three methods of synthesis 1) the translation of concepts from individual studies into one another (reciprocal translational analysis); 2) refutational synthesis, which explores and explains contradictions between individual studies; and 3) lines of argument synthesis, building up a picture of the whole.	Noblit and Hare (1988)
Meta-narrative	Aims to synthesize knowledge across different research paradigms, study disciplines and study designs.	Greenhalgh et al. (2005)
Meta-regression	Meta regression investigates the extent to which statistical heterogeneity between results of multiple studies can be related to one or more characteristics of the studies. It is usually conducted on study-level summary data, because individual observations from all studies are frequently not available.	Thompson and Higgins (2002)
Meta-study	Gives a critical interpretation of existing qualitative research. It contains 3 segments of analysis: meta-data-analysis, meta-method and meta-theory.	(Paterson et al., 2001)
Meta-synthesis	The synonym is qualitative meta-analysis. “It is the bringing together and breaking down of findings, examining them, discovering the essential features, and, in some way, combining phenomena into a transformed whole.” Types of meta-syntheses include theory building, meta-study, grounded formal theory, theory explication, and descriptive study.	(Schreiber et al., 1997, p. 314) Finfgeld (2003)
Mixed methods research synthesis (MMRS)	MMRS is a synthesis in which researchers combine qualitative, quantitative, and mixed methods studies, and apply a mixed methods approach in order to integrate those studies, for the broad purposes of breadth and depth of understanding and corroboration.	Heyvaert et al. (2013)
Mixed-methods systematic reviews (the EPPI center method)	In the mixed-methods systematic reviews, there are three ways in which the reviews are mixed: 1. The types of studies included in the review are mixed; hence, the types of findings to be synthesized are mixed. 2. The synthesis methods used in the review are mixed—statistical meta-analysis and qualitative.	(Harden, 2008)
Mixed methods synthesis	3. The review uses two modes of analysis—theory building and theory testing. The first stage is a traditional systematic review of effectiveness (with or without meta-analysis); the second a synthesis of qualitative research which addresses questions of intervention need, implementation, acceptability, and appropriateness; and, finally a cross-study synthesis which brings the findings of both earlier syntheses together.	Kavanagh et al. (2011)
Narrative synthesis	An approach that relies primarily on the use of words and text to summarize and explain the findings of multiple studies. Whilst narrative synthesis can involve the manipulation of statistical data, the defining characteristic is that it adopts a textual approach to the process of synthesis to ‘tell the story’ of the findings from the included studies. It is used to synthesize both quantitative and qualitative evidence. It brings out context and characteristics of each study.	Popay et al. (2006)
Qualitative synthesis	In a qualitative synthesis, primary qualitative studies are integrated to develop a theory or evidence-based interventions. It systematically searches for research on a topic, and draws the findings from individual studies together (sometimes called a qualitative systematic review). It is quite a generic term, without specific protocol. It involves the use of other methods.	Boeije et al. (2011)
Qualitizing quantitative data	Involves converting quantitative data into narrative data that can be analyzed qualitatively	Tashakkori and Teddlie (1998)
Quasi-random allocation	Methods of allocating people to a trial that are not random, but were intended to produce similar groups when used to allocate participants. Quasi-random methods include: allocation by the person’s date of birth, by	Cochrane Collaboration (2013b) and 3ie (2012b)

	the day of the week or month of the year, by a person's medical record number, or just allocating every alternate person. This group of studies is often called quasi-RCTs or quasi-experimental. Some examples of quasi-experimental methods include propensity score matching, regression discontinuity and instrumental-variable regressions.	
Observational study	A study in which the investigators do not seek to intervene, and simply observe the course of events. Changes or differences in one characteristic (e.g. whether or not people received the intervention of interest) are studied in relation to changes or differences in other characteristic(s) (e.g. whether or not they died), without action by the investigator. There is a greater risk of selection bias than in experimental studies. It is also called a non-experimental study.	Cochrane Collaboration (2013b)
Randomized (random)	The process of randomly allocating participants into one of the arms of a controlled trial. There are two components to randomization: the generation of a random sequence, and its implementation, ideally in a way so that those entering participants into a study are not aware of the sequence (concealment of allocation).	Cochrane Collaboration (2013b)
Randomized control trial	An experiment in which two or more interventions, possibly including a control intervention or no intervention, are compared by being randomly allocated to participants. In most trials one intervention is assigned to each individual but sometimes assignment is to defined groups of individuals (for example, in a household) or interventions are assigned within individuals (for example, in different orders or to different parts of the body). It is also called a Randomized clinical trial (RCT).	Cochrane Collaboration (2013b)
Rapid evidence assessment	Rapid evidence assessments are used to summarize the available research evidence within the constraints of a given timetable, typically three months or less. Rapid evidence assessments differ from full systematic reviews in terms of the time available to prepare them and the extent of the literature searches and other review activities. Whilst attempting to be as comprehensive as possible, rapid evidence assessments must make compromises to meet their tight deadlines, therefore, they may fail to identify potentially relevant studies. They are useful to policy makers who need to make decisions quickly, but should be viewed as provisional appraisals, rather than full systematic reviews.	CRD (2009)
Re-analysis	It is used for verification or corroboration of the original interpretations rather than to address new research questions.	Heaton (2004, p. 45)
Realist synthesis/ Realist review	Realist synthesis/review is a broad approach to evidence review that focuses primarily on understanding the causal mechanisms or 'theories of change' that underlie a particular type of intervention or program. The basic evaluative question – what works? – changes to 'what is it about this program that works for whom in what circumstances?'	Pawson et al. (2004, 2005)
Research synthesis	A research synthesis is a general term used to describe the 'bringing together' of a body of research on a particular topic.	
	Research synthesis attempts to integrate empirical research for the purpose of creating generalizations. It pays attention to relevant theories, critically analyzes the research they cover, tries to resolve conflicts in the literature, and attempts to identify central issues for future research.	Cooper and Hedges (1994, p. 6)
Review of reviews	This describes a systematic review that includes only other systematic reviews. In theory the systematic reviews included in the review should have covered most of the primary studies available. Reviews of reviews are likely to be helpful when a review question is very broad and a number of systematic reviews have already been conducted in the topic area. However, the different inclusion criteria adopted by the various reviews can make their synthesis problematic.	CRD (2009)
Scoping review	Involves a search of the literature to determine what sorts of studies addressing the systematic review question have been carried out, where they are published, in which databases they have been indexed, what sorts of outcomes they have assessed, and in which population. A scoping review determines the size and nature of the evidence base for a particular topic area, which can in turn be used to identify gaps in the literature and make recommendations for future primary research. The literature search should be as extensive as possible, including a range of relevant databases, hand searching and attempts to identify unpublished literature. Scoping reviews differ from standard systematic reviews in that they do not attempt to synthesize the evidence. A scoping review might be useful to research bodies that are planning a primary study, or to assess the feasibility of a full systematic review. It is not appropriate to use a scoping review to answer a clinical question.	Petticrew and Roberts (2008)
Supplementary analysis	Supplementary analysis considers a more in-depth investigation of an emergent issues or aspect of the data that was not addressed, or was only partially addressed by the primary research. It is related to the analytical remit of the primary study, aiming to extend rather than exceed the original work.	Heaton (2004, p. 41)
Survey	The collection of information using (1) a pre-defined sampling strategy, and (2) a survey instrument. A survey may collect data from individuals, households or other units such as firms or schools.	3ie (2012b)

Synthesis of qualitative research	It is a generic term denoting the process of integrating qualitative research through appropriate methods that correspond to approaches used in primary research. Some of the methods maintain the qualitative form of the evidence such as meta-ethnography and some involve converting qualitative findings into a quantitative form such as content analysis.	CRD (2009)
Systematic mapping	Systematic map aims to describe the existing literature, and gaps in the literature, in a broad topic area, and the literature quality and content can be analyzed in depth or more superficially as appropriate to individual projects. Systematic maps gather together existing literature in a specific topic area and categorize it according to predefined keywords to create a coded database of literature. The topic area can be broad or narrow depending on the needs of the project in question.	Clapton et al. (2009) Bates et al. (2007)
Thematic analysis	Thematic analysis is a qualitative analytic method for: identifying, analyzing and reporting patterns (themes) within data. It minimally organizes and describes your data set in (rich) detail. However, frequently it goes further than this, and interprets various aspects of the research topic.	Braun and Clarke (2006, p. 6)
Thematic synthesis	Analysis in systematic reviews aiming to bring together and integrate the findings of multiple qualitative studies. Thematic synthesis has three stages: the coding of text 'line-by-line'; the development of 'descriptive themes'; and the generation of 'analytical themes'	Thomas and Harden (2008)
Validity generalization	VG is a particular type of psychometric meta-analysis conducted to determine whether a particular psychological construct, test, or measure has validity in predicting job performance regardless of situation or setting.	Schmidt et al. (1985)

A4. Specific tools for quality assurance of systematic reviews

The Cochrane handbook requires that authors assess not only the quality of evidence that is incorporated in the review, but also the quality of the final review. The most common tools to assess the quality of systematic reviews are Overview Quality Assessment Questionnaire (OQAQ) and Assessment of Multiple Systematic Reviews (AMSTAR) (Pieper et al., 2012). The OQAQ contains nine questions focusing on different aspects of scientific quality of a systematic review (search strategy, selection strategies, quality assessment, pooling and results) and a question that evaluates the overall scientific quality of the review on a 7-point scale (Oxman & G. H. Guyatt, 1991). Appraising the reviews based on the AMSTAR approach means giving answers to 11 questions that cover the following topics: design of the review, nature of the literature search, characteristics and scientific quality of the included studies, appropriateness of the synthesis methods, assessment of the likelihood of publication bias and potential conflicts of interest (Shea et al., 2007). It should be noted that the quality assessment tools are being constantly revised and upgraded. For example, after a critique that it cannot produce quantifiable assessments, the AMSTAR was re-evaluated and transformed into the revised AMSTAR (R-AMSTAR) (Kung et al., 2010).

The application of different scales for appraising the quality of systematic reviews is, however, not widespread. The Web of Knowledge search has shown 97 articles using the AMSTAR and 24 articles using the OQAQ. The widest application of the tools is in the so-called reviews of reviews. Pieper et al. (2012) report that 64% of such reviews uses some of the quality assessment tools.

Published by:
Ministry of Foreign Affairs of Denmark
Evaluation Department
Asiatisk plads 2
1448 Copenhagen K
Denmark
E-mail: eval@um.dk

The publication can be downloaded from:
www.evaluation.dk

ISBN: 978-87-7087-731-2
ISBN: 978-87-7087-732-9 (internet version)

Other Evaluation Studies

- 2009/1: Evaluating Aid Effectiveness in the Aggregate:
Methodological Issues
- 2009/2: Addressing the Underlying and Basic Causes of
Child Undernutrition in Developing Countries:
What Works and Why?
- 2009/3: Experiences with Conducting Evaluations Jointly with
Partner Countries
- 2009/4: Synthesis of Impact Evaluations of Microcredit
- 2009/5: Synthesis of Evaluations on Support to Business
Development
- 2010/1: Evaluating Aid Effectiveness in the Aggregate:
A critical assessment of the evidence
- 2010/2: Gender and Value Chain Development
- 2011/1: CCPA's Open Fun Football Schools Programme
- 2011/2: Agricultural input subsidies in Sub-Saharan Africa
- 2012/1: The macroeconomic impact of HIV/AIDS and HIV/AIDS
interventions
- 2012/2: Economic development and service delivery in fragile
states
- 2012/3: Effective statebuilding? A review of evaluations
of international statebuilding support in fragile contexts
- 2012/4: Danish support to statebuilding and improved
livelihoods in Afghanistan

