HEALTH SERVICES AND DELIVERY RESEARCH

VOLUME 2 ISSUE 30 SEPTEMBER 2014
ISSN 2050-4349

Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project

Geoff Wong, Trish Greenhalgh, Gill Westhorp and Ray Pawson

National Institute for Health Research

DOI 10.3310/hsdr02300

Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project

Geoff Wong,[1]* Trish Greenhalgh,[1] Gill Westhorp[2] and Ray Pawson[3]

[1]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK
[2]Community Matters, Mount Torrens, SA, Australia
[3]School of Sociology and Social Policy, University of Leeds, Leeds, UK

*Corresponding author

Declared competing interests of authors: none

Published September 2014
DOI: 10.3310/hsdr02300

This report should be referenced as follows:

Wong G, Greenhalgh T, Westhorp G, Pawson R. Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project. *Health Serv Deliv Res* 2014;**2**(30).

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hsdr

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its proceeding programmes as project number 10/1008/07. The contractual start date was in April 2011. The final report began editorial review in July 2013 and was accepted for publication in January 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project

Geoff Wong,[1]* Trish Greenhalgh,[1] Gill Westhorp[2] and Ray Pawson[3]

[1]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK
[2]Community Matters, Mount Torrens, SA, Australia
[3]School of Sociology and Social Policy, University of Leeds, Leeds, UK

*Corresponding author

Background: There is growing interest in theory-driven, qualitative and mixed-method approaches to systematic review, such as realist and meta-narrative review. These approaches offer the potential to expand the knowledge base in policy-relevant areas. However, the quality of such reviews can be difficult to assess.

Objectives: The aim of this project was to produce methodological guidance, publication standards and training resources for those seeking to undertake realist and/or meta-narrative reviews.

Methods/design: We (1) collated and summarised existing literature on the principles of good practice in realist and meta-narrative systematic reviews; (2) considered the extent to which these principles had been followed by published and in-progress reviews, thereby identifying how rigour may have been lost and how existing methods could be improved; (3) used an online Delphi method with an interdisciplinary panel of experts from academia and policy, to produce a draft set of methodological steps and publication standards; (4) produced training materials with learning objectives linked to these steps; (5) refined these standards and training materials prospectively on real reviews in progress, capturing methodological and other challenges as they arose; (6) synthesised expert input, evidence review and real-time problem analysis into more definitive guidance and standards; and (7) disseminated outputs to audiences in academia and policy.

Results: An important element of this study was the establishment of an e-mail mailing list to bring together researches in the field (www.jiscmail.ac.uk/RAMESES). Our literature review identified 35 and nine realist and meta-narrative reviews respectively. Analysis and discussion within the project team produced a summary of the published literature, and common questions and challenges into briefing materials for the Delphi panel, comprising 37 and 33 members (for realist and meta-narrative reviews respectively). Within three rounds this panel had reached a consensus on 19 (realist) and 20 (meta-narrative) key publication standards, with an overall response rate of 90% and 91% respectively. The Realist And Meta-narrative Evidence Syntheses – Evolving Standards (RAMESES) publication standards for realist syntheses and meta-narrative reviews were published in open-access journals and quickly became highly accessed. The RAMESES quality standards and training materials drew together the following sources of data: (1) personal expertise as researchers and trainers; (2) data from the Delphi panels; (3) feedback from participants at training sessions we ran; and (4) comments made on RAMESES mailing list. The quality standards and training materials are freely available online (www.ramesesproject.org).

Discussion: The production of these standards and guidance drew on multiple sources of knowledge and expertise, and a high degree of a consensus was achieved despite ongoing debate among researchers about the overall place of these methodologies in the secondary research toolkit. As with all secondary research methods, guidance on quality assurance and uniform reporting is an important step towards improving quality and consistency of studies. We anticipate that as more reviews are undertaken, further refinement will be needed to the publication and quality standards and training materials.

Limitations: The project's outputs are not definitive and in the future updating and further development is likely to be needed.

Conclusion: An initial set of publication standards, quality standards and training materials have been produced for researchers, users and funders of realist or meta-narrative reviews. As realist and meta-narrative reviews are relatively new approaches to evidence synthesis, methodological development is needed for both review approaches.

Funding: The National Institute for Health Research Health Services and Delivery Research programme.

Contents

List of tables

List of figures

List of boxes

List of abbreviations

AGREE	Appraisal of Guidelines for Research and Evaluation
C	context
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CMO	context, mechanism and outcome
CONSORT	Consolidated Standards of Reporting Trials
EQUATOR	Enhancing the QUAlity and Transparency Of health Research
ERIC	Education Resources Information Center
HSDR	Health Services and Delivery Research
M	mechanism
NIHR	National Institute for Health Research
O	outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAMESES	Realist And Meta-narrative Evidence Syntheses – Evolving Standards
RS	realist synthesis (or review)
SCI	Science Citation Index
SQUIRE	Standards for Quality Improvement Reporting Excellence
SSCI	Social Science Citation Index

Plain English summary

Every year, a lot of research is published. No one is able to read all of this research and so some researchers produce summaries – called literature (or systematic) reviews. There are many different ways of doing systematic reviews; realist and meta-narrative review are two relatively new approaches which both seek to explain why and how interventions work. When we started this project, there were no standards setting out how to judge if realist or meta-narrative reviews were of high quality – something we have called quality standards. Nor did any standards exist to guide researchers on how best to write up their reviews for publication – we have called these publication standards. Although there were some training materials for these review methods, more were needed which showed researchers in detail how to rigorously undertake certain parts of a review.

In this project, we developed quality and publication standards and training materials for realist and meta-narrative reviews. We used a range of information to help us choose and agree on what should be in the standards and training materials. We gathered together a group of experts. We set up an e-mail list and invited people to join and contribute. We asked researchers we worked with on realist or meta-narrative reviews for their comments, and we got feedback from researchers we trained in workshops or presented to at conferences. We analysed and wove together all this information to produce quality and publication standards and training materials. We have made all the outputs of our work freely available online (www.ramesesproject.org).

Scientific summary

Background

Academics and policy-makers are increasingly interested in policy-friendly approaches to evidence synthesis which seek to illuminate issues and understand contextual influences on whether, why and how interventions might work. A number of different approaches have been used to address this goal. Qualitative and mixed-method reviews are often used to supplement, extend and in some circumstances replace Cochrane-style systematic reviews. Theory-driven interpretive approaches to such reviews include realist and meta-narrative review. Realist review was originally developed by Pawson for complex social interventions to explore systematically how contextual factors influence the link between intervention and outcome (summed up in the question: what works, how, for whom, in what circumstances and to what extent?) (Pawson R. *Evidence-based Policy: A Realist Perspective*. London: Sage; 2006). Greenhalgh *et al.* developed a meta-narrative review for use when a policy-related topic has been researched in different ways by multiple groups of scientists, especially when key terms have different meanings in different literatures (Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med* 2005;**61**:417–30).

Quality checklists and publication standards are common (and, increasingly, expected) in health services research. They have two main purposes: they help researchers design and undertake robust studies, and they help reviewers and potential users of research outputs assess validity and reliability. This project seeks to produce a set of quality criteria, comparable publication guidance and training materials for realist and meta-narrative reviews.

Objectives

1. To collate and summarise the literature on the principles of good practice in realist and meta-narrative reviews, highlighting in particular how and why these differ from conventional forms of systematic review and from each other.
2. To consider the extent to which these principles have been followed by published and in-progress reviews, thereby identifying how rigour may be lost and how existing principles could be improved.
3. To use an online Delphi method with an interdisciplinary panel of experts from academia and policy, to produce, in draft form, an explicit and accessible set of methodological guidance and publication standards.
4. To produce training materials with learning objectives linked to these steps and standards.
5. To refine these standards and training materials prospectively on real reviews-in-progress, capturing methodological and other challenges as they arise.
6. To synthesise expert input, evidence review and real-time problem analysis into more definitive guidance and standards.
7. To disseminate these guidance and standards to audiences in academia and policy.

Methods

To fulfil objectives 1 and 2, we undertook a narrative review of the literature that was supplemented by collating feedback from presentations and workshops. We synthesised our findings into briefing materials (one for realist synthesis and another for meta-narrative reviews). We recruited members to two Delphi

panels, which had wide representation from researchers, students, policy-makers, theorists and research sponsors. We used the briefing materials to brief the Delphi panel so they could help us in fulfilling objective 3.

For objective 4, we drew not only on our experience in developing and delivering education materials, but also on relevant feedback from the Delphi panel, an e-mail list we set up specifically for this project (www.jiscmail.ac.uk/RAMESES), training workshops and the review teams we supported methodologically. To help us refine our publication standards (objective 5) we captured methodological and other challenges that arose within the realist or meta-narrative review teams to which we provided methodological support.

To produce the definitive publication standards, quality standards and training materials (objective 6), we synthesised expert input (from the Delphi panel), literature review and real-time problem analysis (e.g. feedback from the e-mail list, training sessions and workshops, and presentations).

Throughout this study, we iteratively and contemporaneously fed any data we captured into our draft publication standards, quality standards and training materials, making changes gradually. The definitive guidance and standards were, thus, the product of continuous refinements. We addressed objective 7 through academic publications, online resources and delivery of presentations and workshops.

Results

An important early output of this study was an e-mail mailing list (www.jiscmail.ac.uk/RAMESES) to bring together researchers in the field, especially since before the study began researchers were dispersed across the globe and many were working in isolation. The list at present has over 350 members and it regularly serves as a resource for its members to ask and get help with methodological questions. The Realist And Meta-narrative Evidence Syntheses – Evolving Standards (RAMESES) list will continue to run after the end of this project.

Our literature review identified 35 realist reviews and and nine meta-narrative reviews. Analysis and discussion within the project team produced a summary of the published literature, and common questions and challenges in briefing materials for the Delphi panel, comprising 37 and 33 members (for realist and meta-narrative reviews respectively). There was an overlap in the membership of the panels. Within three rounds the panels had reached a consensus on 19 realist and 20 meta-narrative key publication standards, with an overall response rate of 90% and 91% respectively. The RAMESES publication standards for realist syntheses and meta-narrative reviews have been published in open access journals and the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network (www.equator-network.org).

The quality standards and training materials drew on the following sources of data: (1) personal expertise as researchers and trainers; (2) data from the Delphi panels; (3) feedback from participants at training sessions we ran; and (4) comments made on RAMESES mailing list. We developed eight quality criteria for realist syntheses and nine for meta-narrative reviews. Versions of these quality criteria were developed for researchers, peer reviewers and funders/commissioners of research. For our training materials we used the data we captured to identify the methodological topics that were identified by the majority of reviewers as most challenging. We developed training materials for four methodological topics in realist reviews and three in meta-narrative reviews. The quality standards and training materials are freely available online (www.ramesesproject.org).

Limitations

This project developed quality and publication standards and training materials for realist and meta-narrative reviews. These outputs are not definitive. As practice and experience in the use of these review approaches increases, we anticipate that these standards and materials are likely to require updating and further development.

Conclusions

Realist and meta-narrative reviews are relatively new approaches to systematic review whose overall place in the secondary research toolkit is not yet fully established. As with all secondary research methods, if used, guidance on quality assurance and uniform reporting is an important step towards improving quality and consistency of reviews. This project has developed the first ever set of such standards and materials. However, further methodological development is needed for both review approaches. These developments should help to refine this project's outputs. Formal evaluations of the value of the project's outputs have not been undertaken and may be of value. Capacity building remains an important area for the future. We anticipate that, as more reviews are undertaken, further refinement will be needed to the publication and quality standards and training materials.

Funding

Funding for this study was provided by the Health Services and Delivery Research programme of the National Institute for Health Research.

Chapter 1 Background

Academics and policy-makers are increasingly interested in policy-friendly approaches to evidence synthesis which seek to illuminate issues and understand contextual influences on whether or not, why and how interventions might work.[1–4] A number of different approaches have been used to try to address this goal, such as meta-ethnography, grounded theory, thematic synthesis, textual narrative synthesis, meta-study, critical interpretive synthesis, ecological triangulation and framework synthesis.[5] Qualitative and mixed-method reviews are often used to supplement, extend and, in some circumstances, replace Cochrane-style systematic reviews.[6–12] Theory-driven approaches to such reviews include realist and meta-narrative reviews. Realist review was originally developed by Pawson for complex social interventions to explore systematically how contextual factors influence the link between intervention and outcome (summed up in the question: what works, how, for whom, in what circumstances and to what extent?).[13,14] Greenhalgh *et al.*[15] developed a meta-narrative review for use when a policy-related topic has been researched in different ways by multiple groups of scientists, especially when key terms have different meanings in different literatures.

Quality checklists and publication standards are common (and, increasingly, expected) in health services research – see for example CONSORT (Consolidated Standards of Reporting Trials) for randomised controlled trials,[16] AGREE (Appraisal of Guidelines for Research and Evaluation) for clinical guidelines,[17] PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for Cochrane-style systematic reviews[18] and SQUIRE (Standards for Quality Improvement Reporting Excellence) for quality improvement studies.[19] They have two main purposes: they help researchers design and undertake robust studies, and they help reviewers and potential users of research outputs assess validity and reliability. This project seeks to produce a set of quality criteria and comparable publication standards for realist and meta-narrative reviews.

What are realist and meta-narrative reviews?

Realist and meta-narrative reviews are systematic, theory-driven interpretative techniques, which were developed to help make sense of heterogeneous evidence about complex interventions applied in diverse contexts in a way that informs policy. Interventions have been described as theory incarnate,[20] driven by hypotheses, hunches, conjectures and aspirations about individual and social betterment. Strengthening a review process that helps to sift and sort these theories may be an important step in producing better interventions.

Realist reviews seek to unpack the relationships between context, mechanism and outcomes (sometimes abbreviated as CMO), i.e. how particular contexts have triggered (or interfered with) mechanisms to generate the observed outcomes.[14] Its philosophical basis is realism, which assumes the existence of an external reality (a real world) but one that is filtered (i.e. perceived, interpreted and responded to) through human senses, volitions, language and culture. Such human processing initiates a constant process of self-generated change in all social institutions, a vital process that has to be accommodated in evaluating social programmes.

In order to understand how outcomes are generated, the roles of both external reality and human understanding and response need to be incorporated. Realism does this through the concept of mechanisms, whose precise definition is contested but for which a working definition is, '. . . underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest'.[21] Different contexts interact with different mechanisms to make particular outcomes more or less likely; hence, a realist review produces recommendations of the general format 'in situations [X], complex intervention [Y], modified in this way and taking account of these contingencies, may be appropriate'. Realist reviews can be undertaken in parallel with traditional Cochrane reviews (see, for example,

the complementary Cochrane and realist reviews of school feeding programmes in disadvantaged children).[22,23] The Cochrane review produced an estimate of effect size, whereas the realist review addressed why and how school feeding programmes worked, explained examples of when they did not work, and produced practical recommendations for policy-makers.

Meta-narrative reviews were originally developed by Greenhalgh *et al.*[15,24] to try to explain the apparently disparate data encountered in their review of diffusion of innovation in health-care organisations. Core concepts such as diffusion, innovation, adoption and routinisation had been conceptualised and studied very differently by researchers from a wide range of primary disciplines including psychology, sociology, economics, management and even philosophy. While some studies had been framed as the implementation of a complex intervention in a social context (thus lending themselves to a realist analysis), others had not. Preliminary questions needed to be asked, such as 'What exactly did these researchers mean when they used the terms 'diffusion', 'innovation' and so on?'; 'How did they link the different concepts in a theoretical model – either as a context–mechanism–outcome proposition or otherwise?'; and 'What explicit or implicit assumptions were made by different researchers about the nature of reality?'

These questions prompted the development of meta-narrative review, which sought to illuminate the different paradigmatic approaches to a complex topic area by considering how the same topic had been differently conceptualised, theorised and empirically studied by different groups of researchers. Meta-narrative review is particularly suited to topics where there is dissent about the nature of what is being studied and what is the best empirical approach to studying it. For example, Best *et al.*,[25] in a review of knowledge translation and exchange, asked how different research teams had conceptualised the terms knowledge, translation and exchange – and what different theoretical models and empirical approaches had been built on these different conceptualisations. Thus, meta-narrative review potentially offers another strategy to assist policy-makers to understand and interpret a conflicting body of research and, therefore, to use it more effectively in their work.

The need for standards in theory-driven systematic reviews

Realist and meta-narrative approaches can capitalise on and help build common ground between social researchers and policy teams. Many researchers are attracted to these approaches because they allow systematic exploration of how and why complex interventions work. Policy-makers are attracted to them because they are potentially able to answer questions relevant to practical decisions (not merely 'What is the impact of X?' but 'If we invest in X, to which particular sectors should we target it, how might implementation be improved and how might we maximise its impact?'). As interest in such approaches is burgeoning, it is our experience that these approaches are sometimes being applied in ways that are not always true to the core principles set out in previous methodological guidance.[4,14,15,26] Some reviews published under the realist banner are not systematic, not theory driven and/or not consistent with realist philosophy. The meta-narrative label has also been misapplied in reviews which have no systematic methodology. For these reasons, we believe that the time has come to develop formal standards and training materials.

There is a philosophical problem here, however. Realist and meta-narrative approaches are interpretive processes (that is, they are based on building plausible evidenced explanations of observed outcomes, presented predominantly in narrative form); hence, they do not easily lend themselves to a formal procedure for quality checking. Indeed, we have argued previously that the core tasks in such reviews are thinking, reflecting and interpreting.[4,15] In these respects realist and meta-narrative reviews face a problem similar to that encountered in assessing qualitative research, namely the extent to which guidelines, standards and checklists can ever capture the essence of quality. Some qualitative researchers are openly dismissive of the technical checklist approach as an assurance of quality in systematic review.[27] While we acknowledge such views, we believe that from a pragmatic perspective, formal quality criteria – with appropriate caveats – are likely to add to, rather than detract from, the overall quality of outputs in

this field. Scientific discovery is never the mere mechanical application of set procedures.[28] Accordingly, research protocols should aim to guide rather than dictate.

The online Delphi method

This study used the online Delphi method, and in this section we introduce, explain and justify our use of this method. The essence of the Delphi technique is to engender reflection and discussion among a panel of experts with a view to getting as close as possible to a consensus and documenting both the agreements reached and the nature and extent of residual disagreement.[29] It was used, for example, to set the original care standards which formed the basis of the Quality and Outcomes Framework for UK general practitioners.[30] Factors which have been shown to influence quality in the Delphi process include:

(a) composition (expertise, diversity) of the expert panel
(b) selection of background papers and evidence to be discussed by that panel (completeness, validity, representativeness)
(c) adequacy of opportunities to read and reflect (balance between accommodating experts' busy schedules and keeping to study milestones)
(d) qualitative analysis of responses (depth of reflection and scholarship, articulation of key issues)
(e) quantitative analysis of responses (appropriateness and accuracy of statistical analysis, clarity of presentation when this is fed back)
(f) how dissent and ambiguity are treated (e.g. avoidance of groupthink, openness to dissenting voices).[29,31,32]

Evidence suggests that the online medium is more likely to improve than jeopardise the quality of the consensus development process. Mail-only Delphi panels have been shown to be as reliable as face-to-face panels.[33] Asynchronous online communication has well-established benefits in promoting reflection and knowledge construction.[34] There are over 100 empirical examples of successful online Delphi studies conducted between geographically dispersed participants (for examples see Keeney *et al.*,[32] Elwyn *et al.*,[35] Greenhalgh and Wengraf,[36] Hart *et al.*,[37] Holliday and Robotin,[38] and Pye and Greenhalgh[39]). We were unable to find any online Delphi study which identified the communication medium as a significant limitation. On the contrary, many authors described significant advantages of the online approach, especially when dealing with an international sample of experts. One group commented 'our online review process was less costly, quicker and more flexible with regard to reviewer time commitment, because the process could accommodate their individual schedules'.[38]

Critical commentaries on the Delphi process have identified a number of issues which may prove problematic, for example 'issues surrounding problem identification, researcher skills and data presentation'[29] or defining consensus; issues of anonymity; time requirements for data collection, analysis, feedback to participants and obtaining responses on feedback; defining and selecting experts; enhancing response rates and deciding on how many rounds to undertake.[40] These comments suggest that it is the underlying design and rigour of the research process which is key to the quality of the study and not the medium through which this process happens.

Chapter 2 Objectives

For this project we set out to:

1. collate and summarise the literature on the principles of good practice in realist and meta-narrative reviews, highlighting in particular how and why these differ from conventional forms of systematic review and from each other
2. consider the extent to which these principles have been followed by published and in-progress reviews, thereby identifying how rigour may be lost and how existing principles could be improved
3. produce, in draft form, an explicit and accessible set of methodological guidance and publication standards using an online Delphi method with an interdisciplinary panel of experts from academia and policy
4. produce training materials with learning objectives linked to these steps and standards
5. refine these standards and training materials prospectively on real reviews-in-progress, capturing methodological and other challenges as they arise
6. synthesise expert input, evidence review and real-time problem analysis into more definitive guidance and standards
7. disseminate these guidance and standards to audiences in academia and policy.

Chapter 3 Methods

Overview of methods

We used a range of methods to meet the objectives we set out above. In this section we provide a brief overview of the methods we used and how they related to each other. The following methods sections outline specific aspects of the methods we used in more detail.

To fulfil objectives 1 and 2 we undertook a narrative review of the literature that was supplemented by collating feedback from presentations and workshops. We synthesised our findings into briefing materials [one for realist synthesis (RS) and another for meta-narrative reviews]. We recruited members to two Delphi panels, which had wide representation from researchers, students, policy-makers, theorists and research sponsors. We used the briefing materials to brief the Delphi panel, so they could help us in fulfilling objective 3. For objective 4, we drew not only on our experience in developing and delivering education materials, but also relevant feedback from the Delphi panel, an e-mail list we set up specifically for this project (www.jiscmail.ac.uk/RAMESES), training workshops and the review teams we supported methodologically. To help us refine our publication standards (objective 5) we captured methodological and other challenges that arose within the realist or meta-narrative review teams we provided methodological support to. To produce the definitive publication standards, quality standards and training materials (objective 6), we synthesised expert input (from the Delphi panel), literature review and real-time problem analysis (e.g. feedback from the e-mail list, training sessions and workshops, and presentations). Throughout this project we did not set specific time points when we would refine the drafts of our project outputs. Instead, we iteratively and contemporaneously fed any data we captured into our draft publication standards, quality standards and training materials, making changes gradually. Only our Delphi panels ran within a specific time frame. The definitive guidance and standards were, therefore, the product of continuous refinements. We addressed objective 7 through academic publications, online resources and delivery of presentations and workshops. *Figure 1* provides a pictorial overview of how the different methods we used fed into each other.

Details of literature search methods

Prior to the start of this project we had undertaken initial exploratory searches. These were rapid searches that were not intended to be comprehensive: they involved each project team member identifying in their personal files examples of realist and/or meta-narrative reviews. To identify further reviews, we undertook a search of the reference lists of each retrieved review. This two-step process yielded 13 reviews that were later used by our expert librarian to develop and refine our searches (see *Appendix 1*). We found that the literature in this field was currently small but expanding rapidly, of broad scope, variable quality and inconsistently indexed. Our purpose for identifying published reviews was not to complete a census of realist and meta-narrative studies. We make no claims that the review we undertook was exhaustive, thus we have not and never intended that it should be published as a stand-alone piece of research. The main purpose of this review was to enable us to produce the briefing materials for our two Delphi panels (objective 3), not to produce an exhaustive summary of all research ever published on the topic.
As such, the review we undertook would best be considered as being a rapid, accelerated or truncated narrative review. Such an approach will predictably produce limitations and these are discussed in *Chapter 5, Limitations*.

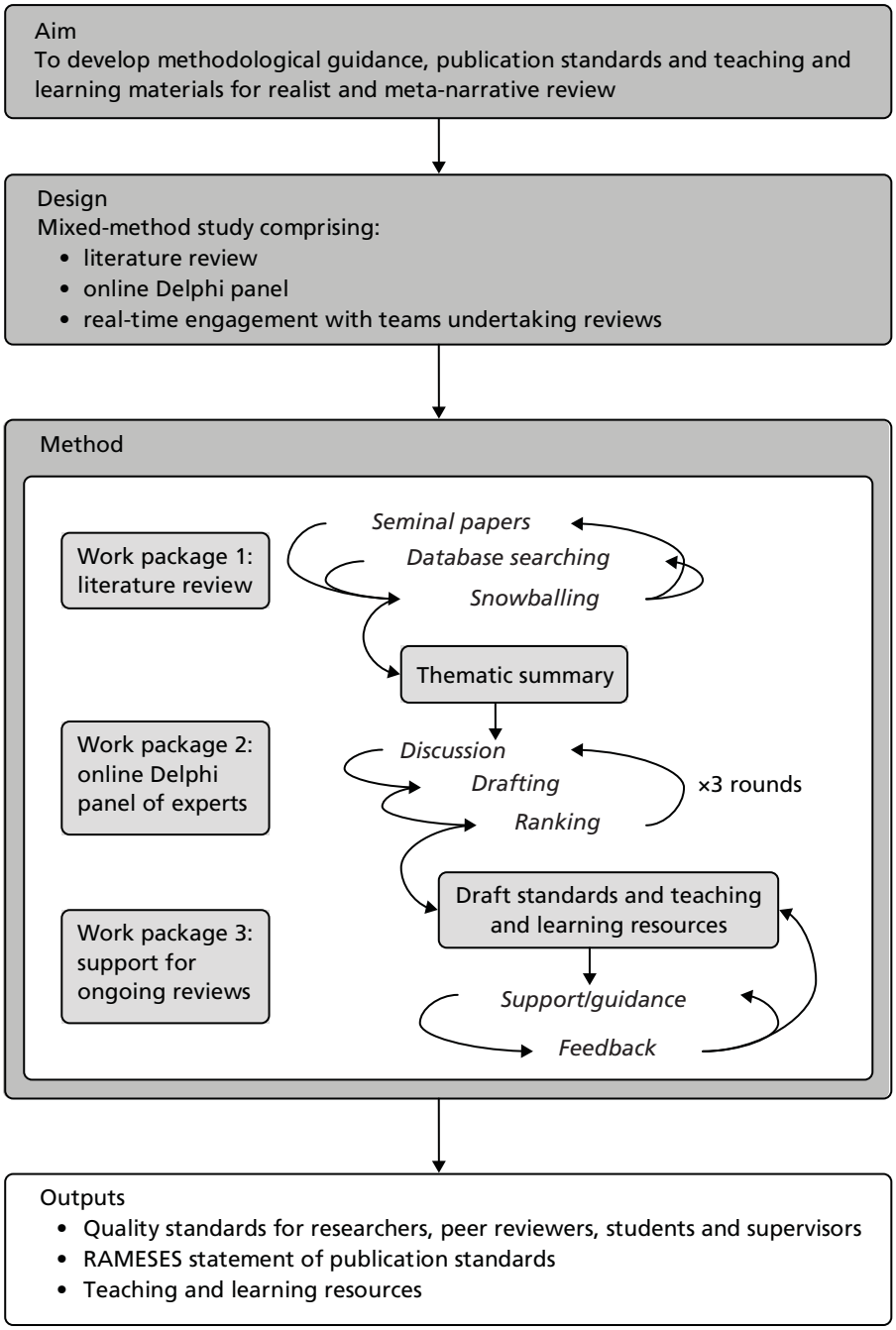


FIGURE 1 Overview of study processes. RAMESES, Realist And Meta-narrative Evidence Syntheses – Evolving Standards.

We wanted our search to allow us to pinpoint real examples (or publications claiming to be examples) that provide rich detail on their usage of those review activities we wish to scrutinise and formalise. To that end, and drawing on a previous study which demonstrated the effectiveness and efficiency of the methods proposed,[41] and employing the skills of a specialist librarian, we searched 16 electronic databases from inception (where applicable) to 15 June 2010. The databases searched are listed below (the number of hits found with each database searched may be found in *Table 1*):

- Academic Search Complete
- Cumulative Index to Nursing and Allied Health Literature (CINAHL)
- The Cochrane Library

- Dissertation Abstracts
- EMBASE
- Education Resources Information Center (ERIC)
- Global Health
- Google
- HealthSTAR
- MEDLINE
- PASCAL [database of INIST (Institut de l'Information Scientifique et Téchnique)]
- PsycINFO
- Scopus
- Sociological Abstracts
- Social Policy and Practice
- Web of Science [Science Citation Index (SCI), Social Science Citation Index (SSCI), Arts and Humanities Citation Index (AHCI)].

TABLE 1 Hits returned for databases searched

Database	Hits returned
Academic Search Complete	69
CINAHL	28
The Cochrane Library	4
Dissertation Abstracts	9
EMBASE	39
ERIC	2
Global Health	10
HeathSTAR	20
MEDLINE	43
PASCAL	6
PsycINFO	14
Scopus	182
Web of Science (SCI and SSCI combined)	190
Sociological Abstracts	0
Social Policy and Practice	0
Google	0
Total retrieved	616
Duplicates	368
Files screened	248

We used the following approaches in our searches:

1. A simple truncated text-word search was conducted on all databases using the following words: (Meta-narrative OR metanarrative OR realist) ADJ (review* OR protocol* OR synthesis OR syntheses OR technic OR technics OR technique*) was used where ADJ (adjacency) was a search operator (Ovid Databases); or (metanarrative OR meta-narrative OR realist) AND (review* OR synthesis OR syntheses OR protocol* OR technic OR technics OR technique*), where adjacency was not a search operator. In the last instance, the search was limited to the title field. The strategy was developed based on a collection of 13 published reviews we had identified in our exploratory searches and was piloted and refined to produce the most sensitive search strategy for the topic.
2. Citation chaining, on databases where this feature was available (Scopus and Web of Science at the time of the search) was performed. Seminal citations were followed, with the reasoning that anyone using realist or narrative techniques would be likely to cite these references.[4,14,15]

Results were kept separate for each database in RefWorks (version 5; RefWorks-COS, Bethesda, MD, USA) reference management software and were then collated into a separate merged file from which duplicates were removed.

No language or study design filters were applied. To construct our sample for further analysis (in which we intended to study both exemplary reviews and those that had flaws), we included any review that claimed to be a realist review or a meta-narrative review. Documents were excluded if they were not a review (e.g. editorials, opinion pieces, commentaries, methods papers) or did not claim to be a realist or meta-narrative review. We did not undertake any independent screening or an audit of a random subset for quality control purposes. The whole searching process from start to the retrieval of all full-text documents took approximately 1 month.

We conducted a thematic analysis of this literature which was initially oriented to addressing seven key areas:

1. What are the strengths and weaknesses of realist and meta-narrative review from both a theoretical and a practical perspective?
2. How have these approaches actually been used? Are there areas where they appear to be particularly fit (or unfit) for purpose?
3. What, broadly, are the characteristics of high- and low-quality reviews undertaken by realist or meta-narrative methods? What can we learn from the best (and worst) examples so far?
4. What challenges have reviewers themselves identified (e.g. in the introduction or discussion sections of their papers) in applying these approaches? Are there systematic gaps between the theory and the steps actually taken?
5. What is the link between realist and meta-narrative review and the policy-making process? How have published reviews been commissioned or sponsored? How have policy-makers been involved in shaping the review? How have they been involved in disseminating and applying its findings? Are there models of good practice (and of approaches to avoid) for academic–policy linkage in this area?
6. How have front-line staff and service users been involved in realist and meta-narrative reviews? If the answer to this is 'usually, not much', how might they have been involved and are there examples of potentially better practice which might be taken forward?
7. How should one choose between realist, meta-narrative and other theory-driven approaches when selecting a review methodology? How might (for example) the review question, purpose and intended audience(s) influence the choice of review method?

The thematic analysis was led by one member of the review team (GWo). He undertook all stages of the review and shared findings with the rest of the project team so that discussion, debate and refinement of his interpretations of the data in the included reviews could take place. Findings were shared by e-mail and, where necessary, face-to-face meetings took place to discuss any interpretations made of the data.

In undertaking our thematic analysis, we familiarised ourselves with the included reviews to identify patterns in the data. We used the questions above, which relate to seven key areas, as a starting point in our sensemaking of the data, and as a project team we were aware that the purpose of the review was to produce briefing documents for the Delphi panels. In these panels we wanted to achieve a consensus on quality and reporting standards, and so what we needed from our review of the literature were data to inform us on what might constitute quality in executing and reporting reviews. We accepted that we might need to refine, discard or add additional questions and topic areas to explore in order to better capture our analysis and understanding of the literature as these emerge from our reading of the papers.

Data were extracted to a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA, USA), which we iteratively refined to capture the data needed to produce our briefing materials. This review was undertaken in a short timeframe, such that the time taken from obtaining full-text documents to producing the final draft for circulation of the briefing documents was approximately 10 weeks. The output of this phase was a provisional summary for each review method that addressed the questions above and highlighted for each question the key areas of knowledge, ignorance, ambiguity and uncertainty. This was distributed to the Delphi panel (as our briefing document) as the starting point for their guidance development work.

Details of online Delphi process

We followed an online adaptation of the Delphi method (see *Chapter 1, The online Delphi method*) which we had developed and used in a previous study to produce guidance on how to critically appraise research on illness narratives.[36] In that study, a key component of a successful Delphi process was recruiting a wide range of experts, policy-makers, practitioners and potential users of the guidance who could approach the problem from different angles and, especially, people who would respond to academic suggestions by asking 'so what?' questions.

Placing the academic–policy/practice tension central to this phase of the research, we planned to construct our Delphi panel to include a majority of experienced academics (e.g. those who have published on theory and method in realist and/or meta-narrative review). We also planned to recruit policy-makers, research sponsors and representatives of third-sector organisations. These individuals were recruited by approaching relevant organisations and e-mail lists [e.g. professional networks of systematic reviewers, CHAIN (Contact, Help, Advice and Information Network; http://chain.ulcc.ac.uk) and INVOLVE (www.invo.org.uk/)]. We approached INVOLVE as we were interested in exploring if we could identify a lay person who might have interest in secondary research and/or informing policy/decision-making through reviews. Those interested in participating were provided with an outline of the study and individuals who indicated greatest commitment and potential to balance the sample were selected. We drew on our own experience of developing standards and guidance, as well as on published papers by CONSORT, PRISMA, SQUIRE, AGREE and other teams working on comparable projects.[16,18,19,42]

The Delphi panel was conducted entirely via the internet using a combination of e-mail and an online survey tool (www.surveymonkey.com). We began with a brainstorm round (round 1) in which participants were invited to submit personal views, exchange theoretical and empirical papers on the topic and suggest items that might could be included in the publication standards. This was done as a warm-up exercise and panel members were sent our own preliminary summary or briefing document (see *Chapter 3, Details of literature search methods*). These early contributions, along with our summary, were collated and summarised in a set of provisional items, which were developed into an online survey and sent electronically (via the online survey tool, SurveyMonkey®; Survey Monkey, Palo Alto, CA, USA) to participants for ranking (round 2). Participants were asked to rank each item twice on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree), once for relevance (i.e. should an item on this theme/topic be included at all in the guidance?) and once for validity (i.e. to what extent do you agree with this item as currently worded?). Those who agreed that an item was relevant, but disagreed on its wording,

were invited to suggest changes to the wording via a free-text comments box. In this second round, participants were again invited to suggest additional topic areas and items.

Each participant's responses were collated and the numerical rankings entered onto an Excel spreadsheet. The response rate, average, mode, median and interquartile range for each participant's response to each item were calculated. Items that score low on relevance were omitted from subsequent rounds. Further online discussion was invited on items that score high on relevance but low on validity (indicating that a rephrased version of the item was needed) and on those where there was wide disagreement about relevance or validity. Following analysis and discussion within the project team, a second list of statements was drawn up and circulated for ranking (round 3). We planned that the process of collation of responses, further e-mail discussion, and reranking would be repeated until a maximum consensus is reached (round 4 et seq.). In practice, very few Delphi panels, online or face to face, go beyond three rounds as participants tend to 'agree to differ' rather than move towards further consensus.[36]

We had planned to report residual non-consensus as such and the nature of the dissent described. Making such dissent explicit tends to expose inherent ambiguities (which may be philosophical or practical) and acknowledges that not everything can be resolved; such findings may be more use to reviewers than a firm statement that implies that all tensions have been fixed.

Preparing teaching and learning resources

A key aim of our project was to produce publicly accessible resources to support training in realist and meta-narrative review. We anticipate that these resources will need to be adapted and perhaps supplemented for different groups of learners, and interactive learning activities added.[43] We developed, and iteratively refined, draft learning objectives, example course materials and teaching and learning support methods. We drew on our previous work on course development, quality assurance and support for interactive and peer-supported learning in health-care professionals for this aspect of our project.[34,43–45]

Real-time refinement

The sponsor of this study, the National Institute for Health Research (NIHR) Health Services and Delivery Research (HSDR) programme, supports secondary research calls for rapid, policy-relevant reviews, some, though not all, of which seek to use realist or meta-narrative methods. We were asked to work with a select sample of teams funded under such calls, as well as other teams engaged in relevant ongoing reviews (selected to balance our sample), to share emerging recommendations and gather real-time data on how feasible and appropriate these recommendations are in a range of different reviews. Over the 27-month duration of this study, we used the feedback we gathered to iteratively refine our draft training materials. Training and support offered to these review teams consisted of three overlapping and complementary packages:

1. An all-comers online discussion forum via JISCmail (www.jiscmail.ac.uk/RAMESES) for interested reviewers who were doing or had previously attempted a realist or meta-narrative review. This was run via light-touch facilitation in which we invited discussion on particular topics and periodically summarise themes and conclusions (a technique known in online teaching as weaving). Such a format typically accommodates large numbers of participants since most people tend to lurk most of the time. Such discussion groups tend to generate peer support through their informal, non-compulsory ethos and a strong sense of reciprocity (i.e. people helping one another out because they share an identity and commitment)[46] and they are often rich sources of qualitative data. We anticipated that this forum would contribute key themes and ideas to the quality and reporting standards and learning materials throughout the duration of the study.

2. Responsive support to our designated review teams. We anticipated that our input to these teams would depend on their needs, interests and previous experience. In our previous dealings with review teams we have been called upon (for example) to assist them in distinguishing context from mechanism in a particular paper, extracting and formalising programme theories, distinguish middle-range theories from macro or micro theories, develop or adapt data extraction tools, advise on data extraction techniques and train researchers in the use of qualitative software for systematic review.
3. A series of workshops for designated review teams and other reviewers. We planned to run a series of workshops both to provide training to fellow reviewers interested in using realist or meta-narrative reviews, but also to get feedback from them about what challenges they faced either learning about or undertaking such reviews.

Chapter 4 Results

In this project we produced three specific outputs for realist and meta-narrative reviews:

1. publication standards
2. quality standards
3. teaching and learning materials (also known as training materials).

We used a range of methods to gather the data that informed the content of each of our intended outputs. This section provides details of the results we obtained from the methods we used and how they contributed to the content of our outputs.

Literature search

Sixteen electronic databases were searched from inception to June 2011 and citation tracking was undertaken generating 248 documents. A flow diagram outlining the disposition of documents can be seen in *Figure 2*. *Table 1* shows the number of hits returned for the databases we searched.

One of the project team (GWo) screened the abstracts and titles and included documents which claimed to be realist or meta-narrative reviews. All the documents judged to be possible realist and meta-narrative reviews were circulated to all the project team and through discussion and debate a final set of included documents were retained. We retrieved what we judged to be 35 possible realist reviews and nine meta-narrative reviews. For the possible realist reviews there was no disagreement between the project team as to inclusion (35 out of 35 were included for analysis). Out of the 11 possible meta-narrative reviews, two were judged not to be meta-narrative reviews, leaving nine documents. *Tables 2* and *3* show characteristics of the documents (review title, type of document, year published and topic area) that we drew on for realist reviews and meta-narrative reviews respectively. We conducted a thematic analysis guided initially by the seven questions set out above (see *Chapter 3, Details of literature search methods*) to produce the briefing documents for the realist and meta-narrative Delphi panels (see *Appendix 2*). All the data we extracted were either entered into an Excel spreadsheet or written up directly into a draft of our briefing documents.

Our briefing documents were based on our thematic analysis which was guided by seven initial key areas (see *Chapter 3, Details of literature search methods* for a list of the key areas). We needed differing levels of immersion and analysis for each of the items we included in our briefing documents. Some were more straightforward to derive from our initial questions and our reading of the literature. We noted that three out of the initial seven questions [(1) What are the strengths and weaknesses of realist and meta-narrative review from both a theoretical and a practical perspective?; (2) How have these approaches actually been used? Are there areas where they appear to be particularly fit (or unfit) for purpose?; and (7) How should one choose between realist, meta-narrative and other theory-driven approaches when selecting a review methodology? How might (for example) the review question, purpose and intended audience(s) influence the choice of review method?] had overlaps and could be collapsed into one topic area for consideration by our Delphi panels. We judged that questions 1, 2 and 7 were related to matching the research question to the method. We noted that in our included reviews, most researchers had also considered this an important topic to address – often through an explanation of why they had deliberately chosen either a realist or meta-narrative review. We therefore included this issue as items 6 and 5 (for meta-narrative and realist reviews respectively) in our briefing document for our Delphi panel. These items asked the Delphi panel members to clarify what a research question would look like in a meta-narrative or realist review.

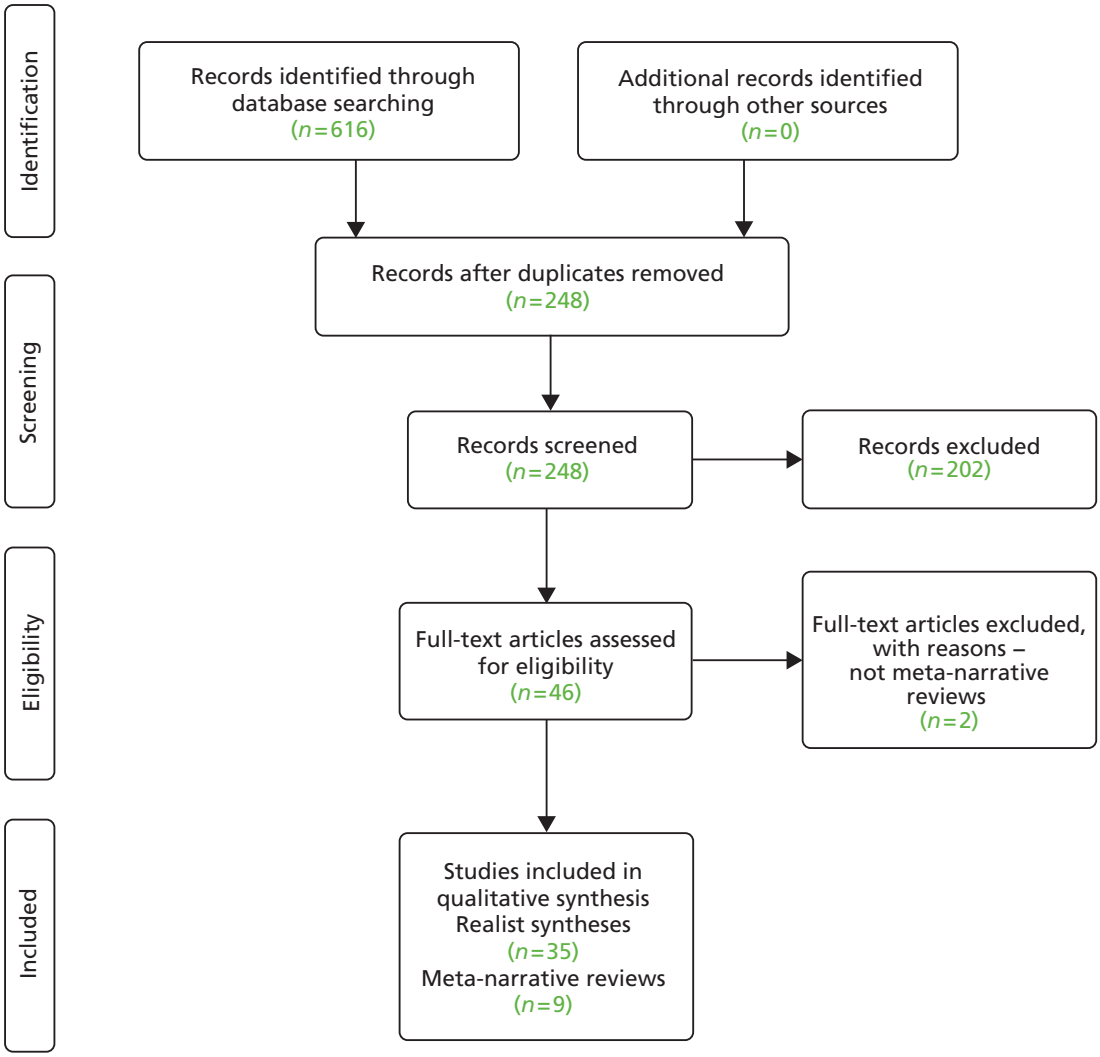


FIGURE 2 Flow diagram outlining the disposition of documents.

When doing realist and meta-narrative reviews ourselves, we had previously noted that such reviews often covered broad topics and needed to be progressively focused. Two of our initial questions related to these issues: (5) What is the link between realist and meta-narrative review and the policy-making process? How have published reviews been commissioned or sponsored? How have policy-makers been involved in shaping the review? How have they been involved in disseminating and applying its findings? Are there models of good practice (and of approaches to avoid) for academic-policy linkage in this area?; and (6) How have front-line staff and service users been involved in realist and meta-narrative reviews? If the answer to this is 'usually, not much', how might they have been involved and are there examples of potentially better practice which might be taken forward? We had asked questions 5 and 6 to ascertain if other researchers had noted this as an important process and, if they had, what approaches had they used. Within our included reviews, the breadth of the initial topic areas had been identified as a challenge and a range of different approaches had been used to focus reviews. The issue of the need to focus reviews thus seemed to us an important one to include in our briefing documents (as items 9 and 8 for meta-narrative and realist reviews respectively). Items 9 and 8 in our briefing documents asked the Delphi panel members to consider if it was important for researchers to explain how and why their review was shaped and contained.

TABLE 2 Characteristics of realist review documents retrieved from literature review (listed by year of publication)

Review title	Type of document	Year published	Topic area
Vocational rehabilitation: what works and in what circumstances[47]	Journal article	2004	Vocational rehabilitation
A systematic review of controlled trials of interventions to prevent childhood obesity and overweight: a realistic synthesis of the evidence[48]	Journal article	2006	Interventions to reduce childhood obesity and overweight
A realist synthesis of evidence relating to practice development: Final report to NHS Education for Scotland and NHS Quality Improvement Scotland[49]	Full report	2006	Practice development
[a]Realist review to understand the efficacy of school feeding programmes[22]	Journal article	2007	Efficacy of school feeding programmes
Evaluating the impact of patient and public involvement initiatives on UK health services: a systematic review[50]	Journal article	2007	Patient and public involvement
Marketing mix standardization in multinational corporations: a review of the evidence[51]	Journal article	2007	Marketing mix standardisation in multinational corporations
Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review[52]	Journal article	2008	Human resource management interventions to improve health workers' performance
Does moving from a high-poverty to lower-poverty neighborhood improve mental health? A realist review of 'Moving to Opportunity'[53]	Journal article	2008	US Moving to Opportunity programme
Primary health care delivery models in rural and remote Australia – a systematic review[54]	Journal article	2008	Primary health-care delivery models in rural and remote Australia
Independent learning literature review (research report DCSF-RR051)[55]	Report	2008	Independent learning in school children
A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries[56]	Journal article	2009	Community health workers
Community-based services for homeless adults experiencing concurrent mental health and substance use disorders: a realist approach to synthesising evidence[57]	Journal article	2009	Community-based services for homeless adults with concurrent mental health and substance use disorders CDs
Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries[58]	Report	2009	Water, sanitation and hygiene interventions in reducing childhood diarrhoea
[a]Internet-based medical education: a realist review of what works, for whom and in what circumstances[43]	Journal article	2009	Internet-based learning
Interventions to promote social cohesion in sub-Saharan Africa[59]	Full report	2010	Interventions to promote social cohesion in subSaharan Africa
Implementation of antiretroviral therapy adherence interventions: a realist synthesis of evidence[60]	Journal article	2010	Antiretroviral adherence interventions
Lean thinking in healthcare: a realist review of the literature[61]	Journal article	2010	Lean thinking
District nurses' role in palliative care provision: a realist review[62]	Journal article	2010	Role of district nurses in palliative care provision

continued

TABLE 2 Characteristics of realist review documents retrieved from literature review (listed by year of publication) (*continued*)

Review title	Type of document	Year published	Topic area
A realist review of evidence to guide targeted approaches to HIV/AIDS prevention among immigrants living in high-income countries[63]	PhD thesis	2010	Evidence to guide targeted approaches to HIV infection or AIDS prevention among immigrants living in high-income countries
Effectiveness of telemedicine: a systematic review of reviews[64]	Journal article	2010	Effectiveness of telemedicine
How equitable are colorectal cancer screening programs which include FOBTs? A review of qualitative and quantitative studies[65]	Journal article	2010	Equitability of colorectal screening programmes
Evidence-based health policy: a preliminary systematic review[66]	Journal article	2010	Evidence-based health policy
Behavioral caregiving for adults with traumatic brain injury living in nursing homes: developing a practice model[67]	Journal article	2010	Behavioural caregiving for adults with traumatic brain injury living in nursing homes
Addressing locational disadvantage effectively[68]	Journal article	2010	Addressing locational disadvantage
Realist review and synthesis of retention studies for health workers in rural and remote areas[69]	Report	2011	Access to health workers in rural and remote areas
[a]Policy guidance on threats to legislative interventions in public health: a realist synthesis[70]	Journal article	2011	Public health legislation
Implementing successful intimate partner violence screening programs in health care settings: evidence generated from a realist-informed systematic review[71]	Journal article	2011	Intimate partner violence
An evidence synthesis of qualitative and quantitative research on component intervention techniques, effectiveness, cost-effectiveness, equity and acceptability of different versions of health-related lifestyle advisor role in improving health[72]	Report	2011	Health-related lifestyle advisors
The gradient in health inequalities among families and children: a review of evaluation frameworks[73]	Journal article	2011	Health inequalities among families and children
Effectiveness of the geriatric day hospital – a realist review[74]	Journal article	2011	Effectiveness of geriatric day hospital
Are journal clubs effective in supporting evidence-based decision making? A systematic review. BEME Guide No.16[75]	Journal article	2011	Effectiveness of journal club in supporting evidence-based decision-making
Conducting a realist review of a complex concept in the pharmacy practice literature: methodological issues[76]	Journal article	2011	Culture in community pharmacy organisations
Getting inside acupuncture trials – exploring intervention theory and rationale[77]	Journal article	2011	Acupuncture
Unleashing their potential: a critical realist scoping review of the influence of dogs on physical activity for dog-owners and non-owners[78]	Journal article	2011	Influence of dogs on physical activity for dog- and non-owners
Social networks, social capital and chronic illness self-management: a realist review[79]	Journal article	2011	Social networks, social capital and chronic illness self-management

AIDS, acquired immunodeficiency syndrome; BEME, Best Evidence Medical Education; FOBT, faecal occult blood test; HIV, human immunodeficiency virus.
a Reviews in which one or more project team members were involved.

TABLE 3 Characteristics of meta-narrative review documents retrieved from literature review (listed by year of publication)

Review title	Type of document	Year published	Topic area
[a]Diffusion of innovations in service organisations: systematic literature review and recommendations for future research[24]	Journal article	2004	Diffusion of innovations
Environmental health and vulnerable populations in Canada: mapping an integrated equity-focused research agenda[80]	Journal article	2008	Environmental health and vulnerable populations
Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method[81]	Journal article	2009	Electronic health record
The health, social care and housing needs of lesbian, gay, bisexual and transgender older people: a review of the literature[82]	Journal article	2009	Health, social care and housing needs of lesbian, gay, bisexual and transgender older people
The role of urban municipal governments in reducing health inequities: a meta-narrative mapping analysis[83]	Journal article	2010	Municipal urban governments in reducing health inequalities
Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature[84]	Journal article	2010	Knowledge exchange processes in organisational policy arenas
[a]Measuring quality in the therapeutic relationship – parts 1 and 2[85,86]	Journal article	2010	Measuring quality in therapeutic relationships
Defining the fundamentals of care[87]	Journal article	2010	Defining the fundamentals of nursing care
How can we improve guideline use? A conceptual framework of implementability[88]	Journal article	2011	Improving guideline use

CD, concurrent disorder.
a Reviews in which one or more project team members were involved.

Question 3 [What, broadly, are the characteristics of high- and low-quality reviews undertaken by realist or meta-narrative methods? What can we learn from the best (and worst) examples so far?] of our initial questions required the most immersion and analysis. With this question we had wanted to understand what processes a review team had to undertake to produce a high-quality review. As a project team we had our own ideas but wanted to explore if these were reflected in our reading of the included reviews. We first had to decide if we could agree among ourselves on which of the included reviews were high, mixed or low quality. To do this, each review was read in detail (GWo) and the characteristics of each review that determined its quality were extracted into an Excel spreadsheet. The headings on this spreadsheet were: study name; type of document; year submitted; topic area; purpose of review; understood method?; methodological comments; lessons for methods; methods for reporting; and challenges reported by reviewers and notes.

Once completed (one for realist reviews and another for meta-narrative reviews), the spreadsheet and the full-text documents were circulated to the rest of the project team and through e-mail discussion and debate, a consensus was achieved. The next process was then to reread each of the included reviews to determine which review processes were necessary to lead to a high-quality review. Again, this was led by GWo and each review process was added to a draft of the briefing documents. These drafts were circulated to the rest of the project team and a consensus achieved through discussion and debate. The briefing materials were the result of seven rounds of revisions.

The contents of our briefing materials were as follows:

- an explanation of how we would like the Delphi panel members to contribute
- background to the review methods
- methodological issues we identified for each method
- a summary of the published examples
- our preliminary thoughts on what might be included as publication standards items.

The complete briefing document circulated to the Delphi panels for realist reviews and meta-narrative reviews can be found in *Appendix 2*.

Delphi panel

Realist review

We ran the realist review Delphi panels between September 2011 and March 2012. We recruited 37 individuals from 27 organisations in six countries. These comprised researchers in: public or population health (8); evidence synthesis (6); health services research (8); international development (2); and education (2). We also recruited experts in research methodology (6), publishing (1), nursing (2) and policy and decision-making (2). We started round 1 in mid-September 2011 and circulated the briefing document to the panel. We sent two chasing e-mails to all panel members, and within 4 weeks all panel members who indicated that they wanted to provide comments had done so. Twenty-two Delphi panel members provided suggestions of items that should be included in the publication standards. We used the suggestions from the panel members and the briefing document as the basis of the online survey for round 2. Round 2 started at the end of November 2011 and ran until early January 2012. Panel members were invited to complete our online survey and asked to rate each potential item for relevance and clarity. A copy of this survey can be found in *Appendix 3*. Two reminder e-mails were sent to the panel members. Once the panel had completed their survey we analysed their ratings for relevance and clarity (*Table 4*).

From round 2, only three items did not achieve a consensus: items 5, 9 and 13 (see *Table 4*). Based on the suggestions made by the panel members we refined the text for these items. We had asked panel members if they had a preference between the terms realist review or RS. Fourteen (39%) preferred RS, 10 (28%) realist review and 12 (33%) had no preference. Our conclusion was that the terms RS and realist review are synonymous. We also produced a post-round briefing document from round 2, which detailed for each item:

- the response rate
- mode
- median
- interquartile range
- the action we took for each item based on the panel's ratings
- an anonymised list of all the free-text comments made.

For round 3, we only asked the panel to consider again the items for which a consensus had not been reached in round 2, namely items 5, 9 and 13. We produced an online survey for round 3 and again asked to rate items 5, 9 and 13 for relevance and clarity. To keep the panel updated we provided them with our post-round briefing document from round 2 (available on request from authors). Round 3 ran from mid-February to mid-March 2012. A copy of this survey can be found in *Appendix 4*. Two reminder e-mails were sent to the panel members. Once the panel had completed their survey we analysed their ratings for relevance and clarity (*Table 5*).

TABLE 4 Summary of results for round 2 Delphi panel for realist review

	Relevance				Content			
Item	Response rate (%)	Mode	Median	Interquartile range	Response rate (%)	Mode	Median	Interquartile range
Title	33/37 (89)	7	7	6–7	31/37 (84)	7	7	6–7
Abstract	34/37 (92)	7	7	7–7	34/37 (92)	7	6.5	5–7
Rationale for review	37/37 (95)	7	7	6–7	35/37 (95)	7	7	5–7
Objective and focus of review	33/37 (89)	7	7	6–7	33/37 (89)	7	7	6–7
Changes in review process[a]	35/37 (95)	7	6	5–7	34/37 (92)	7	5.5	3–6.75
Rationale for using realist review	34/37 (92)	7	7	6–7	33/37 (89)	7	6	5–7
Scoping the literature	35/37 (95)	7	7	5.5–7	37/37 (92)	7	6	5–7
Searching process	34/37 (92)	7	7	6–7	34/37 (92)	7	6	5–7
Selection and appraisal of documents[a]	35/37 (95)	7	7	6–7	35/37 (95)	7	6	4.5–7
Data extraction	34/37 (92)	7	7	6–7	33/37 (89)	7	6	5–7
Analysis and synthesis processes	35/37 (95)	7	7	6–7	35/37 (95)	7	6	5–7
Document flow diagram	35/37 (95)	7	6	6–7	35/37 (95)	7	6	5–7
Document characteristics[a]	35/37 (95)	7	6	5–7	35/37 (95)	7	6	4.5–7
Main findings	34/37 (92)	7	7	6–7	31/37 (84)	7	6.5	5–7
Summary of findings	35/37 (95)	7	7	7–7	34/37 (95)	7	7	6–7
Strength, limitations and future research directions	35/37 (95)	7	7	6–7	35/37 (95)	7	6	6–7
Comparison with existing literature	35/37 (95)	7	6	5–7	35/37 (95)	7	6	5–7
Conclusion and Recommendations	34/37 (92)	7	7	6–7	34/37 (92)	7	6.5	6–7
Funding	35/37 (95)	7	7	7–7	35/37 (95)	7	7	6–7

a Item returned to Delphi panel round 3.

TABLE 5 Summary of results for round 3 Delphi panel for realist review

	Relevance				Content			
Item	Response rate (%)	Mode	Median	Interquartile range	Response rate (%)	Mode	Median	Interquartile range
5. Changes in review process	34/37 (92)	7	7	6–7	34/37 (92)	7	6	6–7
9. Selection and appraisal of documents	33/37 (89)	7	7	6–7	33/37 (89)	7	7	6–7
13. Document characteristics	33/37 (89)	7	7	6–7	33/37 (89)	7	6	6–7

By the end of round 3 a consensus was reached on all items. We produced a post-round briefing document from round 3 and circulated this to all our panel members for the sake of completeness (available on request from authors).

Using the data we gathered from the three rounds of the Delphi panel for realist reviews, we produced a final set of items to be included in the publication standards for realist reviews. These were published simultaneously in January 2013 in *BMC Medicine*[89] and the *Journal of Advanced Nursing*.[90] Our publication standards have also been accepted and listed on the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network (a resource centre for good reporting of health research studies; www.equator-network.org).

Meta-narrative review

We ran the meta-narrative review Delphi panels between September 2011 and March 2012. We recruited 33 individuals from 25 organisations in six countries. These comprised researchers in public or population health researchers (five); evidence synthesis (five); health services research (eight); international development (two); education (two); and also research methodologists (six), publishing (one), nursing (two) and policy and decision-making (two). We started round 1 in mid-September 2011 and circulated the briefing document to the panel. We sent two chasing e-mails to all panel members and within 4 weeks all panel members who indicated that they wanted to provide comments had done so. Twenty-two Delphi panel members provided suggestions of items that should be included in the publication standards. One of these items, on whether or not the concept of epistemic tradition should be included in a meta-narrative review, caused a degree of disagreement within the project team. As a result, we specifically put this issue to the Delphi panel. We used the suggestions from the panel members and the briefing document as the basis of the online survey for round 2. Round 2 started at the end of November 2011 and ran until early January 2012. Panel members were invited to complete our online survey and asked to rate each potential item for relevance and clarity. A copy of this survey can be found in *Appendix 5*. Two reminder e-mails were sent to the panel members. Once the panel had completed their survey we analysed their ratings for relevance and clarity (*Table 6*).

From round 2, only three items did not achieve a consensus: items 6 and 13. Item 5 had reached a consensus on relevance and content on the numerical scores, but there were sufficient concerns raised in the free text that we felt it needed to be amended and returned to the panel (see *Table 6*). Based on the suggestions made by the panel members, we refined the text for these items. We had asked panel members if they had a preference between the terms meta-narrative review or meta-narrative synthesis. Thirteen (41%) preferred meta-narrative synthesis, six (18%) meta-narrative review and 13 (41%) had no preference. Our conclusion was that the terms meta-narrative synthesis and meta-narrative review are synonymous. In response to the question of whether or not epistemic tradition should be included in a meta-narrative review, 16 (60%) agreed that it should be, four (15%) disagreed and seven (26%) did not know. As a result, we decided that epistemic tradition should be included in meta-narrative reviews and was incorporated into item 6. We also produced a post-round briefing document from round 2, which detailed for each item:

- the response rate
- mode
- median
- interquartile range
- the action we took for each item based on the panel's ratings
- an anonymised list of all the free-text comments made.

For round 3, we only asked the panel to consider again the items for which a consensus had not been reached in round 2, namely items 5, 6 and 13. Two additional individuals who had initially decided not to respond to round 2 agreed to provide ratings. To ensure consistency they were briefed on the process and results from round 2. We produced an online survey for round 3 and again asked to rate

TABLE 6 Summary of results for round 2 Delphi panel for meta-narrative review

Item	Relevance				Content			
	Response rate (%)	Mode	Median	Interquartile range	Response rate (%)	Mode	Median	Interquartile range
Title	31/33 (94)	7	7	6–7	31/33 (94)	7	7	6–7
Abstract	32/33 (97)	7	7	6–7	32/33 (97)	7	6.5	6–7
Rationale for review	32/37 (97)	7	7	6–7	32/33 (97)	7	7	6–7
Objectives and focus of review	32/33 (97)	7	7	6–7	32/33 (97)	7	7	6–7
Changes in the review process[a]	31/33 (94)	7	7	6–7	31/33 (94)	7	6	6–7
Rationale or using the meta-narrative approach[a]	27/33 (82)	7	7	6–7	27/33 (82)	7	6	5–7
Evidence of adherence to guiding principles of meta-narrative review	31/33 (94)	7	6	5–7	31/33 (94)	7	6	5–7
Scoping the literature	30/33 (91)	7	7	6–7	30/33 (91)	7	7	6–7
Searching processes	31/33 (94)	7	7	6–7	31/33 (94)	7	7	6–7
Selection and appraisal of documents	31/33 (94)	7	7	6–7	31/33 (94)	7	6	6–7
Data extraction	30/33 (91)	7	6	6–7	30/33 (91)	7	6	5–7
Analysis and synthesis processes	31/33 (94)	7	6	6–7	31/33 (94)	6	6	5.5–7
Document flow diagram	31/33 (94)	7	7	5–7	31/33 (94)	7	6	4.5–7
Document characteristics	31/33 (94)	7	6	5–7	30/33 (91)	7	6	5–7
Main findings	31/33 (94)	7	7	6–7	31/33 (94)	7	7	6–7
Summary of findings	31/33 (94)	7	7	6–7	30/33 (91)	7	7	6–7
Strengths, limitations and future research directions	30/33 (91)	7	7	6–7	30/33 (91)	7	7	6–7
Comparison with existing literature	30/33 (94)	7	6	5–7	30/33 (94)	7	6	5–7
Conclusion and recommendations	31/33 (94)	7	7	6–7	31/33 (94)	7	7	6–7
Funding	29/33 (88)	7	7	6–7	29/33 (88)	7	7	6–7

a Item returned to Delphi panel round 3.

items 5, 6 and 13 for relevance and clarity. To keep the panel updated we provided them with our post-round briefing document from round 2 (available on request from authors). Round 3 ran from mid-February to mid-March 2012. A copy of this survey can be found in *Appendix 6*. Two reminder e-mails were sent to the panel members. Once the panel had completed their survey we analysed their ratings for relevance and clarity (*Table 7*).

By the end of round 3 a consensus was reached on all items. We produced a post-round briefing document from round 3 and circulated this to all our panel members for the sake of completeness (available on request from authors).

TABLE 7 Summary of results for round 3 Delphi panel for meta-narrative review

Item	Relevance				Content			
	Response rate (%)	Mode	Median	Interquartile range	Response rate (%)	Mode	Median	Interquartile range
5. Changes in the review process	29/35 (83)	7	7	6–7	29/35 (83)	7	7	6–7
6. Rationale for using the meta-narrative approach	31/35 (89)	7	7	6–7	31/35 (89)	7	7	6–7
13. Document flow diagram	32/35 (91)	7	7	6–7	31/33 (94)	7	6	6–7

Using the data we gathered from the three rounds of the Delphi panel for realist reviews, we produced a final set of items to be included in the publication standards for realist reviews. These were published simultaneously in January 2013 in *BMC Medicine*[91] and the *Journal of Advanced Nursing*.[92] Our publication standards have also been accepted and listed on the EQUATOR network.

Developing quality standards, teaching and learning resources using real-time refinement

We used a range of sources, in real-time to help us develop and refine our quality standards and teaching and learning resources. The data we used to help us came from the following sources:

- **JISCMail (www.jiscmail.ac.uk/RAMESES).** At the start of the project we set up an e-mail list and membership of this list grew rapidly. As of June 2014, the list has 326 members and there are regular discussions on a range of topics relating to realist and meta-narrative reviews.
- **Methodological support to review teams.** Over the course of this project the project team members have provided differing levels of methodological support to reviewers undertaking realist and meta-narrative reviews. This has ranged from providing answers to questions raised by e-mail or on JISCMail to regular face-to-face meetings. The level of support needed by each team differed considerably depending on each team's initial level of expertise. *Table 8* provides an overview of the projects we provided more in-depth methodological support to and also brief details of the nature of each type of support provided. When providing methodological support we contemporaneously made notes on issues and topics that might be relevant in helping us in this part of the project. An example of the type of records we made of our discussions with a review team we provided methodological support to can be found in *Appendix 7*.

TABLE 8 Overview of reviews the project team provided more in-depth methodological support to

Review title	Research question(s)	Funder	Review type	Type of support provided
Risk models and scores for type 2 diabetes: systematic review	What are the different risk scores for identifying adults at risk of type 2 diabetes and which scores work for whom in what circumstances?	Local primary care trusts/ London Deanery, UK	Realist synthesis linked to systematic review	One of us (TG), as the lead member of this team, provided the following: Training for all other team members on realist review principles Lead researcher on the realist review, undertaking all data extraction, tabulation, analysis and preparation of draft realist section of a mixed Cochrane-style and realist review. One other team member cross-checked this work Writing up the mixed-method review for *British Medical Journal*

TABLE 8 Overview of reviews the project team provided more in-depth methodological support to (*continued*)

Review title	Research question(s)	Funder	Review type	Type of support provided
Uncovering the benefits of participatory research: implications of a realist review for health research and practice	1. What benefits and/or constraints emerge from the collaborative undertaking of health-related research by researchers and those affected by the issues under study and/or those who would apply research results? 2. How can the collaborative research process be theorised and evaluated? 3. How do variations in the programme's context and mechanisms influence the process and outcomes of collaborative health intervention research?	Canadian Institute of Health Research, Canada	Realist synthesis	This novice realist review team was supported in the following ways: • Introduction to realism (bespoke face-to-face teaching – 2 days) • Introduction to realist review method (bespoke face-to-face teaching – 2 days) • Practice with data extraction (bespoke face-to-face teaching – 2 days) • Planning how to start and execute review (bespoke face-to-face teaching, then e-mail and teleconference) • Answering any questions review team had on review processes (e-mail and teleconference) • Answering any questions review team had on data analysis (e-mail and teleconference) • Providing feedback on progress with data extraction (e-mail and teleconference) • Providing feedback with CMO analyses and programme theory development and refinement (e-mail and teleconference) • Preparation of manuscripts for publication (e-mail and teleconference)
Realist review of multicomponent interventions to reduce harms of college binge drinking	What were the underlying theories and assumptions about why these programmes work, and what appear to be the mechanisms and associated contextual influences that led to their intended outcomes?	Dartmouth College, USA	Realist synthesis	This novice realist review team was supported in the following ways: • Introduction to realism (attended workshop run by project team) • Introduction to realist review method (attended workshop run by project team) • Practice with data extraction (e-mail and teleconference) • Planning how to start and execute review (at workshop then e-mail and teleconference) • Answering any questions review team had on review processes (e-mail and teleconference) • Answering any questions review team had on data analysis (e-mail and teleconference) • Providing feedback on progress with data extraction (e-mail and teleconference) • Providing feedback with CMO analyses and programme theory development and refinement (e-mail and teleconference) • Preparation of manuscripts for publication (e-mail and teleconference)

continued

TABLE 8 Overview of reviews the project team provided more in-depth methodological support to (*continued*)

Review title	Research question(s)	Funder	Review type	Type of support provided
How design of places promotes or inhibits mobility of older adults: realist synthesis of 30 years of research	How do characteristics of the environment (place) support mobility and what circumstances appear to facilitate or hinder mobility in older adults?	Centers for Disease Control and Prevention, USA	Realist synthesis	This novice realist review team was supported in the following ways: • Introduction to realism (bespoke face-to-face teaching – 1 day) • Introduction to realist review method (bespoke face-to-face teaching – 1 day) • Practice with data extraction (bespoke face-to-face teaching – 1 day) • Planning how to start and execute review (bespoke face-to-face teaching, then e-mail and teleconference) • Answering any questions review team had on review processes (e-mail and teleconference) • Answering any questions review team had on data analysis (e-mail and teleconference) • Providing feedback on progress with data extraction (e-mail and teleconference) • Providing feedback with CMO analyses and programme theory development and refinement (e-mail and teleconference) • Preparation of manuscripts for publication (e-mail and teleconference)
Systematically synthesizing IMCI implementation: what works for whom and in what circumstances?	1. What were the IMCI interventions in various countries: For which target groups? Over what period of time? With what scope and at what scale (geographical, financial)? What implementation strategies were used? 2. What was the intervention outcome? At what level the outcome was measured? How? Which critical contextual factors influenced IMCI implementation outcomes? Did outcome differ in different contexts? How and why? Which mechanisms were triggered that led to a certain outcome? Which effects had the IMCI implementation strategies on different building blocks of health system and vice versa in different settings?	The Alliance for Health Policy and Systems Research, Switzerland	Realist synthesis	This moderately experienced realist review team was supported in the following ways: • Answering specific questions review team had on data analysis (e-mail and teleconference) – specifically with reference to the nature of mechanisms and the extent to which inference was needed • Answering specific questions on the range of study designs that might be included in a realist review • Providing feedback on progress with data extraction (e-mail and teleconference) • Providing feedback with CMO analyses (e-mail and teleconference)

TABLE 8 Overview of reviews the project team provided more in-depth methodological support to (*continued*)

Review title	Research question(s)	Funder	Review type	Type of support provided
	3. What is the potential to transfer these IMCI implementation strategies to other settings? 4. Which programme theories underlie IMCI implementation strategies and how were these operationalised in different settings?			
Evidence synthesis on the occurrence, causes, consequences, prevention and management of bullying and harassing behaviours to inform decision-making in the NHS	What is known about the occurrence, causes, consequences and management of bullying and inappropriate behaviour in the workplace?	National Institute for Health Research Health Services and Delivery Research programme	Realist synthesis	This novice realist review team was supported in the following ways: • Introduction to realism (attended workshop run by project team, then bespoke face-to-face teaching – 1 day) • Introduction to realist review method (attended workshop run by project team, then bespoke face-to-face teaching – 1 day) • Practice with data extraction (attended workshop run by project team, then bespoke face-to-face teaching – 1 day) • Planning how to start and execute review (workshop, bespoke face-to-face teaching, e-mail, teleconference and two 1-day face-to-face teaching sessions) • Answering any questions review team had on review processes (e-mail and teleconference and two 1-day face-to-face teaching sessions) • Answering any questions review team had on data analysis (e-mail and teleconference and two 1-day face-to-face teaching sessions) • Providing feedback on progress with data extraction (e-mail and teleconference and two 1-day face-to-face teaching sessions) • Providing feedback with CMO analyses and programme theory development and refinement (e-mail and teleconference and two 1-day face-to-face teaching sessions) • Preparation of manuscripts for publication (e-mail and teleconference)
The effective and cost-effective use of intermediate, step-down, hospital at home and other forms of community care as an alternative to	Produce a conceptual framework and summary of the evidence of initiatives that have been designed to provide care closer to home in order to reduce reliance on acute care hospital beds	National Institute for Health Research Health Services and Delivery Research programme	Realist synthesis	This experienced realist review team took part in a 2-day roundtable discussion covering: • methodological issues • practical aspects • extending the scope of realist reviews to economics research

continued

TABLE 8 Overview of reviews the project team provided more in-depth methodological support to (*continued*)

Review title	Research question(s)	Funder	Review type	Type of support provided
acute inpatient care: a realist review				• sustaining the momentum from the RAMESES project
What are the impacts of preschool feeding programmes for disadvantaged young children?	What is the impact of school feeding on growth and educational attainment in preschool children and what explains the successes, failures and partial successes of such programmes	International Initiative for Impact Evaluation (3ie), USA	Realist synthesis	One of us (TG), as the lead on realist review elements for this review, provided the following: • Training for other team members on realist review principles • Co-researcher on the realist elements of the review, undertaking data extraction in parallel with two other researchers, tabulation, analysis and preparation of draft realist section of a mixed Cochrane-style and realist review. Two other team member cross-checked this work Writing up the realist components of the review for publication
Hospital patient safety: a realist analysis	Examine the introduction of three safety interventions (improving leadership, reducing infection rates, and implementing surgical checklists) in seven hospitals	National Institute for Health Research Health Services and Delivery Research programme	Realist synthesis	E-mail advice for team members on practical application of realist review principles
The relevance of complexity concepts and systems thinking to public and population health intervention research: a meta-narrative synthesis	Examine a variety of theoretical frameworks that use the concept of complexity science to help understand the social processes and systems of a constantly evolving environment within which public health interventions have to adapt	Canadian Institute of Health Research, Canada	Meta-narrative review	One of us (TG) was a co-applicant on this study and provided: • training for other team members on meta-narrative review • checking data extraction and synthesis undertaken by other team members • dialogue with all researchers on interpretation and analysis of data Revising drafts of the paper for publication
Mining and aboriginal community health: impacts and interventions	Address the knowledge gap regarding mining impacts on Aboriginal health through a multidisciplinary knowledge synthesis of material from both academic and professional realms as held by Aboriginal communities, mining firms, governments, consultancies and civil society	Canadian Institute of Health Research and Social Sciences and Humanities Research Council, Canada	Meta-narrative review	E-mail advice for team members on practical application of meta-narrative review principles

IMCI, integrated management of childhood illness.

Workshops

We ran a number of methods training workshops during this project and these are listed in *Table 9*. Once again we made contemporaneous notes during these workshops and an example of the notes we made can be found in *Appendix 8*.

Quality standards

The data from the sources above were channelled and collated contemporaneously by GWo and used to initially develop the quality standards for researchers using the realist or meta-narrative method. The initial drafts were circulated within the project team and were iteratively refined for content and clarity. *Box 1* provides an illustration of how we drew on the data sources to produce the quality standards using an example for realist syntheses.

For realist syntheses and meta-narrative reviews we developed two sets of quality standards for each. The two sets have been developed for the following user groups:

1. researchers and peer reviewers using these methods
2. funders/commissioners of research.

Although the core component of the quality standards we have developed are the same for each of the two 'versions' listed above, we have adapted them each in an attempt to make them more focused and useful for the intended users. All the quality standards for realist syntheses and meta-narrative reviews are freely available online.[93]

Quality standards for researchers using the methods and peer reviewers

The quality standards for these user groups are set out using rubrics. By peer reviewers here, we specifically refer to individuals who have been asked to appraise the quality of completed reviews. For each review process that requires a judgement about its quality, we have provided a brief description of why the

TABLE 9 List of workshops

Date	Purpose	Venue
2011 March	Realist review training	Queen Mary, University of London, UK
2011 October	Realist evaluation and review webinar	National Institute for Health Research Health Services and Delivery Research webinar, UK
2011 October	Meta-narrative review training	Queen Mary, University of London, UK
2011 November	Meta-narrative review training	McGill University, Montreal, QC, Canada
2012 March	Realist review training	Karolinska Institute, Stockholm, Sweden
2012 April	Realist review training	University of Leeds, UK
2012 April	Realist review training	University of Sheffield, UK
2012 October	Realist review training	Keele University, UK
2012 October	Plenary: realist synthesis	University of Southern Denmark, Copenhagen, Denmark
2012 November	Realist review training	Queens University Belfast, UK
2012 November	Introduction to realism	Global Health Symposium on Health Systems Research, Beijing, China
2013 March	Realist synthesis and evaluation	Erasmus University, Rotterdam, the Netherlands
2013 April	Realist review training	University of East Anglia, UK
2013 June	Realist review training	University of Leeds, UK

BOX 1 Illustration of the type data we drew on to identify the need for and develop quality standards for realist syntheses

Quality standard: programme theories

Identification of need

As researchers and trainers in RS we had noted that there was some confusion amongst researchers about the nature, need and role of realist programme theory (theories) in realist syntheses. To develop the briefing materials and initial drafts of the reporting standards for realist syntheses, we searched for and analysed a number of published syntheses and noted that our impressions were well founded.

When we ran a 1-day conference in March 2011, the issue of the nature, need and role of realist programme theory (theories) in realist syntheses emerged again. In our Delphi process we encouraged participants to provide free-text comments. These closely reflected the comments we received from our 1-day conference.

Development of the quality criteria

We drew on our content expertise of the topic area and published methodological literature to develop the quality criteria. In addition we found that some of our Delphi panel participants provided us with clear indications that support the criteria we set. For example, we suggested that a RS should develop a programme theory and one that did not was inadequate. Delphi panel participants' free-text comments echoed our suggestion:

> *How could identification of programme theories not be appropriate . . .*
>
> *. . . it cannot be an RS without candidate [programme] theories.*

We were also able to draw on the discussions that took place on JISCMail to find support some of our criteria. For example, under excellent in our suggested quality standards for programme theory, we suggested that: 'The relationship between the programme theory and relevant substantive theory is identified.'

As illustration, a comment from JISCMail that we drew upon to support this criterion was:

> *In a review, one focus[es] first on what is reported but one can – and probably should, in order to produce some added value – reflect the findings and outcomes of the study under review against the theories and/or best practice that already exist.*

process is important and also descriptors of criteria against which a decision about quality might be arrived at. The quality standards for realist syntheses for researchers are set out in *Table 10*, while the quality standards for meta-narrative reviews are presented in *Table 11*.

As an illustrative example to explain the layout of these quality standards, in the quality standard for focusing the reviews (see *Table 10, item 3*) for realist syntheses, this aspect of the review could be judged as being adequate if attempts are made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints. For this aspect of a review to be judged as good we recommend that, as well as fulfilling the criteria for adequate (hence our use of the terms adequate plus), reviews would need to ensure (among others), the focusing process is iterative.

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers

Quality standards for RS (for researchers and peer reviewers)				
1. The research problem				
Realist synthesis is a theory-driven method that is firmly rooted in a realist philosophy of science and places particular emphasis on understanding causation and how causal mechanisms are shaped and constrained by social context. This makes it particularly suitable for reviews of certain topics and question, for example complex social programmes that involve human decisions and actions. A realist research question contains some or all of the elements of what works, how, why, for whom, to what extent and in what circumstances, in what respect and over what duration and applies realist logic to address the question. Above all, realist research seeks to answer the why question. Realist synthesis always has explanatory ambitions. It assumes that programme effectiveness will always be partial and conditional and seeks to improve understanding of the key contributions and caveats				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The research topic is appropriate for a realist approach	The research topic is: • not appropriate for secondary research; and/or • does not require understanding of how and why outcomes are generated	The research topic is appropriate for secondary research. It requires understanding of how and why outcomes are generated and why they vary across contexts	Adequate plus: framing of the research topic reflects a thorough understanding of a realist philosophy of science (generative causation in contexts; mechanisms operating at other levels of reality than the outcomes they generate)	Good plus: there is a coherent argument as to why a realist approach is more appropriate for the topic than other approaches, including other theory-based approaches
The research question is constructed in such a way as to be suitable for a RS	The research question is not structured to reflect the elements of realist explanation. For example, it: • only requires description; and/or • only requires a numerical aggregation of outcomes; and/or • only requires summary of processes; and/or • specifies methods that are inadequate to generate realist understanding (e.g. a thematic analysis of . . .)	The research question includes a focus on how and why the intervention, or programme (or similar classes of interventions or programmes – where relevant) generates its outcomes, and contains at least some of the additional elements, 'for whom, in what contexts, in what respects, to what extent and over what durations'	Adequate plus: the rationale for excluding any elements of 'the realist question' from the research question is explicit the question has a narrow enough focus to be managed within a realist review	Good plus: the research question is a model of clarity and as simple as possible

continued

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers (*continued*)

Quality standards for RS (for researchers and peer reviewers)				
2. Understanding and applying the underpinning principles of realist reviews				
Realist syntheses apply realist philosophy and a realist logic of enquiry. This influences everything from the type of research question to a review's processes (e.g. the construction of a realist programme theory, search, data extraction, analysis and synthesis to recommendations) The key analytic process in realist reviews involves iterative testing and refinement of theoretically based explanations using empirical findings in data sources. The pertinence and effectiveness of each constituent idea is then tested using relevant evidence (qualitative, quantitative, comparative, administrative, and so on) from the primary literature on that class of programmes. In this testing, the ideas within a programme theory are recast and conceptualised in realist terms. Reviewers may draw on any appropriate analytic techniques to undertake this testing				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The review demonstrates understanding and application of realist philosophy and realist logic which underpins a realist analysis	Significant misunderstandings of realist philosophy and/or logic of analysis are evident. Common examples include: • programme/intervention activities or strategies are confused with mechanisms • no attempts are made to uncover mechanisms • outcomes are assumed to be caused by the programme/intervention • relationship(s) between an outcome, its causal mechanism(s) and context(s) are not explained • some theory is provided but this is not explicitly linked to outcome(s)	Some misunderstandings of realist philosophy and/or logic of analysis exist, but the overall approach is consistent enough that a recognisably realist analysis results from the process	The review's assumptions and analytic approach are consistent with a realist philosophy at all stages of the review Where necessary, a realist programme theory is developed and tested	Good plus: review methods, strategies or innovations used to address problems or difficulties within the review are consistent with a realist philosophy of science

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers (*continued*)

Quality standards for RS (for researchers and peer reviewers)				
3. Focusing the review				
Because a realist review may generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it may be necessary to contain a review by progressively focusing both its breadth (how wide an area?) and depth (how much detail?). This important process needs to be considered from the start and may involve iterative rounds of discussion and negotiation with (for example) content experts, funders and/or users. It is typical and legitimate for the review's objectives, question and/or the breadth and depth of the review to evolve as the review progresses				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The review question is sufficiently and appropriately focused	The review question is too broad to be answerable within the time and resources allocated There is no evidence that progressive focusing occurred as the review was undertaken	Attempts are made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints Attempts are documented so that they can be described in publications as appropriate	Adequate plus: the focusing process is iterative. Commissioners of the review are involved in decision-making about focusing decisions made about which avenues are pursued and which are left open for further inquiry are recorded and made available to users of the review	Good plus: the review team draws on external stakeholder expertise to drive the focusing process in order to achieve maximal end-user relevance
4. Constructing and refining a realist programme theory				
Early in the review, the main ideas that went into the making of a class of interventions (the programme theory – which may or may not be realist in nature) are elicited. This initial programme theory sets out how and why a class of intervention is thought to work to generate the outcome(s) of interest. This initial programme theory then needs to be recast in realist terms (a rough outline of the contexts in which, populations for which, and main mechanisms by which, particular outcomes are expected to be achieved). This initial tentative theory will be progressively refined over the course of the review				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
An initial realist programme theory is identified and developed	A realist programme theory is not offered or; a programme theory is offered, but is not converted to a realist programme theory at any stage of the review	An initial programme theory is identified and described in realist terms (that is, in terms of the relationship between contexts, mechanisms and outcomes) The refined theory is consistent with the evidence provided	Adequate plus: the initial realist programme theory is set out at the start and will be refined iteratively as the review team's understanding of the topic grows	Good plus: the relationship between the programme theory and relevant substantive theory is identified implications of the final theory for practice, and for refinements to substantive theory where appropriate, are described the final realist programme theory comprises multiple CMO configurations (describing the ways different mechanisms fire in different contexts to generate different outcomes) and an explanation of the pattern of CMOs

continued

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers (*continued*)

Quality standards for RS (for researchers and peer reviewers)				
5. Developing a search strategy				
Searching in a realist review is guided by the objectives and focus of the review, and revised iteratively in the light of emerging data. Searching is directed at finding data that can be used to test theory, and may lie in a broad range of sources that may cross traditional disciplinary, programme and sector boundaries. The search phase is thus likely to involve searching for different sorts of data, or studies from different domains, with which to test different aspects of any provisional theory				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The search process is such that it would identify data to enable the review team to develop, refine and test programme theory or theories	The search is incapable of supporting a rigorous realist review. Common errors include: • the search is driven by a methodological hierarchy of evidence (e.g. privileging RCTs) rather than the need to identify data to develop, refine or test programme theory/ies • the search process is not informed by the objectives and focus of the review • the database(s) selected are narrow in the subject matter that they contain (e.g. limited to specific topics rather than extending to social science, psychology, etc.) • searching is undertaken once only at the outset of the review and there is no iterative component	Searches are driven by the objectives and focus of the review The search strategy is piloted and refined to check that it is fit for purpose Documents are sought from a wide range of sources which are likely to contain relevant data for theory development, refinement and testing There is no restriction on the study or documentation type that is searched for	Adequate plus: further searches are undertaken in light of greater understanding of the topic area. These searches are designed to find additional data that would enable further theory development, refinement or testing	Good plus: the searching deliberately seeks out data from situations outside the programme under study where it can be reasonably inferred that the same mechanisms(s) might be in operation

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers (*continued*)

Quality standards for RS (for researchers and peer reviewers)				
6. Selection and appraisal of documents				
Realist review requires a series of judgements about the relevance and robustness of particular data for the purposes of answering specific questions within the overall review question An appraisal of the contribution of any section of data (within a document) should be made on two criteria: • *Relevance* – whether it can contribute to theory building and/or testing; and • *Rigour* – whether or not the method used to generate that particular piece of data is credible and trustworthy The selection and appraisal stage may need to run in parallel with the analysis stage				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The selection and appraisal process ensures that sources relevant to the review containing material of sufficient rigour to be included are identified. In particular, the sources identified allow the reviewers to make sense of the topic area; to develop, refine and test theories; and to support inferences about mechanisms	The selection and appraisal process does not support a rigorous and complete realist review. For example: • selection is overly driven by methodological hierarchies (e.g. the restriction of the sources to RCTs to the exclusion of other forms of evidence) • sources are appraised using a technical checklist for a particular method (e.g. assessment of quality for a RCT) rather than by making a defensible judgement on the relevance and rigour of the source • Selection and appraisal processes are overly restrictive and exclude materials that may be useful for a realist analysis • Selection and appraisal processes are not sensitive enough to exclude irrelevant materials	Selection of a document for inclusion into the review is based on what it can contribute to the process of theory development, refinement and/or testing (i.e. relevance) Appraisals of rigour judge the plausibility and coherence of the method used to generate data	Adequate plus: During the appraisal process limitations of the method used to generate data are identified and taken into consideration during analysis and synthesis	Good plus: Selection and appraisal demonstrate sophisticated judgements of relevance and rigour within the domain

continued

TABLE 10 Quality standards for realist reviews for researchers and peer reviewers (*continued*)

Quality standards for RS (for researchers and peer reviewers)				
7. Data extraction				
In a review, data extraction assists analysis and synthesis. Of particular interest to the realist reviewer are data that support the use of realist logic to answer the review's question(s), e.g. data CMO configurations, demiregularities, middle-range and/or programme theories				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The data extraction process captures the necessary data to enable a realist review	The data extraction process does not capture the necessary data to enable a realist review. For example: • data extraction is undertaken mechanically and with no attention to how the data informs the review • no or very limited piloting has been undertaken to test aspects of the data extraction process and improve it	Data extraction focuses on identification and elucidation of CMO configurations and refinement of programme theory Piloting and refinement of the data extraction process has been undertaken where appropriate. Quality control processes are in place to check that all review team members apply common processes and standards in data extraction	Adequate plus: data extraction processes support later processes of analysis (e.g. by organising data into sets relevant for later analysis). The data extracted are comprehensive enough to identify main CMO patterns	Good plus: the data extraction process is continually refined as the review progresses, so as to capture relevant data as the review question is focused and/or programme theory is refined
8. Reporting				
Realist reviews may be reported in multiple formats: lengthy reports, summary reports, articles, websites and so on. Reports should be consistent with the publication standards for RS. (See RAMESES publication standards: Realist syntheses at: http://onlinelibrary.wiley.com/doi/10.1111/jan.12095/full[89], or www.biomedcentral.com/1741-7015/11/21)[89,90]				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The RS is reported using the items listed in the RAMESES reporting standard for realist syntheses	Key items are missing. For example: • no defined research question • limited or no reporting of the review's processes (i.e. methods used) • limited or no explanations and justifications provided for any adaptations made on the realist review process • insufficient detail is reported to enable readers to judge the plausibility and coherence of the findings	Most items reported. In particular the following items should be reported: • Rationale for review • Objectives and focus of review • All method section items (i.e. items 5–11 in the RAMESES publication standards: realist syntheses)[89,90]	All items are reported clearly and in sufficient detail for an external reader to understand and to judge the methods used and the plausibility and coherence of the findings	Good plus: the report is well written and easy to understand. Additional materials are made available for external readers to investigate aspects of the review in more detail

RCT, randomised control trial.

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
1. The research problem				
Meta-narrative review is a relatively new method of systematic review, designed for topics that have been differently conceptualised and studied by different groups of researchers. To understand the many approaches, reviewers have to consciously and reflexively step out of their own world view, learn some new vocabulary and methods, and try to view a topic through multiple different sets of eyes. An overarching narrative of the different perspectives, based on an increased understanding of them, is produced which highlights what different research teams might learn from one another's approaches				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The research topic is appropriate for a meta-narrative approach	The research topic is: • not appropriate for secondary research and/or • does not require understanding of how a topic has been conceptualised and studied differently by different groups	The research topic is appropriate for secondary research. It would benefit from illumination of how a topic has been conceptualised and studied differently by different groups	Adequate plus: • framing of the research topic reflects a thorough understanding of the value, importance and implications of different approaches on research practice and findings	Good plus: • there is a coherent argument as to why a meta-narrative review is more appropriate for the topic than potential alternatives
The research question is constructed in such a way as to be suitable for a meta-narrative review	The research question is not structured to reflect the elements of meta-narrative explanation. For example, it: • requires only description; and/or • requires only a numerical aggregation of outcomes and/or • requires only a summary of processes and/or • specifies methods that are inadequate to generate meta-narrative understanding (e.g. a thematic analysis of . . .)	The research question includes a focus on how a topic has been conceptualised and studied differently by different groups	Adequate plus: • the research question includes an element that addresses the implications of different conceptualisations and approaches to a topic on research findings	Good plus: • the research question is a model of clarity and as simple as possible
2. Understanding and applying the purpose and underpinning principles of meta-narrative reviews				
Meta-narrative review (which is rooted in a constructivist philosophy of science) is inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms. Meta-narrative reviews often look historically at how particular research traditions or epistemic traditions have unfolded over time and shaped the normal science of a topic area The review seeks first to identify and understand as many as possible of the potentially important different research traditions which have a bearing on the topic. In the synthesis phase, by means of an overarching narrative, the findings from these different traditions are compared and contrasted to build a rich picture of the topic area from multiple perspectives. The goal of meta-narrative review is sensemaking of a complex (and perhaps contested) topic area. During analysis and synthesis, six guiding principles (pragmatism, pluralism, historicity, contestation, reflexivity and peer review) should be used				

continued

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers (*continued*)

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The review demonstrates understanding and application of the purpose and principles underpinning a meta-narrative review	Significant misunderstandings of purpose and principles underpinning a meta-narrative review. Common examples include: • analysing only one paradigm/ epistemic tradition • no application of the six underlying principles	Some misunderstandings of purpose and principles underpinning a meta-narrative review, but the overall approach is consistent enough that a recognisable set of distinct meta-narratives together with a higher-order synthesis of these results from the process	The review's assumptions and analytic approach are consistent with the purpose and underpinning principles of a meta-narrative review. In particular, the philosophical position is explicitly constructivist. A sufficient range of paradigms/epistemic traditions has been included to make sense of an unfolding and complex topic area from multiple perspectives and to use contrasts between these as higher-order data	Good plus: • review methods, strategies or innovations used to address problems or difficulties within the review are philosophically coherent and make a clear and illuminative contribution to the knowledge base on the topic area
3. Focusing the review				
A meta-narrative review asks some or all of the following questions: 1. Which research (or epistemic) traditions have considered this broad topic area? 2. How has each tradition conceptualised the topic? 3. What theoretical approaches and methods did they use? 4. What are the main empirical findings? 5. What insights can be drawn by combining and comparing findings from different traditions? Because a meta-narrative review may generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it may be necessary to contain a review by progressively focusing both its breadth (how wide an area?) and depth (how much detail?). This important process needs to be considered from the start and may involve iterative rounds of discussion and negotiation with (for example) content experts, funders and/or users. It is typical and legitimate for the review's objectives, question and/or the breadth and depth of the review to evolve as the review progresses				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The review question is sufficiently and appropriately focused	The review question is too broad to be answerable within the time and resources allocated There is no evidence that progressive focusing occurred as the review was undertaken	Attempts were made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints	Adequate plus: • there is evidence that the focusing process was iterative • commissioners of the review were involved in decision-making about focusing • decisions made about which avenues were pursued and which left open for further inquiry are clearly documented and made available to users of the review	Good plus: • the review team draws on external stakeholder expertise to drive the focusing process in order to achieve maximal end-user relevance

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers (*continued*)

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
4. Scoping the literature				
An important process in a meta-narrative review is to identify a sufficiently broad range of sources to be able to build as comprehensive a map as possible of research undertaken on the topic. This scoping step is used to identify in broad terms the different research traditions, situated in different literatures, which have addressed the topic of interest. Initial attempts to make sense of a topic area may involve not just informal browsing of the literature but also consulting with experts and stakeholders				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The scoping of the literature has been sufficiently and appropriately undertaken	The scoping of the literature has been limited and cursory (e.g. only a single source is used – perhaps the MEDLINE database – and/or the review has inappropriately concentrated on a single research tradition, for example evidence-based medicine)	Attempts made to utilise a broad range of relevant sources and to build as comprehensive a map as possible of the research traditions on the topic	Adequate plus: • a coherent and thorough search strategy, deliberately including exploratory methods such as browsing and modified in the light of emerging findings, is used to identify research traditions	Good plus: • systematic use is made of experts and stakeholders in identifying research traditions
5. Developing a search strategy				
Searching in a meta-narrative review is guided by the objectives and focus of the review, and revised iteratively in the light of emerging data. Searching is directed at finding sufficient data to develop and make sense of the relevant research traditions that have been identified, and may lie in a broad range of sources that may cross traditional disciplinary, programme and sector boundaries. This stage is likely to involve searching for different kinds of data in different ways				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The search process is such that it would identify data to enable the review team to develop and refine the map of seminal papers and primary research studies	The search is incapable of supporting the development of a rigorous meta-narrative review. Errors may include: • the search is driven by a methodological hierarchy of evidence (e.g. privileging RCTs) rather than the need to identify the range of research paradigms (concepts, theories, methods and instruments) that have been brought to bear on a topic • the search process is not informed by the objectives and focus of the review • the database(s) selected are narrow in the subject matter that they contain	Searches are driven by the objectives and focus of the review and are piloted and refined to check that they are fit for purpose Documents are sought from wide range of sources which are likely to contain relevant data on research traditions There is no predefined restriction on the study or documentation type that is searched for	Adequate plus: • further searches are undertaken in light of greater understanding of the topic area, particularly through the use of citation tracking of seminal papers. These searches are designed to find additional data that would allow greater sense to be made of component research traditions and/or draw higher-order insights from contrasts between traditions	Good plus: • the search reflects a high degree of scholarly insight into the key research traditions of the review

continued

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers (*continued*)

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
	(e.g. limited to biomedical topics and approaches rather than extending to social science, psychology, etc.) • Searching is undertaken once only at the outset of the review and there is no iterative component			
6. Selection and appraisal of documents				
Meta-narrative review is not a technical process, rather, it is a process of sensemaking of the literature, selecting and combining data from primary sources to produce an account of how a research tradition unfolded and why, and then (in the synthesis phase) comparing and contrasting findings from these different traditions to build a rich picture of the topic area from multiple perspectives. This process requires a series of judgements about the unfolding of research, in particular traditions, and about the relevance and robustness of particular data within that tradition Meta-narrative review takes its quality criteria from the traditions included in the review. Studies in these separate traditions should be appraised using the quality criteria that a competent peer reviewer in that tradition would choose to use The description of the selection and appraisal process should be sufficiently detailed to enable a reader to judge how likely it is that researchers inadvertently excluded data that may have significantly altered the findings of the review				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The selection and appraisal process ensures that sources relevant to the review containing material likely to help identify, develop and refine understanding of research traditions are included	The selection and appraisal process does not support a rigorous and complete meta-narrative review. For example: • selection is overly driven by methodological hierarchies (in particular the restriction of the sources to RCTs to the exclusion of other forms of evidence) • sources are appraised using a technical checklist focused on methodological procedure rather than by making a defensible judgement on the contribution that a source might make • selection and appraisal processes are overly restrictive and exclude materials that may help sensemaking of research traditions	Selection of a document for inclusion into the review is based on what it can contribute to making sense of research traditions All the key high-quality sources are identified and included in the review and the poor-quality ones accurately excluded	Adequate plus: • during the appraisal process studies in the separate traditions are appraised competently using the quality criteria acceptable to that tradition	Good plus: • the judgements made when appraising papers are a model of good scholarship in the relevant tradition

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers (*continued*)

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
	• selection and appraisal processes are not sensitive enough to exclude irrelevant materials			
7. Data extraction				
In a review, data extraction assists analysis and synthesis. Of particular interest to the meta-narrative reviewer are data elements that would contribute to constructing a story of how research on a topic unfolded over time in a particular tradition				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The data extraction process captures the necessary data to enable a meta-narrative review	The data extraction process does not capture the necessary data to enable a meta-narrative review. For example: • data extraction is undertaken mechanically and with no attention to how the data informs the review • No or very limited piloting is undertaken to test aspects of the data extraction process and improve it	Data extraction focuses on identification and elucidation of data that informs how research on a topic unfolded over time in a particular tradition Piloting and refinement of the data extraction process is undertaken where appropriate Quality control processes are in place to check that all review team members apply common processes and standards in data extraction	Adequate plus: • data extraction processes support later processes of analysis (e.g. by organising data into sets relevant for later analysis). • the data extracted are comprehensive enough to identify important topics that concern a research tradition, for example: o upstream (antecedent) traditions from which these emerged; background philosophical assumptions o research questions and how they were framed o key conceptual and theoretical issues o preferred methodologies, study designs, and quality criteria o key actors and events in the unfolding of the tradition o landmark empirical or theoretical studies o significant findings and how these shaped subsequent work; and	Good plus: • the data extraction process is continually refined as the review progresses, so as to capture relevant data as the review question is focused and/or research traditions identified and elucidated

continued

TABLE 11 Quality standards for meta-narrative reviews for researchers and peer reviewers (*continued*)

Quality standards for meta-narrative reviews (for researchers and peer reviewers)				
			○ key debates and areas of dispute within the tradition, including links with or breaches from other traditions	
8. Synthesis phase				
Having identified the individual meta-narratives, the next phase in a meta-narrative review is to compare and contrast these to generate higher-order data (e.g. to identify and explain conflicting findings)				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The meta-narrative should include a synthesis phase where philosophical, conceptual, methodological and empirical differences between traditions are discussed and explained	The synthesis phase is missing or fails to engage with the underlying philosophical, conceptual or theoretical contrasts between traditions	Some attempt is made to show how different groups of researchers produced different findings as a result of different philosophical assumptions, different ways of conceptualising the topic, different theoretical explanations or different study designs and methods	Adequate plus: • the contrasting accounts of different traditions are synthesised in a way that generates robust higher-order data (e.g. about the contestation between different research storylines at policy level)	Good plus: • the review generates additional philosophical, conceptual, theoretical or methodological insights that inform innovations in research
9. Reporting				
Meta-narrative reviews may be reported in multiple formats – lengthy reports, summary reports, articles, websites and so on. Reports should be consistent with the publication standards for meta-narrative reviews. (See RAMESES publication standards: realist syntheses at: http://onlinelibrary.wiley.com/doi/10.1111/jan.12095/full[89], or www.biomedcentral.com/1741-7015/11/21)[89,90]				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
The meta-narrative review is reported using the items listed in the relevant RAMESES reporting standard	Key items are missing. For example: • no defined research question • limited or no reporting of the review's processes (i.e. methods used) • limited or no explanations and justifications provided for any adaptations made on the meta-narrative review process • insufficient detail is reported to enable readers to judge the plausibility and coherence of the findings	Most items reported. In particular the following items should be reported: • rationale for review • objectives and focus of review • all method section items (i.e. items 5–12 in the RAMESES publication standards: meta-narrative reviews)	All items are reported clearly and in sufficient detail for an external reader to understand and to judge the methods used and the plausibility and coherence of the findings	Good plus: • the report is well written and easy to understand. Additional materials are made available for external readers to investigate aspects of the review in more detail

RCT, randomised control trial.

Quality standards for funders/commissioners of research

As more and more realist syntheses and meta-narrative reviews are being funded/commissioned, decision-makers and peer reviewers at this stage need to make judgements on two broad areas: proposed review processes and methodological expertise. We appreciate that many funding bodies and commissioners will already have processes in place to guide the peer reviewers they appoint. As such we see this set of guidance we have produced not as replacement for, but as supplementation to, any existing organisational peer-review processes and guidance.

The quality standards for realist syntheses for funders/commissioners of research are set out in *Table 12*. Those for meta-narrative reviews are in *Table 13*. These have been abridged and adapted from their respective counterparts in *Tables 10* and *11* to better suit the needs of this user group.

TABLE 12 Quality standards for realist reviews for funders/commissioners of research

Quality standards for RS (for funders/commissioners of research)				
1. The research problem				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is the research topic appropriate for a realist approach?	Research topic: • is not appropriate for secondary research; and/or • does not require understanding of how and why outcomes are generated	Research topic: • is appropriate for secondary research • requires understanding of how and why outcomes are generated and why they vary across contexts	Adequate plus: • framing of the research topic reflects a thorough understanding of a realist philosophy of science	Good plus: • there is a coherent argument as to why a realist approach is more appropriate for the topic than other approaches
Is the research question constructed in such a way as to be suitable for a RS?	The research question is not structured to reflect the elements of realist explanation	The research question includes a focus on how and why the intervention, or programme, generates its outcomes and contains at least some of the additional elements, for whom, in what contexts, in what respects, to what extent and over what durations	Adequate plus: • the rationale for excluding any elements of the realist question from the research question is explicit. • the question has a narrow enough focus to be managed within a realist review	Good plus: • the research question is a model of clarity and as simple as possible
2. Understanding and applying the underpinning principles of realist reviews				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Does the review team demonstrate understanding and application of realist philosophy and realist logic which underpins a realist analysis?	Significant misunderstandings of realist philosophy and/or logic of analysis are evident	Some misunderstandings of realist philosophy and/or logic of analysis exist, but the overall approach is consistent enough that a recognisably realist analysis results from the process	• The review's assumptions and analytic approach are consistent with a realist philosophy at all stages of the review • Where necessary a realist programme theory is developed and tested	Good plus: • proposed review methods, strategies or innovations planned to address problems or difficulties within the review are consistent with a realist philosophy of science

continued

TABLE 12 Quality standards for realist reviews for funders/commissioners of research (*continued*)

Quality standards for RS (for funders/commissioners of research)				
3. Focusing the review				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is, or will, the review question be sufficiently and appropriately focused?	• The review question is too broad to be answerable within the time and resources allocated • There is no evidence that progressive focusing will occur as the review progresses	Process proposed enables the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints	Adequate plus: • the focusing process is iterative • commissioners of the review are involved in decision-making about focusing	Good plus: • the review team draws on external stakeholder expertise to drive the focusing process in order to achieve maximal end-user relevance
4. Constructing and refining a realist programme theory				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Does the review team plan to identify, develop and refine their initial realist programme theory?	There are no plans to identify, develop and refine a realist programme theory	There are plans to identify, develop and refine a realist programme theory	Adequate plus: • the initial realist programme theory is set out at the start and will be refined iteratively as the review team's understanding of the topic grows	Good plus – there are plans to: • identify and explain the relationship between the programme theory and relevant substantive theory • draw on, where necessary, external expertise to develop their programme theory
5. Developing a search strategy				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is the proposed search process such that it would identify data to enable the review team to develop, refine and test programme theory or theories?	The search is incapable of supporting a rigorous realist review	The proposed searches will: • be driven by the objectives and focus of the review • be piloted and refined • seek out documents from a wide range of sources likely to contain relevant data • not be restricted by study or documentation type	Adequate plus: • further searches will be undertaken in light of greater understanding of the topic area	Good plus: • the searching will deliberately seek out data from situations where it can be reasonably inferred that the same mechanism(s) might be in operation

TABLE 12 Quality standards for realist reviews for funders/commissioners of research (*continued*)

Quality standards for RS (for funders/commissioners of research)				
6. Selection and appraisal of documents				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Will the selection and appraisal process ensure that documents of **relevance** to the review containing material of sufficient **rigour** to be included are identified?	The proposed selection and appraisal process does not support a rigorous and complete realist review	Selection of a document for inclusion will be based on: • relevance, i.e. what it can contribute to the process of theory development, refinement and/or testing • rigour – judgements will be made based on the plausibility and coherence of the method used to generate data	Adequate plus: • during the appraisal process limitations of the method used to generate data will be identified and taken into consideration during analysis and synthesis	As for Good
7. Data extraction				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*

Will the data extraction process capture the necessary data to enable a realist review?	• The data extraction process does not capture the necessary data to enable a realist review
	• No piloting of the data extraction process is planned
		The data extraction processes will:
		• focus on identification and elucidation of context-mechanism outcome configurations and refinement of programme theory
		• be piloted and refined where appropriate
		• include quality control processes to ensure uniformity of processes and standards
			Adequate plus:
			• data extraction processes will:
			  ○ support later processes of analysis (e.g. by organising data into sets relevant for later analysis)
			  ○ be comprehensive enough to identify main CMO patterns
				Good plus:
				• there are plans to continually refine the data extraction process as the review progresses, so as to capture relevant data as the review question is focused and/or programme theory is refined

8. Reporting				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Will the review team use the items listed in the RAMESES reporting standard for realist syntheses when reporting their RS?	No information provided	RAMESES reporting standard for realist syntheses will be used for reporting	Adequate plus: • firm commitment made to adhere to all items within the RAMESES reporting standard for realist syntheses	As for Good

TABLE 13 Quality standards for meta-narrative reviews for funders/commissioners of research

Quality standards for meta-narrative reviews (for funders/commissioners of research)				
1. The research problem				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is the research topic appropriate for a meta-narrative approach?	Research topic: • is not appropriate for secondary research; and/or • does not require understanding of how a topic has been conceptualised and studied differently by different groups	Research topic: • is appropriate for secondary research • would benefit from illumination of how a topic has been conceptualised and studied differently by different groups	Adequate plus: • framing of the research topic reflects a thorough understanding of the value, importance and implications of different approaches on research practice and findings	Good plus: • there is a coherent argument as to why a meta-narrative review is more appropriate for the topic than potential alternatives
Is the research question constructed in such a way as to be suitable for a meta-narrative review?	The research question is not structured to reflect the elements of meta-narrative explanation	The research question includes a focus on how a topic has been conceptualised and studied differently by different groups	Adequate plus: • the research question includes an element that addresses the implications of different conceptualisations and approaches to a topic on research findings	Good plus: • the research question is a model of clarity and as simple as possible
2. Understanding and applying the purpose and underpinning principles of meta-narrative reviews				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Does the review team demonstrate an understanding and application of the purpose and principles underpinning a meta-narrative review?	Significant misunderstandings of the purpose and principles underpinning a meta-narrative review	Some misunderstandings of the purpose and principles underpinning a meta-narrative review, but the overall planned approach is consistent enough that a recognisable set of distinct meta-narratives together with a higher-order synthesis of these is likely to results from the process	• The review's assumptions and planned analytic approach are consistent with the purpose and underpinning principles of a meta-narrative review • The philosophical position is explicitly constructivist • A sufficient range of paradigms/epistemic traditions is likely to be included for sensemaking and use made of contrasts between these as higher-order data	Good plus: • review methods, strategies or innovations planned to address problems or difficulties within the review are philosophically coherent and make a clear and illuminative contribution to the knowledge base on the topic area

TABLE 13 Quality standards for meta-narrative reviews for funders/commissioners of research (*continued*)

Quality standards for meta-narrative reviews (for funders/commissioners of research)				
3. Focusing the review				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is, or will, the review question be sufficiently and appropriately focused?	• The review question is too broad to be answerable within the time and resources allocated • There is no evidence that progressive focusing will occur as the review progresses	Attempts will be made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints	Adequate plus: • the focusing process will be iterative and reflexive • commissioners of the review will be involved in decision-making about focusing	Good plus: • the review team will draw on external stakeholder expertise to drive the focusing process in order to achieve maximal end-user relevance
4. Scoping the literature				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Has sufficient and appropriate scoping of the literature been planned?	The planned scoping of the literature appears to be limited and cursory	Attempts will be made to utilise a broad range of relevant sources and to build as comprehensive a map as possible of the research traditions on the topic	Adequate plus: • a coherent and through search strategy will be used, deliberately including exploratory methods such as browsing and will be modified in the light of emerging findings	Good plus: • systematic use will be made of experts and stakeholders in identifying research/epistemic traditions
5. Developing a search strategy				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Is the proposed search process such that it would identify data to enable the review team to develop and refine the map of seminal papers and primary research studies?	The planned search is incapable of supporting the development of a rigorous meta-narrative review	The proposed searches will: • be driven by the objectives and focus of the review • be piloted and refined • seek out documents from a wide range of sources likely to contain relevant data on research traditions • not be restricted by study or documentation type	Adequate plus: • further searches will be undertaken in light of greater understanding of the topic area, particularly through the use of citation-tracking of seminal papers	As for Good

continued

TABLE 13 Quality standards for meta-narrative reviews for funders/commissioners of research (*continued*)

Quality standards for meta-narrative reviews (for funders/commissioners of research)				
6. Selection and appraisal of documents				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Will the selection and appraisal process ensure that sources relevant to the review containing material likely to help identify, develop and refine understanding of research traditions be included?	The selection and appraisal process will not support a rigorous and complete meta-narrative review	Selection of a document for inclusion into the review will: • be based on what it can contribute to making sense of research traditions • accurately include all the key high-quality sources identified and exclude the poor-quality ones	Adequate plus: • during the appraisal process studies in the separate traditions will be appraised using the quality criteria acceptable to that tradition	As for Good
7. Data extraction				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*

Will the data extraction process capture the necessary data to enable a meta-narrative review?	The data extraction process will not capture the necessary data to enable a meta-narrative review	Data extraction processes will:
		• focus on identification and elucidation of data that informs how research on a topic unfolded over time in a particular tradition
		• be piloted and refined where appropriate
		• include quality-control processes to ensure uniformity of processes and standards
			Adequate plus:
			• data extraction processes will:
			  ○ support later processes of analysis (e.g. by organising data into sets relevant for later analysis)
			  ○ be comprehensive enough to identify important topics that concern a research tradition
				Good plus:
				• the data extraction process will be continually refined as the review progresses, so as to capture relevant data as the review question is focused and/or research traditions identified and elucidated

8. Synthesis phase				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Will synthesis of the meta-narratives include discussion and explanation of the philosophical, conceptual, methodological and empirical differences between traditions?	A synthesis phase: • is not planned, or • is planned in such a way that it fails to engage with the underlying philosophical, conceptual or theoretical contrasts between traditions	The planned synthesis phase will attempt to show how different groups of researchers produced different findings as a result of different philosophical assumptions, ways of conceptualising the topic, theoretical explanations or study designs and methods	Adequate plus: • Contrasting accounts of different traditions will be sought out and synthesised in a way that generates robust higher-order data	As for Good
9. Reporting				
Criterion	*Inadequate*	*Adequate*	*Good*	*Excellent*
Will the review team use the items listed in the RAMESES reporting standard for meta-narrative reviews when reporting their meta-narrative review?	No information provided	RAMESES reporting standard for meta-narrative reviews will be used for reporting	Adequate plus: • firm commitment made to adhere to all items within the RAMESES reporting standard for meta-narrative reviews	As for Good

Teaching and learning resources

We developed teaching resources for both realist and meta-narrative reviews. The challenge we faced when tackling this task was that both methods were relatively new and as yet only some methodological development has taken place. As a project team, we discussed at length and repeatedly what our fellow researchers might find helpful, as we were aware that time and resources were limited. We took our inspiration of what kind of teaching and learning resources to produce from the feedback we had obtained from fellow realist researchers on a paper a member of our project team (RP) had co-authored on realist evaluation.[94] The feedback we had been given was that focusing on areas that researchers found challenging and teaching through examples from the literature was helpful. We decided to adopt this format for both sets of our training materials – namely to focus on the aspects of each review method that researchers found the most challenging and to illustrate both good and bad practice with examples from the published literature. From our analysis of the data we gathered from our various sources (see *Chapter 4, Literature search* and *Delphi panel* and *Developing quality standards, teaching and learning resources using real-time refinement*) we noted that there were specific review method issues that fellow researchers found the most challenging and focused our teaching and learning materials on these.

For realist reviews, the challenging issues we covered were:

- Focusing reviews.

  Because a RS will generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it may be necessary to contain a review. Many different aspects of a realist review might need to be focused. Focusing may also take place at different time points in the review process.

- Programme theory.

  Realist synthesis has most often been used to make sense of complex interventions. These interventions or programmes often have multiple components (which interact in non-linear ways), outcomes (some intended and some not) and long pathways to the desired outcome(s). The term programme theory refers to an abstracted description and/or diagram that lays out what a programme (or family of programmes or intervention) comprises and how it is expected to work. Programme theory serves two main functions in a RS. The first is to sketch the terrain that will be investigated and, in the process, to assist in refining the elements and scope for the review. The second is to provide a structure for review findings.

- Developing a search strategy.

  What constitutes the right evidence is different in a RS than it is in other form of review. Data that may usefully contribute to a RS are:

  - decided not by research type (e.g. randomised controlled trial) but by relevance to the review question
  - not restricted to research into or evaluations of programs per se, but related to the programme theory that underpins the programme
  - not necessarily about the whole research question, but relevant to a subsection of it
  - drawn not necessarily from a whole text/document, but from a subsection of it relevant to a particular aspect of the review question
  - able to shed light on any aspect of context (C), mechanism (M) or outcome (O) for any element of the theory
  - different for theory building (which does not need to be as rigorous) as opposed to theory testing (which needs to be sufficiently rigorous to support the conclusion being drawn on for the review).

- Selection and appraisal of documents.

  Realist synthesis requires a series of judgements about the relevance and robustness of particular data items for the purposes of answering a specific question. A wide range of documents may contain data that contribute to a RS. Hence, rejecting a document on a global assessment of its methodological quality is illogical. Instead, inclusion and exclusion decisions are based on two criteria:

  - *Relevance* – whether it can contribute to theory building and/or testing.
  - *Rigour* – whether or not the methods used to generate the relevant data are credible and trustworthy.

- Applying realist principles in analysis.

  The basic analytic task in a realist review is to *find and align the evidence* to demonstrate that particular mechanisms generate particular outcomes and to demonstrate which aspects of context matter. Working from the basic analytic structure described above, it follows that relevant mechanisms cannot be identified without reference to outcomes (mechanisms are what cause outcomes) and that relevant aspects of context cannot be identified without reference to mechanisms. An ideal RS provides evidence for outcomes, evidence to support the existence of the hypothesised mechanisms, evidence that those mechanisms cause those outcomes, evidence that features of context exist and evidence that those features of context affect whether and which mechanisms fire.

The challenging issues faced by meta-narrative reviewers were very similar and concerned:

- Understanding and applying the underpinning principles of meta-narrative reviews

  Meta-narrative review (which is rooted in a constructivist philosophy of science) was inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms.[15] Meta-narrative reviews often look historically at how particular research traditions or epistemic traditions have unfolded over time and shaped the normal science of a topic area. The review seeks first to identify and understand as many as possible of the potentially important different research traditions that have a bearing on the topic. In the synthesis phase, by means of an overarching narrative, the findings from these different traditions are compared and contrasted to build a rich picture of the topic area from multiple perspectives. The goal of meta-narrative review is sensemaking of a complex (and perhaps contested) topic area.

- Focusing reviews

  Because a meta-narrative review will generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it may be necessary to contain a review. Many different aspects of a realist review might need to be focused. Focusing may also take place at different time points in the review process.

- Finding the most relevant evidence

  Three specific processes will help the meta-narrative reviewer find the most relevant evidence:

  - scoping the literature
  - developing and pursuing a search strategy
  - selecting and appraising the documents.

Through an iterative cycle of drafts, feedback on drafts and revisions we developed the final structure of our teaching materials. We drew on our collective experiences in teaching and learning as well as knowledge of the educational literature to develop these materials. A particular challenge we faced when developing these materials was deciding on who our exact audience would be, i.e. the novice, intermediate or advanced reviewer. We finally decided that we would focus on providing materials for the more novice end of this spectrum, as many of the enquiries we (as a project team) were getting asking for help were from novice review teams, and we and our fellow realist and meta-narrative reviewers had identified that capacity building was a real and significant issue. Each of the teaching and learning resources has a similar structure and cover:

- objectives
- an explanation on why the topic area is important to get right
- what would constitute high quality for this topic area
- one or more worked examples (drawn from the published literature) of how the topic area in a review might be improved
- example(s) from the published literature of how the topic area has been tackled successfully
- learning activities (realist review only)
- reflection activities.

A list of suggested further reading and resources is provided within each of the teaching and learning materials documents. The teaching and learning materials for RS and meta-narrative are in *Appendix 9* and *10*, respectively, and are freely available online.[93]

In addition to these teaching and training materials, one of our project team (Professor Ray Pawson) has written a book on realist research methods that also provides more in-depth discussions on various aspects of realist review.[95]

Chapter 5 Discussion

In this project we have developed publication standards, quality standards, and teaching and learning resources for realist and meta-narrative reviews. Both are relatively new systematic review methods in health services research. Realist and meta-narrative reviews potentially offer great promise in unpacking the black box of the many complex interventions that are increasingly being used to improve health and patient outcomes. We see this project as a start to the long journey of advancing the rigour of how realist and meta-narrative reviews are carried out and reported.

Both realist and meta-narrative reviews are methods that have grown out of the increasing call for secondary research methods to address issues around the implementation of interventions.[96] They are not the only review methods that try to address this challenge, other examples include meta-ethnography, grounded theory, thematic synthesis, textual narrative synthesis, meta-study, critical interpretive synthesis, ecological triangulation and framework synthesis. With this growth in possible review methods, one unintended consequence has been that there may now be too much choice and it is not immediately apparent which method should be used and when.[5] A detailed discussion of this issue is beyond the scope of this report, but excellent resources exist that may help in the choice of review methods.[97,98]

As relatively experienced users of these methods, we had noted a number of common and recurrent challenges that face grant awarding bodies, peer reviewers, reviewers and users. These centred on two closely related questions: how to judge if a realist or meta-narrative review, or a proposal for such a review, is of high quality (including, for completed reviews, how credible and robust findings are) and how to undertake such reviews. Our experience suggests that we can go a long way towards answering these questions by developing resources that helps fellow reviewers to give due consideration to the theoretical and conceptual underpinnings of realist and meta-narrative reviews, outlined briefly below.

Realist review is based on a realist philosophy of science, which permeates and informs its underlying epistemological assumptions, methodology and quality considerations. Meta-narrative review takes a more constructivist philosophical position, though it is compatible with approaches which propose the existence of a social reality independent of our constructions of it. The meta-narrative approach seeks to tease out and explore the full range of philosophical positions represented in the primary literature.

One of the most common misapplications we have noted is that reviewers have not always appreciated the underlying philosophical basis of these review methods (and the implications of this for how the review should be conducted). Instead, they have based their reviews explicitly or implicitly on fundamentally different philosophical assumptions – most commonly the positivist notion that generalisable truths are best generated from controlled experiments, especially randomised trials.

Even when a realist philosophy of science has been adhered to in a realist review, reviewers – ourselves included – often struggle with recurring conceptual and methodological issues. Mechanisms present a particular challenge in realist review – how to define them, where to locate them, how to identify them and how to test and refine them. Both review methods trade on the use of theoretical explanations to make sense of the observed data. Realist reviewers commonly grapple with how to define a theory (what, for example, is the difference between a programme theory and a middle-range theory?) and what level of abstraction is appropriate in different circumstances. On a more pragmatic level, those who seek to produce theory-driven reviews of heterogeneous topic areas wrestle with a broad range of how to issues: how to define the scope of the review; how and to what extent to refine this scope as the review unfolds; what literature(s) to search and how; how to critically appraise what is often a very diverse sample of primary studies; how to collate, analyse and synthesise findings; and how to make recommendations that are academically defensible and useful to policy-makers and so on. We believe that the resources we have produced from this project will go some way to addressing the challenges we have highlighted above.

In undertaking this project we were faced with one main dilemma that related to how best to allocate time and resources to the multiple work packages. For example, we could easily have spent more time on our narrative review but this may potentially have been at the expense of relatively neglecting our Delphi panels, support to review teams or developing teaching materials. In retrospect our project was very ambitious in its aims and as such we had to prioritise some aspects of the project above others. For example, we felt that it was more important to devote more time to getting right our Delphi process, so that we had a solid a consensus on which to develop our quality and publication standards and (to a lesser extent) our teaching materials. This meant that our narrative review had to be rapid/truncated/abbreviated (see *Chapter 3, Details of literature search methods* and *Chapter 4, Literature search* for more details). Another example of prioritisation was in the breadth and depth of our training materials. Entire textbooks could be written for these, but instead we chose to focus on common challenges. Our hope is that we have started the journey to addressing some of the issues around all new methods – namely how do you judge quality, how do you report it and how do you do X, Y or Z. We do however fully accept that more is needed and as such we have provided recommendations in *Research recommendations and implications for practice*.

Changes to protocol

Near the start of this project we published our project protocol.[99] During the course of the project we varied two aspects of our protocol, as described in the sections below.

Real-time piloting of the provisional standards, guidance and training materials

Our intention had been that over the 27-month duration of this study, we would recruit two cohorts of review teams. With the first cohort, we would use provisional standards, guidance and training materials developed from our initial review of the literature. Whereas, with the second cohort, we would pilot the standards, guidance and training materials which had been produced/refined via the Delphi process. After following the two cohorts of review teams through their reviews, we would then further revise the outputs as a master document before considering how to modify these for different audiences.

However, there were a number of issues that made our plans impractical and potentially misleading. Firstly, it was not immediately apparent from our literature review what the main methodological and training challenges were. Secondly, we had no control over when review teams wanted us to provide them with methodological support. It was, therefore, difficult for us to assemble together the necessary cohorts and have our initial drafts ready. We found ourselves providing methodological support almost on a continuous basis but to different teams at different times on a wide range of different methodological aspects. Getting clear starting and finishing points in time for any cohort we could assemble was impossible if we wanted to be responsive to the needs of review teams. Finally, we noted early on in our project that while the literature was useful in helping us to identify methodological and training challenges, our fellow reviewers were better. We found that review teams possessed an invaluable store of knowledge about the challenges they faced. As we supported review teams, communicated via e-mail and met them at conferences and workshops, we were able to harness and gather more and more information about what they found really challenging. We thus made the decision that iterative refinement (building on the gradually accumulating experiences of fellow researchers) might prove to be a more fruitful way of developing our resources rather than what we had originally planned.

Fishbowl exercise

Approximately halfway through the study period, we had planned to formally present our emerging findings to a panel of external researchers in order to collate additional feedback. We had planned this event as a precaution against any groupthink that we might encounter within our project team. We discussed the need for such an event with our project steering group, especially in the light of the fact that we had been able to recruit what we considered to be a very diverse range of individuals to our

Delphi panels. With the agreement of our project steering group we decided that there was little merit in holding such an event.

Limitations

To develop the briefing materials for our Delphi panels we undertook a narrative review. This review has limitations that are likely to have introduced a number of biases and so – potentially at least – limit the inferences that can be made from the included reviews. For example, the search process for the review, despite being developed by an expert librarian, was not exhaustive. All the screening for inclusion and exclusion was undertaken by one screener and no quality checks were undertaken. Both processes may well mean that we are likely to have missed some reviews. Once reviews had been included, data extraction was undertaken by one researcher, and omissions in data extraction are likely to have occurred. However, all the included reviews and the data extraction spreadsheet were circulated to all project team members and so a degree of informal quality checking did occurr. Deciding what should be included in the Delphi panels' briefing materials was undertaken by the entire project team. We are aware that any item or topic included in the briefing materials was done so as a result of our subjective interpretations, raising questions about reproducibility. However, the briefing documents we produced were not an end product in itself, but the starting point for the Delphi panels to build a consensus. As such, we expected that changes would occur as we ran each round of the Delphi process and, so, were not as concerned that any omissions as a result of the review's limitations process would have that large an impact on the final publication and quality standards.

We recognise that there is much more to cover in terms of the breadth and depth of the teaching and learning resources we have produced. Because realist and meta-narrative reviews are both relatively new review methods, the wish list we were able to elicit from our fellow reviewers when using these methods is quite long. Given the time and resources allocated for this project we elected to focus on providing depth, rather than breadth on the issues that were most challenging. With time, we hope to use the community of practice we have developed to address more and more methodological challenges.

As experience grows with the use of these methods, it is very likely that the resources we have produced will need to be updated. We welcome and invite methodological development in realist and meta-narrative reviews. We expect that what we have produced should be gradually refined and updated as methodological developments take place with increasing use of realist and meta-narrative reviews. Thus we view the publication and quality standards and teaching and learning resources more as a starting point than definitive resources that must not be altered in any way.

We are aware that both realist and meta-narrative reviews are used for secondary research on a wide range of topics and by reviewers from a broad range of disciplines. The level of expertise of the users of our resources will also vary considerably, from novice to seasoned reviewer. These two aspects mean that some latitude is needed in the use of the resources we have produced. For example, not all the publication standard's items will always be applicable when reporting all reviews. Or when assessing the quality of a review, there may be justifiable reasons for a review to not meet some quality criteria. We have tried to anticipate the varied uses that realist and meta-narrative review might be put to by providing a degree of flexibility in our standards. For example, in our publication standards, if adaptations are made to the review method (as originally described), then reviewers are invited to provide an explanation for any such adaptations.

Research recommendations and implications for practice

In common with many quality and reporting standards there is a dearth of research to demonstrate that such standards necessarily change practice and improve the quality of research.[18] This will also be true for the standards we have produced and, therefore, research to demonstrate a change in practice and improvement in the quality of realist and meta-narrative reviews is needed.

As experience with realist and meta-narrative reviews grows and more are undertaken, new methodological insights are likely to occur. These need to be captured and analysed to determine if the quality and publication standards we have produced continue to be fit for purpose or need to be updated. Ideally, further funding might enable a project similar to this one, i.e. RAMESES II, to address the updating of the standards, though as much groundwork has already been done a more truncated project may suffice.

Our training materials are focused on what we were able to identify as being the processes that fellow reviewers found the most challenging to execute. There are additional processes that we have not focused on and further work is needed to identify these. With our training materials we have chosen to produce learning materials that teach by using examples from published reviews and through use of learning and reflection activities. They have not been formally evaluated and are likely to benefit from iterative cycles of evaluation and updating based on findings.

Finally, both realist and meta-narrative reviews are relatively new approaches and as with any approach, capacity building is an issue. This project has enabled the project team to support and build capacity with a number of researchers and set up an e-mail mailing list to bring researchers together. A pressing need for the future is to maintain the momentum generated by this project. To this end, the JISCMail e-mail list continues to run but we invite any researchers interested in either method to join us in helping to build capacity.

Chapter 6 Conclusion

In conclusion, while realist and meta-narrative reviews hold much promise for developing theory and informing policy in some of the health sector's most pressing questions, misunderstandings and misapplications of these methods are common. To try to address these problems we have produced publication and quality standards, and teaching and learning materials. We hope that our resources will be the start of an iterative journey of refinement and development of better resources for realist and meta-narrative reviews. Acknowledging that research should never be static, the RAMESES project does not seek to produce the last word on this topic but to capture current expertise and establish an agreed state of the science on which future researchers will no doubt build.

Acknowledgements

We wish to thank the following:

- Project steering group – Drs C. Meads, J. Lancaster and D. Noble.
- Literature searching – Jeanette Buckingham.

For their feedback, comments and time:

- All contributors to the RAMESES JISCMail.
- All Delphi panel members (see *Appendix 11*).
- All the review teams we supported.

Contributions of authors

Dr Geoff Wong is a co-investigator in the RAMESES project and is a senior lecturer in primary health care at Queen Mary University of London, London, UK.

Professor Trish Greenhalgh is the principal investigator on the RAMESES project and is Professor of Primary Health Care and Dean for Research impact at Queen Mary University of London, London, UK.

Dr Gill Westhorp is an evaluation expert and has used both the realist evaluation and synthesis methods on a range of projects internationally. She is the Director of Community Matters, a research consultancy in Adelaide, SA, Australia.

Professor Ray Pawson is a co-investigator in the RAMESES project and is Professor of Social Research Methodology at the University of Leeds, Leeds, UK.

Geoff Wong carried out the literature review with the help of Jeanette Buckingham (Librarian, John W. Scott Health Science Library, University of Alberta, AB, Canada). Geoff Wong, Trish Greenhalgh, Gill Westhorp and Ray Pawson analysed the findings from the review and produced the materials for the Delphi panel. They also analysed the results of the Delphi panel and real-time data to produce the publication standards, the methodological guidance and teaching and learning materials. Geoff Wong, Trish Greenhalgh, Gill Westhorp and Ray Pawson conceived of the study and participated in its design. Geoff Wong co-ordinated the study and ran the Delphi panels. All authors read and approved the final manuscript.

Publications

Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: meta-narrative reviews. *BMC Med* 2013;**11**:20.

Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: realist syntheses. *BMC Med* 2013;**11**:21.

Jagosh J, Macaulay AC, Salsberg J, Bush PL, Henderson J, Sirett E, *et al.* Uncovering the benefits of participatory research: Implications of a realist review for health research and practice. *Milbank Q* 2012;**90**:311–46.

Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011;**343**:d7163.

References

1. Berwick DM. The science of improvement. *JAMA* 2008;**299**:1182–4. http://dx.doi.org/10.1001/jama.299.10.1182
2. Lavis JN. How can we support the use of systematic reviews in policymaking? *PLOS Med* 2009;**6**:e1000141. http://dx.doi.org/10.1371/journal.pmed.1000141
3. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy* 2005;**10**(Suppl. 1):6–20. http://dx.doi.org/10.1258/1355819054308576
4. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;**10**(Suppl. 1):21–34. http://dx.doi.org/10.1258/1355819054308530
5. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 2009;**9**:59. http://dx.doi.org/10.1186/1471-2288-9-59
6. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* 2005;**10**:45–53. http://dx.doi.org/10.1258/1355819052801804
7. Lucas PJ, Baird J, Arai L, Law C, Roberts HM. Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Med Res Methodol* 2007;**7**:4. http://dx.doi.org/10.1186/1471-2288-7-4
8. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. *Health Res Policy Syst* 2006;**4**:20. http://dx.doi.org/10.1186/1478-4505-4-20
9. Popay J, Rogers A, Wiliams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qual Health Res* 1998;**8**:341–51. http://dx.doi.org/10.1177/104973239800800305
10. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 2008;**8**:45. http://dx.doi.org/10.1186/1471-2288-8-45
11. Tricco AC, Tetzlaff J, Moher D. The art and science of knowledge synthesis. *J Clin Epidemiol* 2011;**64**:11–20. http://dx.doi.org/10.1016/j.jclinepi.2009.11.007
12. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, *et al.* Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. *Int J Technol Assess Health Care* 2008;**24**:133–9. http://dx.doi.org/10.1017/S0266462308080185
13. Pawson R. Evidence-based policy: the promise of 'realist synthesis'. *Evaluation* 2002;**8**:340–58. http://dx.doi.org/10.1177/135638902401462448
14. Pawson R. *Evidence-based Policy: a Realist Perspective*. London: Sage; 2006.
15. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med* 2005;**61**:417–30. http://dx.doi.org/10.1016/j.socscimed.2004.12.001
16. Moher D, Hopewell S, Schulz K, Montori V, Gotzsche P, Devereaux P, *et al.* CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *BMJ* 2010;**340**:c869. http://dx.doi.org/10.1136/bmj.c869

17. AGREE collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003;**12**:18–23. http://dx.doi.org/10.1136/qhc.12.1.18

18. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gotzsche P, Ioannidis J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;**339**:b2700. http://dx.doi.org/10.1136/bmj.b2700

19. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S, for the SQUIRE Development Group. Publication guidelines for improvement studies in health care: evolution of the SQUIRE project. *Ann Intern Med* 2008;**149**:670–6. http://dx.doi.org/10.7326/0003-4819-149-9-200811040-00009

20. Weiss C. Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In Connell JP, Kubisch AC, Schorr LB, Weiss CH, editors. *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington, DC: Aspen Institute; 1995. pp. 65–92.

21. Astbury B, Leeuw F. Unpacking black boxes: mechanisms and theory building in evaluation. *Am J Eval* 2010;**31**:363–81. http://dx.doi.org/10.1177/1098214010371972

22. Greenhalgh T, Kristjansson E, Robinson V. Realist review to understand the efficacy of school feeding programmes. *BMJ* 2007;**335**:858–61. http://dx.doi.org/10.1136/bmj.39359.525174.AD

23. Kristjansson E, Robinson V, Petticrew M, Macdonald B, Krasevec J, Janzen L, *et al.* School feeding for improving the physical and psychosocial health of disadvantaged elementary school children. *Cochrane Database Syst Rev* 2007;**1**:CD004676.

24. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organisations: systematic literature review and recommendations for future research. *Milbank Q* 2004;**82**:581–629. http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x

25. Best A, Terpstra J, Moor G, Riley B, Norman C, Glasgow R. Building knowledge integration systems for evidence-informed decisions. *J Health Organ Manag* 2009;**23**:627–41. http://dx.doi.org/10.1108/14777260911001644

26. Pawson R, Greenhalgh T, Harvey G, Walshe K. *Realist synthesis – an introduction*. ESRC Working Paper Series. London: ESRC; 2004.

27. McClure M. 'Clarity bordering on stupidity': where's the quality in systematic review? *J Educ Policy* 2005;**20**:393–416. http://dx.doi.org/10.1080/02680930500131801

28. Popper K. *The Logic of Scientific Discovery*. Vienna: Verlag von Julius Springer; 1935.

29. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;**32**:1008–15.

30. Campbell S, Roland M, Shekelle P, Cantrill J, Buetow S, Cragg D. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual Health Care* 1999;**8**:6–15. http://dx.doi.org/10.1136/qshc.8.1.6

31. Hsu C-C, Sandford B. The Delphi Technique: Making Sense of Consensus. *Pract Assess Res Eval* 2007;**12**:1.

32. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud* 2001;**38**:195–200. http://dx.doi.org/10.1016/S0020-7489(00)00044-4

33. Washington D, Bernstein S, Kahan J, Leape L, Kamberg C, Shekelle P. Reliability of clinical guideline development using mail-only versus inperson expert panels. *Med Care* 2003;**41**:1374–81. http://dx.doi.org/10.1097/01.MLR.0000100583.76137.3E

34. Russell J, Elton L, Swinglehurst D, Greenhalgh T. Using the online environment in assessment for learning: a case-based study of a web-based course in primary care. *Assess Eval Higher Educ* 2006;**31**:465–78. http://dx.doi.org/10.1080/02602930600679209

35. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, *et al.* Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006;**333**:417. http://dx.doi.org/10.1136/bmj.38926.629329.AE

36. Greenhalgh T, Wengraf T. Collecting stories: is it research? Is it good research? Preliminary guidance based on a Delphi study. *Med Educ* 2008;**42**:242–7. http://dx.doi.org/10.1111/j.1365-2923.2007.02956.x

37. Hart L, Bourchier S, Jorm A, Kanowski L, Kingston A, Stanley D, *et al.* Development of mental health first aid guidelines for Aboriginal and Torres Strait Islander people experiencing problems with substance use: a Delphi study. *BMC Psychiatry* 2010;**10**:78. http://dx.doi.org/10.1186/1471-244X-10-78

38. Holliday C, Robotin M. The Delphi process: a solution for reviewing novel grant applications. *Int J Gen Med* 2010;**3**:225–30. http://dx.doi.org/10.2147/IJGM.S11117

39. Pye J, Greenhalgh T. First aid kits for recreational dive boats: a Delphi study. *J Travel Med* 2010;**8**:311–17. http://dx.doi.org/10.1016/j.tmaid.2010.07.001

40. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *J Adv Nurs* 2006;**53**:205–12. http://dx.doi.org/10.1111/j.1365-2648.2006.03716.x

41. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;**331**:1064–5. http://dx.doi.org/10.1136/bmj.38636.593461.68

42. MacDermid J, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham I. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res* 2005;**5**:18. http://dx.doi.org/10.1186/1472-6963-5-18

43. Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Med Educ* 2010;**10**:12. http://dx.doi.org/10.1186/1472-6920-10-12

44. Greenhalgh T, Toon P, Russell J, Wong G, Plumb L, Macfarlane F. Transferability of principles of evidence based medicine to improve educational quality: systematic review and case study of an online course in primary health care. *BMJ* 2003;**326**:142–5. http://dx.doi.org/10.1136/bmj.326.7381.142

45. Wong G, Greenhalgh T, Russell J, Boynton P, Toon P. Putting your course on the web: lessons from a case study and systematic literature review. *Med Educ* 2003;**37**:1020–3. http://dx.doi.org/10.1046/j.1365-2923.2003.01673.x

46. Russell J, Greenhalgh T, Boynton P, Rigby M. Soft networks for bridging the gap between research and practice: illuminative evaluation of CHAIN. *BMJ* 2004;**328**:1174. http://dx.doi.org/10.1136/bmj.328.7449.1174

47. Floyd M, Pilling D, Garner K, Barrett P. Vocational rehabilitation: what works and in what circumstances. *Int J Rehabil Res* 2004;**27**:99–103. http://dx.doi.org/10.1097/01.mrr.0000127638.09376.cd

48. Connelly J, Duaso M, Butler G. A systematic review of controlled trials of interventions to prevent childhood obesity and overweight: a realistic synthesis of the evidence. *Public Health* 2007;**121**:510–17. http://dx.doi.org/10.1016/j.puhe.2006.11.015

49. McCormack B, Dewar B, Wright J, Garbett R, Harvey G, Ballantine K. *A Realist Synthesis of Evidence Relating to Practice Development: Final Report to NHS Education for Scotland and NHS Quality Improvement Scotland*. Edinburgh: NHS Education for Scotland; 2006 May.

50. Daykin N, Evans D, Petsoulas C, Sayers A. Evaluating the impact of patient and public involvement initiatives on UK health services: a systematic review. *Evid Policy* 2007;**3**:47–65. http://dx.doi.org/10.1332/174426407779702201

51. Birnik A, Bowman C. Marketing mix standardization in multinational corporations: a review of the evidence. *Int J Manag Rev* 2007;**9**:303–24. http://dx.doi.org/10.1111/j.1468-2370.2007.00213.x

52. Dieleman M, Gerretsen B, van der Wilt G. Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review. *Health Res Policy Syst* 2009;**7**:7. http://dx.doi.org/10.1186/1478-4505-7-7

53. Jackson L, Langille L, Lyons R, Hughes J, Martin D, Winstanley V. Does moving from a high-poverty to lower-poverty neighborhood improve mental health? A realist review of 'Moving to Opportunity'. *Health Place* 2009;**15**:961–70. http://dx.doi.org/10.1016/j.healthplace.2009.03.003

54. Wakerman J, Humphreys J, Wells R, Kuipers P, Entwistle P, Jones J. Primary health care delivery models in rural and remote Australia – a systematic review. *BMC Health Serv Res* 2008;**8**:276. http://dx.doi.org/10.1186/1472-6963-8-276

55. Meyer B, Haywood N, Sachdev D, Faraday S. *Independent Learning Literature Review (Research Report DCSF-RR051)*. Nottingham: DCSF Publications; 2008.

56. Kane S, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M. A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Serv Res* 2010;**10**:286. http://dx.doi.org/10.1186/1472-6963-10-286

57. O'Campo P, Kirst M, Schaefer-McDaniel N, Firestone M, Scott A, McShane K. Community-based services for homeless adults experiencing concurrent mental health and substance use disorders: a realist approach to synthesizing evidence. *J Urban Health* 2009;**86**:965–89. http://dx.doi.org/10.1007/s11524-009-9392-1

58. Waddington H, Snilsveit B, White H, Fewtrell L. *Water, Sanitation and Hygiene Interventions to Combat Childhood Diarrhoea in Developing Countries*. New Delhi: International Initiative for Impact Evaluation; 2009.

59. King E, Samii C, Snilstveit B. *Interventions to Promote Social Cohesion in Sub-Saharan Africa*. New Delhi: International Initiative for Impact Evaluation; 2010.

60. Leeman J, Chang Y, Lee E, Voils C, Crandell J, Sandelowski M. Implementation of antiretroviral therapy adherence interventions: a realist synthesis of evidence. *J Adv Nurs* 2010;**66**:1915–30. http://dx.doi.org/10.1111/j.1365-2648.2010.05360.x

61. Mazzocato P, Savage C, Brommels M, Aronsson H, Thor J. Lean thinking in healthcare: a realist review of the literature. *Qual Saf Health Care* 2010;**19**:376–82. http://dx.doi.org/10.1136/qshc.2009.037986

62. Walshe C, Luker K. District nurses' role in palliative care provision: a realist review. *Int J Nurs Stud* 2010;**47**:1167–83. http://dx.doi.org/10.1016/j.ijnurstu.2010.04.006

63. McMahon T. *A Realist Review of Evidence to Guide Targeted Approaches to HIV/AIDS Prevention Among Immigrants Living in High-income Countries*. PhD thesis. Adelaide, Australia: Flinders University; 2010.

64. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Infirm* 2010;**79**:736–71. http://dx.doi.org/10.1016/j.ijmedinf.2010.08.006

65. Javanparast S, Ward P, Young G, Wilson C, Carter S, Misan G, *et al.* How equitable are colorectal cancer screening programs which include FOBTs? A review of qualitative and quantitative studies. *Prev Med* 2010;**50**:165–72. http://dx.doi.org/10.1016/j.ypmed.2010.02.003

66. Morgan G. Evidence-based health policy: a preliminary systematic review. *Health Educ J* 2010;**69**:43–7. http://dx.doi.org/10.1177/0017896910363328

67. McLean AM, Koppang J. Behavioral caregiving for adults with traumatic brain injury living in nursing homes: developing a practice model. *J Theory Constr Test* 2010;**14**:17–22.

68. Ware V-A, Gronda H, Vitis L. *Addressing Locational Disadvantage Effectively*. Melbourne, Australia: Australian Housing and Urban Research Institute Research Synthesis Service; 2010.

69. Dieleman M, Kane S, Zwanikken P, Gerretsen B. *Realist Review and Synthesis of Retention Studies for Health Workers in Rural and Remote Areas*. Geneva: World Health Organization; 2011.

70. Wong G, Pawson R, Owen L. Policy guidance on threats to legislative interventions in public health: a realist synthesis. *BMC Public Health* 2011;**11**:222. http://dx.doi.org/10.1186/1471-2458-11-222

71. O'Campo P, Kirst M, Tsamis C, Chambers C, Ahmad F. Implementing successful intimate partner violence screening programs in health care settings: evidence generated from a realist-informed systematic review. *Soc Sci Med* 2011;**72**:855–66. http://dx.doi.org/10.1016/j.socscimed.2010.12.019

72. Carr S, Lhussier M, Forster N, Geddes L, Deane K, Pennington M, *et al.* An evidence synthesis of qualitative and quantitative research on component intervention techniques, effectiveness, cost-effectiveness, equity and acceptability of different versions of health-related lifestyle advisor role in improving health. *Health Technol Assess* 2011;**15**(9). http://dx.doi.org/10.3310/hta15090

73. Davies JK, Sherriff N. The gradient in health inequalities among families and children: a review of evaluation frameworks. *Health Policy* 2011;**101**:1–10. http://dx.doi.org/10.1016/j.healthpol.2010.09.015

74. Gunawardena I. Effectiveness of the geriatric day hospital – a realist review. *Rev Clin Gerontol* 2011;**21**:267–9.

75. Harris J, Kearley K, Heneghan C, Meats E, Roberts N, Perera R, *et al.* Are journal clubs effective in supporting evidence-based decision making? A systematic review. BEME Guide No. 16. *Med Teach* 2011;**33**:9–23. http://dx.doi.org/10.3109/0142159X.2011.530321

76. Jacobs S, Ashcroft D, Hassell K. Conducting a realist review of a complex concept in the pharmacy practice literature: methodological issues. *Int J Pharm Prac* 2010;**18**(Suppl. 1):1–2.

77. Price S, Long AF, Godfrey M, Thomas KJ. Getting inside acupuncture trials – exploring intervention theory and rationale. *BMC Complement Altern Med* 2011;**11**:22. http://dx.doi.org/10.1186/1472-6882-11-22

78. Toohey A, Rock M. Unleashing their potential: a critical realist scoping review of the influence of dogs on physical activity for dog-owners and non-owners. *Int J Behav Nutr Phys Act* 2011;**8**:46. http://dx.doi.org/10.1186/1479-5868-8-46

79. Vassilev I, Rogers A, Sanders C, Kennedy A, Blickem C, Protheroe J, *et al.* Social networks, social capital and chronic illness self-management: a realist review. *Chronic Illn* 2011;**7**:60–86. http://dx.doi.org/10.1177/1742395310383338

80. Masuda J, Zupancic T, Poland B, Cole D. Environmental health and vulnerable populations in Canada: mapping an integrated equity-focused research agenda. *Can Geogr* 2008;**52**:427–50. http://dx.doi.org/10.1111/j.1541-0064.2008.00223.x

81. Greenhalgh T, Potts HW, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. *Milbank Q* 2009;**87**:729–88. http://dx.doi.org/10.1111/j.1468-0009.2009.00578.x

82. Addis S, Davies M, Greene G, MacBrideStewart S, Shepherd M. The health, social care and housing needs of lesbian, gay, bisexual and transgender older people: a review of the literature. *Health Soc Care Community* 2009;**17**:647–58. http://dx.doi.org/10.1111/j.1365-2524.2009.00866.x

83. Collins P, Hayes M. The role of urban municipal governments in reducing health inequities: a meta-narrative mapping analysis. *Int J Equity Health* 2010;**9**:13. http://dx.doi.org/10.1186/1475-9276-9-13

84. Contandriopoulos D, Lemire M, Denis J-L, Tremblay E. Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature. *Milbank Q* 2010;**88**:444–83. http://dx.doi.org/10.1111/j.1468-0009.2010.00608.x

85. Greenhalgh T, Heath I. Measuring quality in the therapeutic relationship Part 1: objective approaches. *Qual Saf Health Care* 2010;**19**:475–8. http://dx.doi.org/10.1136/qshc.2010.043364

86. Greenhalgh T, Heath I. Measuring quality in the therapeutic relationship Part 2: subjective approaches. *Qual Saf Health Care* 2010;**19**:479–83. http://dx.doi.org/10.1136/qshc.2010.043372

87. Kitson A, Conroy T, Wengstrom Y, Profetto-McGrath J, Robertson-Malt S. Defining the fundamentals of care. *Int J Nurs Pract* 2010;**16**:423–34. http://dx.doi.org/10.1111/j.1440-172X.2010.01861.x

88. Gagliardi A, Brouwers M, Palda V, Lemieux-Charles L, Grimshaw J. How can we improve guideline use? A conceptual framework of implementability. *Implement Sci* 2011;**6**:26. http://dx.doi.org/10.1186/1748-5908-6-26

89. Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: realist syntheses. *BMC Med* 2013;**11**:21. http://dx.doi.org/10.1186/1741-7015-11-21

90. Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: realist syntheses. *J Adv Nurs* 2013;**69**:1005–22. http://dx.doi.org/10.1111/jan.12095

91. Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: meta-narrative reviews. *BMC Med* 2013;**11**:20. http://dx.doi.org/10.1186/1741-7015-11-20

92. Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: meta-narrative reviews. *J Adv Nurs* 2013;**69**:987–1004. http://dx.doi.org/10.1111/jan.12092

93. Wong G, Westhorp G, Greenhalgh T, Pawson R. *The RAMESES Project: Project Outputs.* 2013. URL: www.ramesesproject.org/index.php?pr=Project_outputs (accessed 19 August 2014).

94. Pawson R, Manzano-Santaella A. A realist diagnostic workshop. *Evaluation* 2012;**18**:176–91. http://dx.doi.org/10.1177/1356389012440912

95. Pawson R. *The Science of Evaluation: a Realist Manifesto*. London: Sage; 2013.

96. Eccles M, Armstrong D, Baker R, Cleary K, Davies H, Davies S, *et al.* An implementation research agenda. *Implement Sci* 2009;**4**:18. http://dx.doi.org/10.1186/1748-5908-4-18

97. Gough D, Oliver S, Thomas J. *An Introduction to Systematic Reviews*. London: Sage; 2012.

98. Pluye P, Hong Q. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annu Rev Public Health* 2014;**35**:29–45. http://dx.doi.org/10.1146/annurev-publhealth-032013-182440

99. Greenhalgh T, Wong G, Westhorp G, Pawson R. Protocol – Realist and Meta-narrative Evidence Synthesis: Evolving Standards (RAMESES). *BMC Med Res Methodol* 2011;**11**:115. http://dx.doi.org/10.1186/1471-2288-11-115

Appendix 1 Realist and meta-narrative reviews identified from initial exploratory searches

Connelly J, Duaso M, Butler G. A systematic review of controlled trials of interventions to prevent childhood obesity and overweight: a realistic synthesis of the evidence. *Public Health* 2007;**121**:510–17.

Dieleman M, Gerretsen B, van der Wilt G. Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review. *Health Res Policy Syst* 2009;**7**.

Dieleman M, Kane S, Zwanikken P, Gerretsen B. *Realist Review and Synthesis of Retention Studies for Health Workers in Rural and Remote Areas*. Geneva: World Health Organization; 2011.

Greenhalgh T, Kristjansson E, Robinson V. Realist review to understand the efficacy of school feeding programmes. *BMJ* 2007;**335**:858–61.

Jackson L, Langille L, Lyons R, Hughes J, Martin D, Winstanley V. Does moving from a high-poverty to lower-poverty neighborhood improve mental health? A realist review of 'Moving to Opportunity'. *Health Place* 2009;**15**:961–70.

Kane S, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M. A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Serv Res* 2010;**10**.

King E, Samii C, Snilstveit B. *Interventions to Promote Social Cohesion in Sub-Saharan Africa*. New Delhi: International Initiative for Impact Evaluation; 2010.

Leeman J, Chang Y, Lee E, Voils C, Crandell J, Sandelowski M. Implementation of antiretroviral therapy adherence interventions: a realist synthesis of evidence. *J Adv Nurs* 2010;**66**:1915–30.

McCormack B, Dewar B, Wright J, Garbett R, Harvey G, Ballantine K. *A Realist Synthesis of Evidence Relating to Practice Development: Final Report to NHS Education for Scotland and NHS Quality Improvement Scotland*. Edinburgh: NHS Education for Scotland; 2006.

O'Campo P, Kirst M, Schaefer-McDaniel N, Firestone M, Scott A, McShane K. Community-based services for homeless adults experiencing concurrent mental health and substance use disorders: a realist approach to synthesizing evidence. *J Urban Health* 2009;**86**:965–89.

Waddington H, Snilsveit B, White H, Fewtrell L. *Water, Sanitation and Hygiene Interventions to Combat Childhood Diarrhoea in Developing Countries*. New Delhi: International Initiative for Impact Evaluation; 2009.

Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Med Educ* 2010;**10**:12.

Wong G, Pawson R, Owen L. Policy guidance on threats to legislative interventions in public health: a realist synthesis. *BMC Public Health* 2011;**11**.

Appendix 2 Briefing document for realist and meta-narrative review Delphi panel

RAMESES

Realist and Meta-narrative Evidence Synthesis: Evolving Standards

Delphi Panel Briefing Document: developing publication standards for realist and meta-narrative reviews

Trish Greenhalgh, Geoff Wong, Gill Westhorp, Ray Pawson

CONTENTS

What we would like you to do, how and when

The task is to produce consensus publication standards for two sorts of systematic review: realist and meta-narrative. You have agreed to be a member of our Delphi panel. A Delphi panel is a way of working towards consensus on a topic or question. It consists of a number of rounds. In a preliminary round, you will be asked to suggest topics which you would like to see covered (or statements you would like to see included). In each subsequent round (usually two more), you will be asked to do a task which involves *scoring* a draft set of statements. There will be a deadline for this, because we can't analyse the responses until everyone has replied.

After each *scoring* round, you will be sent your own scores *and* the average score for everyone in the group. If you find you are an 'outlier', you have two choices: amend your score (after reflecting on the statement and why you scored it as you did) – or stand your ground and argue your case to the group (they won't know how you scored the statement). Even if you scored a statement similarly to the group average, you may be swayed to change your score by arguments put subsequently.

Each statement is scored on two dimensions: [a] relevance (should we include this topic / theme at all?) and [b] content (should we word it like this?). High scores for relevance *and* content mean the statement will be included 'as is'. High scores for relevance but low scores for content means we need to word the statement differently (we'll ask for suggestions). Low scores for relevance mean the statement gets dropped. But when some panel members score a statement high and others score it low, we need a discussion. For references on the validity and methodology of the Delphi process, please ask us.

Here's what we'd like you to do:

- Pull out now if you've changed your mind (so you don't count as a 'withdrawal')
- For ROUND 1, please read this background paper (and, if you've got time, the full study protocol and the other documents we have provided)
- Respond within one month to Geoff *only* by hitting the reply button with your suggestions.
- Wait while we analyse all the responses and build the draft statements
- Respond to the ROUND 2 email (expected mid-October 2011) within one month by looking at the statements and entering your scores for each (we'll give you a link to an online questionnaire)
- Wait again while we analyse the data and send you back your scores
- Join in an email discussion on how we might amend the statements
- Repeat the last three steps for ROUND 3 (expected late November 2011)

This Delphi panel is part of the wider RAMESES project, which has six work packages: [a] produce publication standards for realist and meta-narrative reviews; [b] refine and extend existing methodological guidance; [c] develop, pilot and run training modules; [d] run a JISCmail discussion list (www.jiscmail.ac.uk/RAMESES); [e] support teams undertaking reviews; and [f] contribute to the academic literature (e.g. on the methodology of doing [a] to [e]). The RAMESES study protocol is appended.

Authorship policy

We want to acknowledge the input of everyone who contributes to RAMESES. We propose two levels of authorship:

a. People who contribute materially and significantly to conceptualising the study, undertaking the research, analysing the data or writing up will be named as co-authors alongside us on publications. The format of the author list will be "Smith A, Jones B, Bloggs D on behalf of the RAMESES group".
b. Members of the Delphi panel who do not fulfil the above criteria will be listed directly below the authors in the following format: "The RAMESES group comprised: Aardvark H, Bloggs D ...etc to Zindel B".

Please let us know if you are looking for a formal authorship role or if at any stage you believe you deserve to join the author list. We will also be alert to input from Delphi panel members above and beyond what is expected of an ordinary participant. It is quite possible that the RAMESES statement will have a large number of authors and we are comfortable with that.

Whatever your level of input to this project, you won't get paid unless you were costed on the grant application. Nevertheless your input is greatly valued.

Briefing on meta-narrative reviews

Background

Meta-narrative review is a new method of systematic review, designed for topics which have been *differently conceptualised and studied* by different groups of researchers. Here's an example. Many groups have studied the building of dams in India. Some have conceptualised this dam-building as engineering; others as colonialism; others as a threat (or promise) to the local eco-system; others as inspiration for literature and drama, and so on. If we were to summarise this topic area in a way that was faithful to what each different group set out to do, we would have to start by asking how each of them approached the topic, what aspect of 'dams in India' they chose to study and how. In order to understand the many approaches, we would have to consciously and reflexively *step out of* our own world-view, learn some new vocabulary and methods, and try to view the topic of 'dams in India' through multiple different sets of eyes. When we had begun to understand the different perspectives, we could summarise them in an over-arching narrative, highlighting what the different research teams might learn from one another's approaches.

(NOTE: some reviewers might be interested only in summarising the findings of randomised controlled trials of 'dam present' versus 'dam absent' on a predefined outcome, and if that was the focus of the review, a Cochrane review with statistical meta-analysis would be the gold standard approach. The meta-narrative approach is only intended for those reviews where the underlying research goal is to identify and explore the *diversity* of research approaches to a topic.)

Methodological issues in meta-narrative review

The methodology of meta-narrative review was developed by Trish Greenhalgh and her team in 2004 when reviewing the literature on diffusion of service-level innovations in healthcare.[1] A methods paper was published in Social Science and Medicine in early 2005.[2] The inspiration for this method was Kuhn's 1962 book *The Structure of Scientific Revolutions*, which argued that science progresses in paradigms (i.e. particular ways of viewing the world, including assumptions about how the world works) and that one scientific paradigm gives way to another as scientific progress renders yesterday's assumptions and practices obsolete.[3] Newton's theories and methods, for example, became less and less able to answer the emerging questions of particle physics, leading Einstein to develop his theory of relativity. Meta-narrative review looks historically at how particular research traditions have unfolded over time and shaped the kind of questions being asked and the methods used to answer them. A research tradition is a series of linked studies, each building on what has gone before and taking place within a coherent paradigm (that is, within a shared set of assumptions and preferred methodological approach shared by a group of scientists).

While researching the background for the RAMESES project, we came across *meta-triangulation review*, another synthesis method described by Marianne Lewis and Andrew Grimes in the Academy of Management Journal in 1999.[4] We had been unaware of this approach when we published our original work on meta-narrative review but have subsequently communicated with Prof Lewis, who has offered her input to the RAMESES study (subject to other commitments). The (many) similarities and (few) differences

between these two approaches are shown in Appendix 1. When collecting meta-narrative reviews for the RAMESES study, we looked at examples of meta-triangulation reviews but decided to exclude these because meta-triangulation review seeks to understand and analyse topics at a the level of theoretical differences between paradigms, whereas meta-narrative review is more interested in the research tradition as its unit of analysis (a working definition of a research tradition is "what researchers get up to *within* a paradigm"). However, the methodology of meta-triangulation review offers some transferable insights which will help us refine the quality criteria for a meta-narrative review (Appendix 1).

Summary of published examples of meta-narrative reviews

With the help of a specialist informaticist/librarian (Jeanette Buckingham), we identified a sample of 9 published papers which were described as meta-narrative reviews. These were examined independently by Geoff and Trish. As expected, the 9 reviews covered a range of complex topic areas which had been differently studied by different groups of researchers (e.g. electronic patient records, environmental health, fundamentals of nursing care, knowledge translation and exchange). Most were published after 2009, and we know of several more reviews which are ongoing or in press. We considered that five of our sample of 9 were "true" meta-narrative reviews, defined by three working criteria: [a] the authors clearly understood the need to consider the topic from multiple paradigmatic perspectives and used the research tradition (or something comparable to it) as their unit of analysis; [b] the authors made efforts to step out of their own world-view and recognise and value alternative world-views; [c] the synthesis included a comparison of how the topic area was approached from at least two contrasting perspectives. Another of the 9 reviews appeared to "almost" meet these criteria. Three papers described as meta-narrative reviews did not meet even these fairly loose criteria.

Preliminary thoughts on publication standards for meta-narrative reviews

Our analysis of these published reviews, along with our discussions with review teams who are currently undertaking meta-narrative reviews, have surfaced the following issues and implications for the RAMESES project. These are preliminary – we hope the Delphi panel members will add to and/or challenge them.

1. TERMINOLOGY. Key terms were used inconsistently by review teams (partly because we had omitted to define some of them in our original methods papers).

   => *We need a glossary and set of definitions.*

2. PHILOSOPHICAL BASIS. The philosophical assumptions of meta-narrative review (e.g. Kuhn's notion that science progresses in paradigms, and a paradigm is a particular conceptual lens which shapes what counts as knowledge) appear to be widely misunderstood. Misunderstanding or undervaluing the importance of the philosophical basis of meta-narrative reviews and its implications appeared to lead to mis-application of the method.

   => *We need to find ways of making the philosophy accessible and its implications clear.*

3. CLASSIFICATION. Some review teams did not appear to understand the fundamental difference between the meta-narrative method and an old-fashioned narrative review. The term "meta-narrative" seemed to be used as a synonym for a form of thematic analysis or on one which offered an under-theorised "thematic analysis" (by which was meant that the findings section listed the themes raised by the empirical papers).

   => We need to include very clear criteria for classifying a review as "meta-narrative" and an alert that the term is sometimes misused.

4. UNIT OF ANALYSIS. Some review teams were confused about "the research tradition" as the unit of analysis.

   => We need to clarify what a research tradition is <u>and</u> the importance of inter-disciplinary working.

5. TITLE. Some but not all meta-narrative reviews were described as such in the title.

   => We need to encourage authors to do this.

6. RESEARCH QUESTION. Some review teams either asked no research question or asked a question which did not seem amenable to being answered using meta-narrative review. Successful reviews had sought to make sense of a topic by appropriately adapting of one or more of the 'generic' questions that underlie the meta-narrative method (How has the topic been conceptualised in each separate research tradition? What are the key concepts, theories, assumptions? What are the preferred study designs and ways of knowing? What are the main empirical findings? What can we learn from the range of different approaches?)

   => We need to clarify what a research question (and sub-questions) would look like in a meta-narrative review. We also need to highlight the kinds of questions which are UNsuited to this kind of review.

7. METHODS. Some review teams appear to have cut and pasted the methods section from a published meta-narrative review virtually verbatim into their own paper, thus *claiming* to have followed all the recommended steps even when it was clear that they had not. This suggests that some journal editors and peer reviewers are unable to judge whether the method is being followed or not. Some review teams described a "modified" meta-narrative approach but did not say how and why they modified it.

   => We need to include techniques for confirming that the methods were actually followed and an alert to the cut-and-paste ruse. We need to include the instruction that if teams modify the method, they have to say <u>how and why</u> they modified it.

8. SEARCHING. A number of reviews did not undertake any iterative searching. Rather, they used a one-off, predefined search strategy (as is standard in many Cochrane reviews). But because meta-narrative review is aimed at making sense of the

literature, it may only become clear which data are needed as a review progresses. Hence searching needs to be not only iterative, but also purposive and flexible (e.g. in terms of inclusion criteria).

=> We need to encourage review teams to begin with a broad, "browsing" search and progressively refine this in the light of emerging data. If iterative searching is not undertaken, such a decision should be justified.

9. FOCUSING THE REVIEW. Several teams reported difficulties in making the review manageable within the time and resource available. This is an inherent problem since meta-narrative review is designed to make sense of large and contested bodies of literature. Where reported, containment and focusing was achieved through discussion within the review team and with reference to interested parties (e.g. service users, experts in the field).

   => In the publication standards, we need to expect a statement of how the review was shaped and contained. In the methodological advice we need to suggest approaches to this.

10. APPRAISING PRIMARY STUDIES. Review teams appraised studies in different ways. Some used a flexible approach, using judgement to include (or exclude) and appraise studies in an iterative manner as their data extraction and synthesis unfolded. Others preferred to develop a formal list of inclusion and exclusion criteria and used this systematically (and somewhat inflexibly) to rule papers in or out of their dataset *before* reading them in detail. We strongly favour the former method, which aligns with the interpretive basis of meta-narrative review. Meta-narrative review assesses studies using the quality standards accepted within a paradigm. The purpose of this is to aid the sense-making process as it allows (for example) a review team to say "in this tradition, the X, Y and Z are considered to be high quality studies".

    => We need to encourage an interpretive and iterative approach to assessing primary studies for inclusion.

11. FINDINGS. Some review teams did not provide sufficient detail to support the inferences in their findings section.

    => We need to include clear guidance on how we expect review teams to present and justify their findings in a way that allows others to judge their coherence and plausibility.

12. CONCLUSIONS. Some but not all teams provided a clear line of reasoning linking findings to conclusions and recommendations.

    => We need to require conclusions should be 'traceable' back to detailed presentation of the mapping and analysis of the research traditions and their underpinning paradigms.

Briefing on realist reviews

Background

The realist research question is often summarised as "What works for whom under what circumstances, how and why)?" Realist inquiry considers the interaction between context, mechanism and outcome. In a realist world, intervention X is *not* thought of as having effect size Y with confidence interval Z. Rather, intervention X (e.g. a programme introduced by policymakers who seek to create a particular outcome) alters context (for example by making new resources available), which then triggers mechanism(s) which produce both intended and unintended outcomes. X may work very will in one context but poorly or not at all in another context.

Realist inquiry seeks to unpack the context-mechanism-outcome relationship, thereby *explaining* examples of success, failure, and various eventualities between. Theoretical explanations of this kind are referred to as "middle-range theories" (i.e. ones which *"...involve abstraction... but [are] close enough to observed data to be incorporated in propositions that permit empirical testing".*[5]

The basis of realist inquiry is a realist philosophy, whose key tenets are as follows (feel free to challenge these – this is just to get us going):

1. There is a [social] reality which can't be measured directly (because it is processed through our brains, language, culture and so on) but can be known indirectly. Realism thus sits, broadly speaking, between positivism ('there is a real world which we can apprehend directly through observation') and constructivism ('given that all we can know has been interpreted through human senses and the human brain, we cannot know for sure what the nature of reality is').
2. Social programmes (including complex interventions) may change the social context (for example by introducing legislation) or may change the resources or opportunities available to participants and, in that sense, change the *context* for those participants.
3. To understand the relationship between context and outcome, realism uses the concept of *mechanisms*, defined as "...underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest." [6]

The realist approach has informed empirical studies (realist evaluation),[7] and offers the potential for insights (e.g. in relation to complex interventions and the implementation of research findings) that go beyond the narrowly experimental paradigm of the randomised controlled trial.[8]

Methodological issues in realist review

"Realist synthesis" was first described by Ray Pawson in 2002,[9] updated in an ESRC-commissioned monograph in 2004[10], published as a book in 2006[11] and summarised in a short methods paper in the Journal of Health Services Research and Policy in 2005.[12]

A realist review (or realist synthesis) applies realist philosophy to the synthesis of findings from primary studies that have a bearing on a single research question. Reviews begin by eliciting from the literature the main ideas that have gone into the making of a class of interventions (the programme theory). This programme theory can be thought of as setting out how and why a class of intervention is though to 'work' to generate the outcome(s) of interest. The pertinence and effectiveness of each constituent idea is then tested using the available evidence (qualitative, quantitative, comparative, administrative, etc.) that has gathered in the primary literature on that family of programmes. In this testing, the ideas within a programme theory are re-cast and conceptualised in realist terms and for each idea reviewers have to seek out the contextual (C) influences that have triggered the relevant mechanism(s) (M) to generate the outcome(s) (O) of interest. Synthesis consists of comparing 'how the programme was supposed to operate' to the 'empirical evidence on the actuality' – all along CMO lines. Analytic purchase comes from the ability to describe and understand the many contingencies that need to be put in place (or avoided) to improve the likelihood of such interventions generating their intended outcomes – in other words, explaining how an intervention might change the context or provide resources in such a way as to most likely trigger the right mechanism(s) to produce the desired outcome.

Summary of published examples of realist reviews

We identified a sample of 35 published papers which were described as realist reviews. These were examined in detail by Geoff, and aspects of his analysis checked by Trish, Ray and Gill. They were published between 2004 and 2011 and covered a broad range of topics (e.g. health, education, human resources). We classified 8 of these 35 as high-quality realist reviews, five as having many but not all features of a high-quality realist reviews and 22 lacking many substantial aspects of a realist analysis. Our classification of these reviews was based on our judgment of whether a realist analysis (the application of realist logic and concepts in a review and synthesis) had been undertaken.

Preliminary thoughts on publication standards for realist reviews

Our analysis of these published reviews, along with our discussions with review teams who are currently undertaking realist reviews, have surfaced the following preliminary issues and implications for the RAMESES project. Many of these are similar to the problems found in meta-narrative reviews.

1. TERMINOLOGY. Key terms were used inconsistently by review teams (partly because even when we defined them clearly in our own publications, other conflicting definitions exist in the literature or teams chose to redefine the concepts themselves).

   => *We need a glossary and set of definitions (and we need to make sure we don't just privilege what we ourselves have written before).*

2. CLASSIFICATION. Currently the number of reviews which we have judged to have significant limitations from a realist perspective outnumber those which we have judged to be robust realist reviews.

=> We need an accessible way of determining the quality of realist analysis within realist reviews.

3. PHILOSOPHICAL BASIS. In our judgement, the commonest flaw in our sample of published realist reviews was lack of appreciation of the philosophical basis of realism and the implications of this for the review methodology. These reviews used the term "realist" to mean (variously) "qualitative", "narrative", "non-Cochrane" and/or did not explain its methods in detail.

   => We need to explain the philosophy and its implications for the methodology. We need to devise ways in which assessors can determine whether review teams understood the implications of realist philosophy and its application.

4. THE REVIEW TEAM. High-quality realist reviews tended to have been undertaken by a team of reviewers with relevant methodological expertise. This appeared to have allowed them to "bounce ideas off" each other to focus the review and apply the realist logic of analysis appropriately and effectively.

   => We need to highlight the opportunities offered by working in a review team.

5. RESEARCH QUESTION. Some review teams either asked no research question or asked a question which did not seem amenable to being answered using realist review (e.g. one that could not be mapped to the generic question "what works, for whom, in what circumstances, to what extent, how and why?").

   => We need to clarify what a research question (and sub-questions) would look like in a realist review. We also need to highlight the kinds of questions which are UNsuited to this kind of review.

6. METHODS. There was a mismatch between what review teams said they had done and what the findings section suggested had actually been done. Sometimes, a review would explain that realist review had been chosen as the preferred method for one of the following reasons; heterogeneous data, a wish to synthesise quantitative and qualitative data, and/or to address the "what works for whom and in what circumstances" question. An explanation would then be provided as to what the realist review method is. This section (if provided) often suggested that a realist analysis had been undertaken, but the findings would consist of a thematic and/or narrative synthesis. In some reviews which made claims to be "realist", realist concepts were not mentioned at all or incorrectly conceptualised (e.g. mechanism was confused with intervention). Some review teams described a "modified" realist approach but did not say how or why they modified it.

   =>We need to include techniques for confirming that realist methods were actually followed and an alert to the cut-and-paste ruse. We need to include the instruction that if teams modify the method, they have to say <u>how</u> they modified it and <u>why</u>.

7. SEARCHING AND INCLUSION CRITERIA. A number of reviews did not undertake any iterative searching. Whilst this may not always be necessary, it is highly likely that as the process of theory-building and theory-testing progresses, additional searching will be needed after the initial papers have been identified. Some of the realist reviews in our sample searched for and included only randomised controlled trials (RCTs), and these found that they had too little detail in their included studies to build and test theory.

   => We need to encourage review teams to search iteratively, purposively and continuously throughout the review, and refine searches in the light of emerging data. We should discourage including only RCTs in a realist review, since the data needed to enable reviewers to make coherent and plausible theoretical inferences can come from a wide variety of sources.

8. FOCUSING THE REVIEW. Several teams reported difficulties in making the review manageable within the time and resource available. This is not surprising since a realist review on any topic is potentially endless as more and more refined explanations are sought to explain increasingly diverse aspects of a review. Where reported, as with meta-narrative review, containment and focusing were achieved through discussion within the review team and with reference to interested parties (e.g. decision makers, experts in the field).

   => In the reporting standards, we need to expect a statement of how and why the review was shaped and contained. In the methodological advice we need to suggest approaches to this.

9. THEORISING. In reviews which had some but not all characteristics of a robust realist review, what was missing was the use of theory to try to provide an overarching coherent and plausible explanation of the observed patterns of outcomes. An important aspect of this theory development is that it is iteratively tested against the reported data in the included studies.

   => We need to include an expectation for this type of theorising in the publication standards (and explain how to do it in the methodological guidance and training materials).

10. APPRAISING PRIMARY STUDIES. Review teams appraised studies in different ways. Some used a flexible approach, implicitly or explicitly following Pawson's judgement-dependent criteria of "relevance" and "rigour", and appraised studies in parallel with their data extraction and synthesis. Others preferred to apply a formal critical appraisal checklist and used questions on this checklist as a tool for excluding studies before undertaking the detailed synthesis work. We strongly favour the former method, which aligns with the explanatory basis of realist synthesis. Quality appraisal may need to be iterative, because as the process of theory-building and/or refinement unfolds, a different section of an included study may yield relevant data.

=> We need to encourage approaches that assess the relevance and rigour of primary studies for inclusion. Studies should be included in a review if they are able to make a contribution to theory building or testing (relevance). We should caution against excluding studies based on an overall assessment of their 'quality', since small segments of included studies may contain useful insights for theory-building even when other aspects of them are flawed. We should encourage a parallel rather than sequential approach to appraising papers and synthesising the insights from them.

11. FINDINGS. Some review teams did not provide sufficient detail to support the inferences in their findings section.

    => We need to include clear guidance on how we expect review teams to present and justify their findings in a way that allows others to judge their coherence and plausibility.

12. RECOMMENDATIONS. Few studies contained sufficient detail on contextual influences. The explanations in realist reviews are highly dependent on contextual influences. It follows that recommendations must be contingent (for example only under certain contexts will a particular mechanism be triggered to generate the desired outcome) rather than a list of "dos and don'ts".

    => We need to stipulate the format of recommendations in a relist review (e.g. "In pursing programme theory A, attend to the following contingencies, context and implementation features B, C. D, E, ….N.").

Appendix: Meta-triangulation vs. meta-narrative review

	META-TRIANGULATION REVIEW[4]	META-NARRATIVE REVIEW[2]
Purpose	To build theory. "Studying multifaceted phenomena characterized by expansive and contested research domains"	To build a rich, multifaceted picture of a complex topic, especially when a summary is needed for policy decisions
Philosophical basis	Constructivist (Kuhn's philosophy of science)	Constructivist (Kuhn's philosophy of science)
Intended audience	Academics	Policymakers
Type of insights	Analytic	Predominantly descriptive but recognises potential for analytic, theory-building insights
Examples of topics reviewed	Theoretical topics at high level of abstraction e.g. power, strategy	Policy and/or practice-relevant topics e.g. electronic records, knowledge translation
Empirical data	Included only as an aid to theorising	Included as substantive component of review
Unit of analysis	**Paradigm:** "the assumptions, practices and agreements among a scholarly community"	**Research tradition:** the historical unfolding of research on a particular theme by a group of scientists, which occurs within a paradigm
Key stages	**GROUNDWORK** Define phenomenon of interest Focus paradigmatic lenses **SEARCH** Collect data interpretable from multiple lenses **MAPPING PARADIGMS** Plan paradigm itinerary (ordered use of different paradigmatic lenses) Code data Write paradigm accounts **THEORY BUILDING** Explore metaconjectures Attain meta-paradigm perspective Reflect critically on the process	**GROUNDWORK** Assemble multidisciplinary team Outline research question Agree outputs with funder **SEARCH** Browse literature to identify research traditions Search within each tradition to identify seminal conceptual and theoretical papers Search systematically for empirical papers **MAPPING RESEARCH TRADITIONS** Describe paradigmatic basis for each tradition Highlight the 'storyline' of each tradition (key issues and discoveries as they unfolded) Appraise and summarise primary studies **SUMMARY / SYNTHESIS** Summarise each research tradition separately, highlighting similarities and differences View discrepancies as higher-order data; explain as contestation between paradigms **RECOMMENDATIONS** Consider implications for sponsor / audience
Principles and approaches to assure quality of the review	**Reflexivity:** Theorist should be fully aware of own assumptions **Systematic cross-paradigm synthesis techniques:** e.g. paradigm bridging (seeking commonalities), paradigm bracketing (highlighting differences), interplay (exploring tensions); meta-theorizing (exploring patterns that span conflicting understandings)	**Pragmatism:** What to include is not self evident **Pluralism:** Include multiple perspectives and ask what we can learn from each **Historicity:** Trace research traditions over time **Contestation:** Use "conflicting findings" in a positive way to generate new insights **Peer review:** Present emerging findings periodically to a critical external audience

Reference List

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

(2) Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. Soc Sci Med 2005; 61(2):417-430.

(3) Kuhn TS. The structure of scientific revolutions. Chicago: University of Chicago Press; 1962.

(4) Lewis MW, Grimes AJ. Meta-triangulation: Building theory from multiple paradigms. Academy of Management Review 1999; 24(4):672-690.

(5) Merton.R. On Theoretical Sociology. Five Essays, Old and New. New York: Free Press; 1967.

(6) Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. American Journal of Evaluation 2010; 31:363-381.

(7) Pawson R, Tilley N. Realistic evaluation. London: Sage; 1997.

(8) Berwick DM. The science of improvement. JAMA 2008; 299(10):1182-1184.

(9) Pawson R. Evidence-based policy: the promise of 'realist synthesis'. Evaluation 2002; 8(3):340-358.

(10) Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist synthesis - an introduction. ESRC Working Paper Series. London: ESRC; 2004.

(11) Pawson R. Evidence-based policy: A realist perspective. London: Sage; 2006.

(12) Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005; 10 Suppl 1:21-34.

Appendix 3 'Paper' version of round 2 online Delphi panel survey for realist reviews

RAMESES Delphi - Round 2

Introduction

Thank you for continuing to help us with the RAMESES project.

In Round 1 of our Delphi process, we had asked panel members for suggestions of Items to include in the RAMESES publication standards for meta-narrative and realist syntheses – 'RAMESES statement'. What we hope to produce are publication standards rather than detailed guidance on how to conduct a meta-narrative or realist synthesis. Your comments related to how to conduct reviews have however been captured for later use when we develop our training materials. We hope to make our standards relevant to researchers, journal editors, peer-reviewers and funders.

We have collated all your responses and compiled a list of potential Items for inclusion in the 'RAMESES statement'. In Round 2, we would be grateful if you would please rate each Item for:

- Relevance (should we include an Item on this theme/topic at all?)
- Content (should we word this Item like this?)

There will be a free text box for you to make comments on any aspect of an Item. To help you understand why an Item has been included we have also provided a brief explanation.

This survey will take you between 30 to 60 minutes to complete.

You may at any time stop and return to where you left off by clicking on the unique web link you were sent inviting you to take part in this survey. You may also go back to previous items if you wish.

We would be most grateful if you would please try to complete the survey by Sunday 8th January 2012 at the latest.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

PART 1 - Realist Synthesis

The questions in PART 1 cover potential Items for inclusion in the RAMESES publication standards for Realist Synthesis only.

The first four Items are topics for consideration in the Introductory section.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 1: Title - Realist Synthesis

Item 1: Title

In the title, identify the document as a Realist Synthesis or Review.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Our background searching has shown that some realist reviews are not flagged as such in the title and may also be inconsistently indexed, and hence are more difficult to locate during searching. The terms realist synthesis and realist review are both in widespread use. Consistent use of one term is likely to aid indexing and identification.
We are interested to find out which term you prefer and why.

Please choose your preferred term:

◯ Realist Synthesis

◯ Realist Review

◯ No preference

Reason(s) for choice of term (optional):

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 2: Abstract - Realist Synthesis

Item 2: Abstract

As far as possible taking account of journal-specific formatting and content requirements, the abstract should contain brief details of the study context, review question or objectives; search strategy; selection and appraisal of documents; analysis and synthesis methods; results; and conclusions/implications.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Apart from the title, an abstract is the only source of information accessible to searchers unless the full paper is obtained. The information in it must allow reviewers and/or users to decide if the review is relevant to their needs.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 3: Rationale for review - Realist Synthesis

Item 3: Rationale for review

Explain why the review was done and what it is likely to add to existing understanding of the topic area.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
As with all research, a background section explaining what is already known and what the researchers considered the 'knowledge gaps' to be is a helpful orientation. Some realist reviews are done with the goal of adding to the academic literature (e.g. a thesis); others may be undertaken for a specific purpose (e.g. to inform policy in a particular setting).

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 4: Objectives and focus of review - Realist Synthesis

Item 4: Objectives and focus of review

State the purpose of the review and the review question(s). Define and justify the scope of the review.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A realist review research question contains some or all of the elements of 'What works, how, why, for whom, to what extent and in what circumstances, in what respect and over what duration?' and applies realist logic to address the question (see Item 11).

Because a realist review may generate a potentially infinite number of things that might be explored and explained, and because resources and timescale are invariably finite, the review must be 'contained' by progressively focusing both its breadth (how wide an area?) and depth (how much detail?). This important process may involve discussion and negotiation with (for example) context experts, funders and/or users. It is typical and legitimate for the review question and/or the breadth and depth of the review to evolve as the review progresses. How and why it evolved is usually worth reporting either here or in item 16.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Methods section - Realist Synthesis

The following questions cover potential Items for inclusion in the Methods section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 5: Protocol - Realist Synthesis

Item 5: Protocol

The final protocol (i.e. the account of what was actually done) should be reproduced, at least in summary form, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the protocol given. It may also be appropriate to publish the original protocol (e.g. as set out in the grant proposal or developed in the early stages of the review).

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The study protocol for a realist review differs in significant respects from that in a traditional meta-analytic review. As noted above (in Item 4), the research question and scope (and, by implication, all subsequent steps) of a realist review can (and often should) evolve over the course of the review. However, this does not mean the review can meander uncontained. An accessible summary of what was done, in what order, and why is essential for interpreting the review. Comparing the original protocol with the final account of what was done may provide transparency on how the review's processes has evolved in its bid to build understanding of the topic area.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 6: Rational for using Realist Synthesis

Item 6: Rationale for using Realist Synthesis

Explain why realist synthesis was used.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Realist synthesis is a theory-driven method that is firmly rooted in a realist philosophy of science. It places particular emphasis on understanding causation (in this case, understanding how programs and policies generate outcomes through human decisions) and how causal mechanisms are shaped and constrained by social context. This makes it particularly suitable for reviews of certain topics and questions – for example, complex social programmes that involve human decisions and actions. It also makes realist review less suitable than other review methods for certain topics and questions – for example those which seek primarily to determine the average effect size of a simpler intervention administered in a limited range of conditions. The most common limitation of published 'realist' reviews is inadequate engagement with the philosophical principles of the realist approach and the implications these have, firstly, for understanding programs and how they work, and secondly, for cumulating evidence and explanation.

The published literature on realist review indicates that some review teams have deliberately adapted the method as first described by Pawson. The description and rationale for any adaptations made should be provided. Such information will allow criticism, debate and counter criticism amongst review teams and users on the suitability of such adaptations.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 7: Scoping the literature - Realist Synthesis

Item 7: Scoping the literature

Describe and justify the initial process of exploratory scoping of literature.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
This step is used to build an understanding of the programme or intervention(s) of interest and identify provisional programme theories. If identification of programme theories is not deemed to be appropriate at this stage, this should be justified. Findings from the scoping exercise should be reported in the Results section.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 8: Searching processes - Realist Synthesis

Item 8: Searching processes

State how the search was done and provide details on all the information sources accessed in the review. Where electronic search strategies were used, the information should include (for example) name of database, dates of coverage, limits applied, and date last searched. Contact with relevant content experts should be indicated.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:

Searching should be guided by the objectives and focus of the review, and revised iteratively in the light of emerging data. Data relevant to a realist review may lie in a broad range of sources that may cross traditional disciplinary, program and sector boundaries. Searching is thus likely to involve searching for different sorts of data, or studies from different domains, with which to test different aspects of any provisional theory.

A single maximally sensitive search is most unlikely to be sufficient. Search methods using forward and backward citation tracking are more likely to help in finding the documents necessary to develop and then test provisional theories. Realist reviews do not exclude sources solely on the basis of their study design, hence 'methodological filters' (for example, to identify randomised controlled trials) should used with caution, if at all.

Searching is likely to be iterative because as the review progresses new or refined elements of theory may be required to explain particular findings, or to examine specific aspects of particular processes. As new elements of theory are included, searches for evidence to support, refute or refine those elements may be required. If undertaken the process used for any such additional searches should be clearly documented. A search strategy that does not change as the review progresses may suggest insufficient reflection on emerging findings.

Sufficient detail should be given to enable the reader to judge whether searching was sufficiently extensive and directed to locate key sources needed for theory building and/or testing.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 9: Selection and appraisal of documents - Realist Synthesis

Item 9: Selection and appraisal of documents

Explain how judgements were made about documents to be included and excluded, and justify these.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Realist review is not a technical process. Rather, it requires a series of judgements about the relevance and robustness of particular data for the purposes of answering a specific question.

Within any one document, there may be several pieces of data that serve different purposes in the review, such as helping to build one theory, refining another theory and so on. Therefore the selection (for inclusion or exclusion) and appraisal of the 'worth' of any document cannot be based on an overall assessment of document 'quality'. An appraisal of the 'worth' of any section of data (within a document) should be made on two criteria:

- Relevance – whether it can contribute to theory building and/or testing; and
- Rigour – whether the method used to generate that particular piece of data is credible and trustworthy.

A wide range of documents can potentially contribute to a realist review. For example, outcome and impact studies, qualitative interviews, ethnography, questionnaire surveys, mixed-method case studies, and close reading of policies, business plans, websites, project initiation documents and 'grey literature' write-ups of programmes may all contribute in different ways to identifying and elucidating programme theories. Because of this range and realist review's focus on relevance and rigour, it can initially be difficult to 'whittle down' the number of documents that are potentially eligible for inclusion in a review. This process can only occur as the data sources are analysed in detail. Thus, in practice, the selection and appraisal stage may need to run in parallel with the analysis stage.

Description of the selection and appraisal process should be sufficiently detailed to enable a reader to estimate how likely it is that researchers inadvertently excluded data that may have significantly altered the findings of the review.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 10: Data extraction - Realist Synthesis

Item 10: Data extraction

Describe and explain which data were extracted from the included documents and justify this selection.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
There are two purposes for data extraction:
1) to assist data analysis and synthesis and
2) to add to the transparency of the review process.

The extracted data may consist of descriptions (e.g. of the detail of what was done in a programme), findings (e.g. cure rates, mortality) or explanations about how and why the programme may have worked in particular contexts. Of particular interest to the realist reviewer are data which support the use of realist logic to answer the review's question(s) – e.g. data on context, mechanisms, and outcome configurations, demi-regularities, middle-range and/or programme theories. Realist review is used for a wide range of research questions, so it is impossible to be prescriptive about which data should be extracted. However, the link between the research question and the type of data extracted should be clear.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 11: Analysis and synthesis processes - Realist Synthesis

Item 11: Analysis and synthesis processes

Describe the analysis and synthesis processes in detail. This section should include information on the constructs that are analysed, describe the analytic process, and document and justify any changes in this process as the study unfolded.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
In a realist review, the analysis and synthesis processes occur iteratively and may be sequential or in parallel. At the centre of any realist analysis is the application of a realist philosophical 'lens' to data. A realist analysis of data specifically seeks to analyse data using realist concepts. Specifically, realism adheres to a generative explanation for causation – i.e. an outcome (O) of interest was generated by relevant mechanism(s) (M) being triggered by context (C). Within or across the documents, recurrent patterns (or demi-regularities) of outcomes and their associated mechanisms and contexts (CMO configurations) are likely to occur.

During synthesis the goal is to make sense of the analysed data using theory, at at least one of two levels. Firstly, theory (or theories) may be sought, developed and/or refined to explain how it is that a programme (or part of a programme) achieves its outcomes (that is, the mechanism(s) operating within a program) and the contexts in which those mechanisms do and do not fire. This provides a realist program theory. Secondly, theory (or theories) may be sought, developed and/or refined to explain, at a somewhat more general level, the pattern of C, M and Os. A full realist analysis addresses both these levels and attempts to make sense of the relationship between these two levels. Syntheses which address only one level may also be considered realist syntheses assuming that they apply and demonstrate application of a realist philosophy of science. The level(s) of analysis chosen will depend on the review's focus. The theories used may have been developed and/or refined from the data and/or be refinement of existing substantive theory.

The key analytic process in realist review involves iterative testing and refinement of theoretically based explanations using empirical findings in data sources. Reviewers may draw on any appropriate analytic techniques to undertake this testing. Explanation and justification for the choice of techniques should be provided.

Ideally a description should be provided on how the all the individuals involved in the review have been involved in the analysis and synthesis processes, and how these evolved as the review took shape.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Results section - Realist Synthesis

The following questions cover potential Items for inclusion in the Results section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 12: Document flow diagram - Realist Synthesis

Item 12: Document flow diagram

Ideally within a flow diagram, provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin. A template (which may need further modification to suit the data) is given in Figure X.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A flow diagram provides an accessible summary of the sequence of steps and gives and indication of the volume of data included and excluded at each step.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 13: Document characteristics - Realist Synthesis

Item 13: Document characteristics

Information on the characteristics of the documents included in the review should be provided.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Characteristics of documents might include for example (where applicable) full citation, country of origin, study design and (where applicable) main findings. A clear summary of the characteristics of included sources adds to the transparency of the review and may help readers judge the coherence and plausibility of inferences. Reviewers may wish to report data source characteristics within one or more tables.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 14: Main findings - Realist Synthesis

Item 14: Main findings

Present the key findings with a specific focus on theory building and testing.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The defining feature of a realist review is the nature of the theory(ies) it offers. Such a theory explains why a social programme / intervention generates particular outcomes in particular contexts, in terms of one or more mechanisms – that is how the programme's infrastructure and resources trigger particular decisions or behaviours in human participants. Programme theories are usually 'middle-range' – that is, specific enough to generate propositions that can be tested about aspects of the programme but sufficiently abstract to be applicable to other programmes. Mechanisms are contingent: they are causal processes that have a tendency to occur in a particular set of conditions, but which do not always occur (because the circumstances have to be right for any particular mechanism to operate, and because many mechanisms can operate concurrently, sometimes cancelling each other out).

The validity of a 'realist' review which talks about programme theories or mechanisms but which expresses these as simple and linear relationships between variables should be questioned.

The findings of a realist review consist largely of inferences about the links between context, mechanism and outcome and the theory(ies) that account for these links. It is important that where inferences are made this is clearly articulated. It is also important to include where possible as much of an explanation to show how these inferences were arrived at.

Transparency of the review process can be demonstrated, for example, by including such things as a detailed worked example, verbatim quotes from primary sources, or an exploration of disconfirming data (i.e. findings which appeared to refute the programme theory but which, on closer analysis, could be explained by other contextual influences).

When presenting inferences about context-mechanism-outcome configurations, reviewers should be clear about what they have categorised as context, what as mechanism and what as outcome, and justify this taxonomy. In a realist review a mechanism involves the interaction between particular inputs (or resources) and human reasoning which produces a particular outcome (or not).

More than one piece of data might be needed to support an inference. It is sometimes appropriate to build the argument for an inference as an unfolding narrative in which successive data sources increase the strength of the inference. Provide enough details about each data item to identify its source and enable readers to make judgements about its relevance and rigour.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Discussion section - Realist Synthesis

The following questions cover potential Items for inclusion in the Discussion section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 15: Summary of findings - Realist Synthesis

Item 15: Summary of findings

Summarise the main findings with attention to the research question, focus of the review, and intended audience.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
In order to place the findings in the context of the wider literature and policy need, it is necessary to summarise briefly what has been found. This section should be succinct and balanced, explaining one or more key theories which emerged from the analysis and highlighting the strength of evidence for the main inferences. This should be done with careful attention to the needs of the main users of the review.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 16: Strengths, limitations and future research directions - Realist Sy...

Item 16: Strengths, limitations and future research directions

Discuss both the strengths of the review and its limitations. These should include (but need not be limited to):
[a] consideration of all the steps in the review process and
[b] comment on the adequacy and trustworthiness of the explanatory insights which emerged. The limitations identified may point to areas where further research is needed.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Realist reviews may be constrained by time and resources, by the skill mix and collective experience of the research team and/or by anticipated or unanticipated challenges in the data. These should be made explicit so that readers can interpret the findings in the light of them. A common challenge in realist reviews is that in order to focus the review, some material is omitted at each successive stage. Some aspects of the topic area therefore end up being reviewed in detail and rich explanatory insights produced for these. Other aspects are neglected (relatively or absolutely). It is thus inevitable that in generating illumination, the review will also cast shadows. These should be highlighted in the discussion so as to indicate areas where other reviews might focus.

Limitations imposed by any modifications made to the review process should also be reported and justified.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 17: Comparison with existing literature - Realist Synthesis

Item 17: Comparison with existing literature

Compare and contrast the review's findings with the existing literature on the same topic matter.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Comparing and contrasting the findings from a review with the existing literature may help readers to put these into context. For example, this Item might cover questions such as; how does this review compare to other reviews (e.g. were they theory-driven?); what does this review add, and which body of work in particular does it add to?; has this review reached the same or different conclusion to previous reviews?; and has it answered a question previously identified as important by leaders in the field?

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 18: Conclusion and recommendations - Realist Synthesis

Item 18: Conclusion and recommendations

List the main implications that are justified by the data and place these in the context of other relevant literature. If appropriate, offer recommendations.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A clear line of reasoning is needed to link implications with the findings presented in the results section. If the review is small and preliminary, or if the strength of evidence behind the inferences is weak or moderate, firm implications for practice and policy may be inappropriate.

If recommendations are given, these should be presented appropriately. The explanations in realist analysis are highly dependent on contextual influences. It follows that recommendations must be contingent (for example only under certain contexts will a particular mechanism be triggered to generate the desired outcome) rather than statements that X should or should not be done.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 19: Funding - Realist Synthesis

Item 19: Funding

Details should be provided for the funding source (if any) for the review, the role played by the funder (if any) and any conflicts of interests of the reviewers.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The source of funding for a review and/or personal conflicts of interests may influence the research question, methods, data analysis and conclusions. No review is a 'view from nowhere', and readers will be better able to interpret the review if they know why it was done and for which sponsor.

If a review is published, the process for reporting funding and conflicts of interest as set out by the publication concerned should be followed.

Please comment on item, including wording (optional):

Appendix 4 'Paper' version of round 3 online Delphi panel survey for realist reviews

RAMESES Delphi - Round 3

Introduction

Thank you for continuing to help us with the RAMESES project.

In Round 2 of our Delphi process, you rated a list of potential Items for inclusion in the 'RAMESES statement' for Relevance and Content.

We have analysed you ratings and free text comments and only six Items (three each for realist synthesis and meta-narrative review) needed to be included in Round 3 for your attention. We would be grateful if you would please rate these six Items for:

- Relevance (should we include an Item on this theme/topic at all?)
- Content (should we word this Item like this?)

There will be a free text box for you to make comments on any aspect of an Item. To help you understand why an Item has been included we have also provided a brief explanation.

As this survey has only six Items, it will only take a few minutes of your time.

You may at any time stop and return to where you left off by clicking on the unique web link you were sent inviting you to take part in this survey. You may also go back to previous items if you wish.

We would be most grateful if you would please try to complete the survey by Sunday 11th March 2012 at the latest.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 3

PART 1 - Realist Synthesis

The questions in PART 1 cover potential Items for inclusion in the RAMESES publication standards for Realist Synthesis only.

Our previous formulations of the following two items from the Methods section of the RAMESES publication standards did not achieve consensus in Round 2. We would appreciate further attention to our revised efforts.

For each Item we have provided you with our new suggested wording as well as the results of the ratings and original wording from Round 2.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 3

Item 5: Changes in the review process

Item 5: Changes in the review process

Any changes made to the review process that was initially planned should be briefly described and justified.

ORIGINAL TEXT IN ROUND 2
Item 5: Protocol
The final protocol (i.e. the account of what was actually done) should be reproduced, at least in summary form, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the protocol given. It may also be appropriate to publish the original protocol (e.g. as set out in the grant proposal or developed in the early stages of the review).

RATINGS FROM ROUND 2

Relevance
Response rate (%): 35/37 (95)
Mode: 7
Median: 6
Inter-quartile range: 5 to 7

Content:
Response rate (%): 34/37 (92)
Mode: 7
Median: 5.5
Inter-quartile range: 3 to 6.75

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A realist review can (and often should) evolve over the course of the review. For example changes to the research question or scope is likely to have an impact on many of the review's subsequent processes. However, this does not mean the review can meander uncontained. At the very least, an accessible summary of what was planned and how and why this differed from what was done should be provided as this may assist interpretation.

ORIGINAL TEXT IN ROUND 2
The study protocol for a realist review differs in significant respects from that in a traditional meta-analytic review. As noted above (in Item 4), the research question and scope (and, by implication, all subsequent steps) of a realist review can (and often should) evolve over the course of the review. However, this does not mean the review can meander uncontained. An accessible summary of what was done, in what order, and why is essential for interpreting the review. Comparing the original protocol with the final account of what was done may provide transparency on how the review's processes has evolved in its bid to build understanding of the topic area.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 3

Item 9: Selection and appraisal of documents - Realist Synthesis

Item 9: Selection and appraisal of documents

Explain how judgements were made about documents to be included and excluded, and justify these.

RATINGS FROM ROUND 2

Relevance
Response rate (%): 35/37 (95)
Mode: 7
Median: 7
Inter-quartile range: 6 to 7

Content:
Response rate (%): 35/37 (95)
Mode: 7
Median: 6
Inter-quartile range: 4.5 to 7

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

RAMESES Delphi - Round 3

Explanation:
Realist review is not a technical process. Rather, it requires a series of judgements about the relevance and robustness of particular data for the purposes of answering a specific question.

Within any one document, there may be several pieces of data that serve different purposes in the review, such as helping to build one theory, refining another theory and so on. Therefore the selection (for inclusion or exclusion) and appraisal of the 'worth' of any document cannot be based on an overall assessment of document 'quality'. An appraisal of the 'worth' of any section of data (within a document) should be made on two criteria:

- Relevance – whether it can contribute to theory building and/or testing; and
- Rigour – whether the method used to generate that particular piece of data is credible and trustworthy.

A wide range of documents can potentially contribute to a realist review. For example, outcome and impact studies, qualitative interviews, ethnography, questionnaire surveys, mixed-method case studies, and close reading of policies, business plans, websites, project initiation documents and 'grey literature' write-ups of programmes may all contribute in different ways to identifying and elucidating programme theories. Because of this range and realist review's focus on relevance and rigour, it can initially be difficult to 'whittle down' the number of documents that are potentially eligible for inclusion in a review. This process can only occur as the data sources are analysed in detail. Thus, in practice, the selection and appraisal stage may need to run in parallel with the analysis stage.

It is unlikely that authors will be able to describe each decision involved, but the broad processes used to determine relevance and assess rigour (for example, using quality standards appropriate to particular kinds of research to appraise documents or sections of documents, discussion and/or debate within a review team of a document's findings or consulting experts about technical aspects of methods or findings) should be described. Whilst the description of the processes followed will not allow the reader to draw firm conclusions about judgements made, it will give an indication of the coherence, plausibility and appropriateness of the processes used to inform those judgements.

ORIGINAL TEXT IN ROUND 2
[NB: Only the final paragraph of the Round 2 text has been amended. The original Round 2 text is below.]
Description of the selection and appraisal process should be sufficiently detailed to enable a reader to estimate how likely it is that researchers inadvertently excluded data that may have significantly altered the findings of the review.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 3

Results section - Realist Synthesis

The following question covers a potential Item for inclusion in the Results section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 3

Item 13: Document characteristics - Realist Synthesis

Item 13: Document characteristics

Provide information on the characteristics of the documents included in the review.

RATINGS FROM ROUND 2

Relevance
Response rate (%): 35/37 (95)
Mode: 7
Median: 6
Inter-quartile range: 5 to 7

Content:
Response rate (%): 35/37 (95)
Mode: 7
Median: 6
Inter-quartile range: 4.5 to 7

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A clear summary of the characteristics of included sources can add to the transparency of the review and may help readers judge the coherence and plausibility of inferences. Characteristics of documents might include, where applicable: full citation, country of origin, study design, summary of key main findings and how the document contributed to the review. While considering specific requirements of any particular publication, reviewers may wish to tabulate key characteristics.

ORIGINAL TEXT IN ROUND 2
Characteristics of documents might include for example (where applicable) full citation, country of origin, study design and (where applicable) main findings. A clear summary of the characteristics of included sources adds to the transparency of the review and may help readers judge the coherence and plausibility of inferences. Reviewers may wish to report data source characteristics within one or more tables.

Please comment on item, including wording (optional):

Appendix 5 'Paper' version of round 2 online Delphi panel survey for meta-narrative reviews

RAMESES Delphi - Round 2

PART 2 - Meta-Narrative Synthesis

The questions in PART 2 cover potential Items for inclusion in the RAMESES publication standards for Meta-Narrative Synthesis only.

The first four Items are topics for consideration in the Introductory section.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 1: Title - Meta-Narrative Synthesis

Item 1: Title

In the title, identify the document as a Meta-narrative Synthesis or Review.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Our background searching has shown that some meta-narrative reviews are not flagged as such in the title and may also be inconsistently indexed, and hence are more difficult to locate during searching. Most authors currently use the term 'meta-narrative review'. Consistent use of one term is likely to aid indexing and identification. We are interested to find out which term you prefer and why.

Please choose your preferred term:

◯ Meta-Narrative Synthesis

◯ Meta-Narrative Review

◯ No preference

Reason(s) for choice of term (optional):

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 2: Abstract - Meta-Narrative Synthesis

Item 2: Abstract

As far as possible (taking account of journal-specific formatting and content requirements), the abstract should contain brief details of the study context, review question or objectives; search strategy (including literatures / disciplinary areas searched); selection and appraisal of documents; analysis and synthesis methods; results; and conclusions/implications.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Apart from the title, an abstract is the only source of information accessible to searchers unless the full paper is obtained. The information in it must allow reviewers and/or users to decide if the review is relevant to their needs.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 3: Rationale for review - Meta-Narrative Synthesis

Item 3: Rationale for review

Explain why the review was done and what it is likely to add to existing understanding of the topic area.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
As with all research, a background section explaining what is already known and what the researchers considered the 'knowledge gaps' to be is a helpful orientation. Some meta-narrative reviews are done with the goal of adding to the academic literature (e.g. a thesis); others may be undertaken for a specific purpose (e.g. to inform policy in a particular setting).

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 4: Objectives and focus of review - Meta-Narrative Synthesis

Item 4: Objectives and focus of review

State the purpose of the review and the review question(s). Define and justify the scope of the review.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A meta-narrative review asks some or all of the following questions:
[1] Which research (or epistemic) traditions have considered this broad topic area;
[2] How has each tradition conceptualised the topic (for example including assumptions about the nature of reality, preferred study designs and ways of knowing)?;
[3] What theoretical approaches and methods did they use?;
[4] What are the main empirical findings?; and
[5] What insights can be drawn by combining and comparing findings from different traditions?'

Because a meta-narrative review may generate a potentially infinite number of things that might be explored and explained, and because resources and timescale are invariably finite, the review must be 'contained' by progressively focusing both its breadth (how wide an area?) and depth (how much detail?). This important process may involve discussion and negotiation with (for example) context experts, funders and/or users. It is typical and legitimate for the review question and/or the breadth and depth of the review to evolve as the review progresses. How and why it evolved is usually worth reporting.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Methods section - Meta-Narrative Synthesis

The following questions cover potential Items for inclusion in the Methods section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 5: Protocol - Meta-Narrative Synthesis

Item 5: Protocol

The final protocol (i.e. the account of what was actually done) should be reproduced, at least in summary form, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the protocol given. It may also be appropriate to publish the original protocol (e.g. as set out in the grant proposal or developed in the early stages of the review).

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The study protocol for a meta-narrative review differs in significant respects from that in a traditional systematic review with meta-analysis. As noted above (Item 4), the research question and scope (and, by implication, all subsequent steps) of a meta-narrative review can (and often should) evolve over the course of the review. However, this does not mean the review can meander uncontained. An accessible summary of what was done, in what order, and why is essential for interpreting the review. Comparing the original protocol with the final account of what was done may provide transparency on how the review's processes has evolved in its bid to build understanding of the topic area.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 6: Rational for using Meta-Narrative approach

Item 6: Rationale for using Meta-Narrative Synthesis

Explain why meta-narrative review was used.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Meta-narrative review (which is rooted in a constructivist philosophy of science) is inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms (see definition below). Meta-narrative review looks historically at how particular research (or epistemic*) traditions have unfolded over time and shaped the 'normal science' of a topic area.

Some definitions:
• A paradigm is a particular way of viewing the world, including assumptions about how the world works, what are the important questions in a particular topic area, and what study designs and methods are best for adding to the knowledge base.
• A research tradition is a series of linked studies, each building on what has gone before, usually situated within a coherent paradigm, though an interdisciplinary tradition may bridge more than one paradigm.
• Normal science is a paradigm along with the practices and empirical approaches which are taken for granted by scientists within a particular tradition.

Meta-narrative review is therefore best suited to studying topic areas that have been differently conceptualised and studied by different groups over time. The review seeks first to identify and understand all the different research traditions which have a bearing on the topic, and then to synthesise them by means of an over-arching narrative. The goal of meta-narrative review is sense-making of a complex (and perhaps contested) topic area.

*We would value your opinion on whether a meta-narrative should also look historically at how particular epistemic traditions have unfolded over time:

Should a meta-narrative also consider epistemic traditions?

◯ Yes
◯ No
◯ Don't know

Reason(s) for choice of term (optional):

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 7: Evidence of adherence to guiding principles of meta-narrative revie...

Item 7: Evidence of adherence to guiding principles of meta-narrative review

Present evidence in your methodology to show how the six guiding principles (pragmatism, pluralism, historicity, contestation, reflexivity and peer review) have been followed.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Meta-narrative review is based on six guiding principles:
• Principle of pragmatism: what to include is not self-evident. The reviewer must be guided by what will be most useful to the intended audience(s);
• Principle of pluralism: the topic should be illuminated from multiple angles and perspectives, using the established quality criteria appropriate to each. Hence, reviewers should avoid beginning with a single 'preferred' perspective or methodological hierarchy and proceed to judge work in other traditions using these external benchmarks. Research which lacks rigour must be rejected, but the grounds for rejection should be intrinsic to the relevant tradition, not imposed on it;
• Principle of historicity: each research tradition should be described as it unfolded over time, highlighting significant individual scientists, events and discoveries which shaped the tradition;
• Principle of contestation: 'conflicting data' from different research traditions should be examined to generate higher-order insights (e.g. about how different research teams framed the issue differently or made different assumptions about the nature of reality);
• Principle of reflexivity: throughout the review, reviewers must continually reflect, individually and as a team, on the emerging findings;
• Principle of peer review: emerging findings and the draft summary must be presented to an external audience and their feedback used to guide further reflection and analysis.

The published literature on meta-narrative review indicates that some review teams have deliberately adapted the method as first described by Greenhalgh et al. Whilst evolution and/or adaptation of the method is to be welcomed in principle, the description and rationale for any adaptations made should be provided to allow readers to judge their coherence with the underlying principles of meta-narrative review (and hence with its constructivist philosophical basis).

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 8: Scoping the literature - Meta-Narrative Synthesis

Item 8: Scoping the literature

Describe and justify the initial process of exploratory scoping of literature.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
One of the main challenges in meta-narrative review is to identify a sufficiently broad range of sources so as to be able to build as comprehensive a map as possible of research undertaken on the topic. This scoping step is used to identify in broad terms the different research traditions, sited in different literatures, which have addressed the topic of interest. Findings from this scoping phase should normally be reported in the Results section.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 9: Searching processes - Meta-Narrative Synthesis

Item 9: Searching processes

State how the search was done and provide details on all the information sources accessed in the review. In particular, state how seminal sources were identified. Where electronic search strategies were used, the information should include (for example) name of database, dates of coverage, limits applied, and date last searched. Contact with relevant content experts should be indicated.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Searching should be guided by the objectives and focus of the review, and revised iteratively in the light of emerging data. By definition, a meta-narrative review seeks to identify and combine different research traditions, hence different search strategies will need to be developed as appropriate to the different literatures. This stage is likely to involve searching for different kinds of data in different ways.

A single maximally sensitive search is most unlikely to be sufficient. Search methods using forward and backward citation tracking are more likely to help in finding key documents. In particular, potential seminal sources (conceptual, theoretical or empirical studies which have defined the tradition and inspired later work) may be identified from judicious searching of the reference lists of later studies. Once identified, seminal sources should be citation-tracked to identify further sources which drew on these.

Meta-narrative reviews do not approach the literature with a pre-defined 'preferred' study design. Rather, any preferred study design(s) should be identified from quality standards developed within a particular research tradition. 'Methodological filters' (for example, to identify randomised controlled trials) should used only when these have been designated as a quality feature by the scientists within that tradition.

Searching is necessarily iterative, since the reviewer must move between the seminal source(s) and papers which subsequently cited that source, so as to build a historical picture of how research unfolded in each tradition. The process used for any such additional searches should be clearly documented. A search strategy that does not change as the review progresses may suggest insufficient attention to emerging findings.

Sufficient detail should be given to enable the reader to judge whether searching was sufficiently extensive and directed to locate key sources needed for elucidating all the key research traditions.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 10: Selection and appraisal of documents - Meta-Narrative Synthesis

Item 10: Selection and appraisal of documents

Explain how judgements were made about documents to be included and excluded, and justify these.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Meta-narrative review is not a technical process. Rather, it is a process of sense-making of the literature, selecting and combining data from primary sources to produce a historical account of how a research tradition unfolded and why, and then (in a second phase) comparing and contrasting findings from these different traditions to build a rich picture of the topic area from multiple angles. This process requires a series of judgements about the unfolding of research in particular traditions, and about the relevance and robustness of particular data within that tradition.

Meta-narrative review takes its quality criteria from the traditions included in the review, and in particular from seminal papers which have been accepted by others within that tradition as authoritative. A meta-narrative review might, for example, include a meta-narrative from clinical epidemiology in which randomised controlled trials and meta-analyses of these are greatly valued; it might also include a meta-narrative from critical sociology in which theory-driven qualitative studies are greatly valued. Studies in these separate traditions should be appraised using the quality criteria which a competent peer-reviewer in that tradition would use.

Description of the selection and appraisal process should be sufficiently detailed to enable a reader to estimate how likely it is that researchers inadvertently excluded data that may have significantly altered the findings of the review.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 11: Data extraction - Meta-Narrative Synthesis

Item 11: Data extraction

Describe and explain which data were extracted from the included documents and justify this selection.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The type of data that may be collected in meta-narrative review can be very diverse. There are two purposes for data extraction: 1) to assist data analysis and synthesis and 2) to provide transparency of the review process.

In a meta-narrative review the data elements extracted would go to constructing a story of how research on a topic unfolded over time in a particular tradition. This may include:
• upstream (antecedent) traditions from which these emerged; background philosophical assumptions;
• research questions and how they were framed; conceptual and theoretical issues;
• preferred methodologies, study designs and quality criteria;
• key actors (e.g. leading scientists or commentators) and events (e.g. conferences) in the unfolding of the tradition;
• landmark empirical or theoretical studies;
• significant findings and how these shaped subsequent work; and
• key debates and areas of dispute within the tradition, including links with or breaches from other traditions.

Meta-narrative review is used for a wide range of research questions, so it is impossible to be prescriptive about which data should be extracted. However, the link between the research question and the type of data extracted should be clear.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 12: Analysis and synthesis processes - Meta-Narrative Synthesis

Item 12: Data collection process

Describe the analysis and synthesis processes in detail. This section should include information on the process by which the account of each meta-narrative (i.e. the story of each unfolding research tradition) was built up and how the separate meta-narratives were compared and contrasted. Document and justify any changes in this process as the study unfolded.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
If exploration of a range of research traditions on the topic is not deemed to be appropriate, the work is probably not a meta-narrative review.

A meta-narrative review should comprise two specific stages, though these will usually overlap as they will necessarily influence one another iteratively.

In the analysis stage, reviewers should seek to identify and map out specific meta-narratives (that is, unfolding stories of research traditions over time), focusing in particular on the concepts, theories, methods and instruments which have characterised the tradition, major findings in that tradition and foci of dissent and disagreement.

The process of building this unfolding storyline is essentially interpretive and hence follows the principles of interpretivist analysis, including immersion in the data by repeated reading and/or analysis of quantitative data; reflexivity and discussion among researchers; consideration of how each new data item fits with an emerging picture of the whole; and checking where appropriate that the story is considered valid by members within the designated research tradition. Both quantitative and qualitative traditions and data will need to be incorporated in the storyline. Explanation and justification for the need to use any analytic methods should be provided.

The synthesis stage involves comparing and contrasting the meta-narratives so as to identify differences between how the different groups have conceptualised the topic (including differences in philosophical position), how they have theorised it, and the methodological approaches and study designs used. Differences in findings between meta-narratives are higher-order data and should be analysed interpretively to produce further insights (e.g. about differences in underlying assumptions or methodological approaches between different research traditions).

Synthesis across traditions may occur at a high level of abstraction (i.e. at the level of concepts and theories) and involve one or more of the following:
- paradigm bridging (seeking commonalities in underlying conceptual and theoretical assumptions),
- paradigm bracketing (highlighting differences in these assumptions),
- interplay (exploring tensions);
- meta-theorising (exploring patterns that span conflicting understandings)

Synthesis may also occur at a more concrete level and summarise empirical findings, using techniques including statistical aggregation, qualitative aggregation, and narrative summary.

A description should be provided of how the all the individuals involved in the review have been involved in the analysis and synthesis processes, and of the nature of any input from external advisors / peer reviewers from the included traditions.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Results section - Meta-Narrative Synthesis

The following questions cover potential Items for inclusion in the Results section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 13: Document flow diagram - Meta-Narrative Synthesis

Item 13: Document flow diagram

Ideally within a flow diagram, provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin. A template (which may need further modification to suit the data) is given in Figure X.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A flow diagram provides an accessible summary of the sequence of steps and gives and indication of the volume of data included and excluded at each step.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 14: Document characteristics - Meta-Narrative Synthesis

Item 14: Document characteristics

Information on the characteristics of the documents included in the review should be provided.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Characteristics of documents might include for example (where applicable) full citation, country of origin, study design and main findings. A clear summary of the characteristics of included sources adds to the transparency of the review and may help readers judge the coherence and plausibility of inferences. Reviewers may wish to report data source characteristics within one or more tables.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 15: Main findings - Meta-Narrative Synthesis

Item 15: Main findings

Present the main findings with a specific focus on the key meta-narratives that have a bearing on the topic area, and the commonalities and differences between them.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The defining feature of a meta-narrative review is illumination of a complex topic area from multiple angles. Each meta-narrative should first be presented as a coherent individual account which conveys the underpinning 'normal science' of the relevant research tradition (concepts, theories, preferred methods) and the key empirical findings in that tradition. Findings and inferences from the synthesis across the different meta-narratives may be presented as an over-arching narrative which retains the integrity of the separate research traditions but draws out what might be learnt from the commonalities and differences between them.

The outputs of paradigm bridging, paradigm bracketing, interplay and meta-theorising should be presented as appropriate to summarise the conceptual and theoretical basis of the meta-narratives. The outputs of statistical aggregation, qualitative aggregation, and narrative summary of disaggregated data should be presented as appropriate to summarise the empirical findings. In each case, data from the primary documents should be presented and sourced to illustrate how inferences have been made and justify these. The more detail is given, the more readers will be able to judge the validity of the inferences.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Discussion section - Meta-Narrative Synthesis

The following questions cover potential Items for inclusion in the Discussion section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 2

Item 16: Summary of findings - Meta-Narrative Synthesis

Item 16: Summary of findings

Summarise the main findings with attention to the research question, focus of the review, and intended audience.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
In order to place the findings in the context of the wider literature and policy need, it is necessary to summarise briefly what has been found. This section should be succinct and balanced, highlighting the key meta-narratives which emerged from the analysis and the key points of commonality and contestation between them. This should be done with careful attention to the needs of the main users of the review.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 17: Strengths, limitations and future research directions - Meta-Narr...

Item 17: Summary of evidence

Discuss both the strengths of the review and its limitations. These should include (but need not be limited to) [a] consideration of all the steps in the review process and [b] comment on the adequacy and trustworthiness of the explanatory insights which emerged. The limitations identified may point to areas where further research is needed.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Meta-narrative reviews may be constrained by time and resources, by the skill mix and collective experience of the research team and/or by anticipated or unanticipated challenges in the data. These should be made explicit so that readers can interpret the findings in the light of them. A common challenge in meta-narrative reviews is that in order to focus the review, some material is omitted at each successive stage. Some aspects of the topic area therefore end up being reviewed in detail and rich explanatory insights produced for these. Other aspects are neglected (relatively or absolutely). It is thus inevitable that in generating illumination, the review will also cast shadows. These should be highlighted in the discussion so as to indicate areas where other reviews might focus.

Limitations imposed by any modifications made to the review process should also be reported and justified.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 18: Comparison with existing literature - Meta-Narrative Synthesis

Item 18: Comparison with existing literature

Compare and contrast the review's findings with the existing literature on the topic area.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
Comparing and contrasting the findings from a review with the existing literature may help readers to put these into context. This section might cover questions such as; how does this review compare to other reviews; what does this review add, and which body of work in particular does it add to?; has this review reached the same or different conclusion to previous reviews?; and has it answered a question previously identified as important by leaders in the field?

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 19: Conclusion and recommendations - Meta-Narrative Synthesis

Item 19: Conclusion and recommendations

List the main implications that are justified by the findings and place these in the context of other relevant literature. If appropriate, offer recommendations.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A clear line of reasoning is needed to link implications with the findings presented in the results section. If the review is small and preliminary, or if the strength of evidence behind the inferences is weak or moderate, firm statements about the implications for practice and policy may be inappropriate.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 2

Item 20: Funding - Meta-Narrative Synthesis

Item 20: Funding

Details should be provided for the funding source (if any) for the review, the role played by the funder (if any) and any conflicts of interests of the reviewers.

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
The source of funding for a review and/or personal conflicts of interests may influence the research question, methods, data analysis and conclusions. No review is a 'view from nowhere', and readers will be better able to interpret the review if they know why it was done and for which sponsor.

If a review is published, the process for reporting funding and conflicts of interest as set out by the publication concerned should be followed.

Please comment on item, including wording (optional):

Appendix 6 'Paper' version of round 3 online Delphi panel survey for meta-narrative reviews

RAMESES Delphi - Round 3

PART 2 - Meta-Narrative Synthesis

The questions in PART 2 cover potential Items for inclusion in the RAMESES publication standards for Meta-narrative Synthesis only.

Our previous formulations of the following two items from the Methods section of the RAMESES publication standards did not achieve consensus in Round 2. We would appreciate further attention to our revised efforts.

For each Item we have provided you with our new suggested wording as well as the results of the ratings and original wording from Round 2.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 3

Item 5: Changes in the review process

Item 5: Changes in the review process

Any changes made to the review process that was initially planned should be briefly described and justified.

ORIGINAL TEXT IN ROUND 2
The final protocol (i.e. the account of what was actually done) should be reproduced, at least in summary form, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the protocol given. It may also be appropriate to publish the original protocol (e.g. as set out in the grant proposal or developed in the early stages of the review).

RATINGS FROM ROUND 2

Relevance
Response rate (%): 31/33 (94)
Mode: 7
Median: 7
Inter-quartile range: 6 to 7

Content:
Response rate (%): 31/33 (94)
Mode: 7
Median: 6
Inter-quartile range: 6 to 7

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A meta-narrative review can (and often should) evolve over the course of the review. For example changes to the research question or scope is likely to have an impact on many of the review's subsequent processes. However, this does not mean the review can meander uncontained. At the very least, an accessible summary of what was planned and how and why this differed from what was done should be provided as this may assist interpretation.

ORIGINAL TEXT IN ROUND 2
The study protocol for a meta-narrative review differs in significant respects from that in a traditional systematic review with meta-analysis. As noted above (Item 4), the research question and scope (and, by implication, all subsequent steps) of a meta-narrative review can (and often should) evolve over the course of the review. However, this does not mean the review can meander uncontained. An accessible summary of what was done, in what order, and why is essential for interpreting the review. Comparing the original protocol with the final account of what was done may provide transparency on how the review's processes has evolved in its bid to build understanding of the topic area.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 3

Item 6: Rational for using Meta-Narrative approach

Item 6: Rationale for using Meta-Narrative Synthesis

Explain why meta-narrative review was considered the most appropriate method to use.

ORIGINAL TEXT IN ROUND 2
Explain why meta-narrative review was used.

RATINGS FROM ROUND 2

Relevance
Response rate (%): 27/33 (82)
Mode: 7
Median: 7
Inter-quartile range: 6 to 7

Content:
Response rate (%): 27/33 (82)
Mode: 7
Median: 6
Inter-quartile range: 5 to 7

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

RAMESES Delphi - Round 3

Explanation:
Meta-narrative review (which is rooted in a constructivist philosophy of science) is inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms (see definition below). Meta-narrative reviews often look historically at how particular research or epistemic traditions have unfolded over time and shaped the 'normal science' of a topic area.

Some definitions:
• A paradigm is a particular way of viewing the world, including assumptions about how the world works, what are the important questions in a particular topic area, and what study designs and methods are best for adding to the knowledge base.
• A research tradition is a series of linked studies, each building on what has gone before, usually situated within a coherent paradigm, though an interdisciplinary tradition may bridge more than one paradigm.
• An epistemic tradition is the unfolding of the underpinning set of philosophical assumptions which drive the development of theory and method; scholarship may progress via debate around these assumptions even in the absence of new empirical studies.
• Normal science is a paradigm along with the practices and empirical approaches which are taken for granted by scientists within a particular tradition.

Meta-narrative review is therefore best suited to studying topic areas that have been differently conceptualised and studied by different groups. The review seeks first to identify and understand as many as possible of the potentially important different research traditions which have a bearing on the topic, and then to synthesise them by means of an over-arching narrative. The goal of meta-narrative review is sense-making of a complex (and perhaps contested) topic area.

ORIGINAL TEXT IN ROUND 2
[NB: Main change has been the addition of a definition of epistemic tradition.]
Meta-narrative review (which is rooted in a constructivist philosophy of science) is inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms (see definition below). Meta-narrative review looks historically at how particular research or epistemic traditions have unfolded over time and shaped the 'normal science' of a topic area.

Some definitions:
• A paradigm is a particular way of viewing the world, including assumptions about how the world works, what are the important questions in a particular topic area, and what study designs and methods are best for adding to the knowledge base.
• A research tradition is a series of linked studies, each building on what has gone before, usually situated within a coherent paradigm, though an interdisciplinary tradition may bridge more than one paradigm.
• Normal science is a paradigm along with the practices and empirical approaches which are taken for granted by scientists within a particular tradition.

Meta-narrative review is therefore best suited to studying topic areas that have been differently conceptualised and studied by different groups over time. The review seeks first to identify and understand all the different research traditions which have a bearing on the topic, and then to synthesise them by means of an over-arching narrative. The goal of meta-narrative review is sense-making of a complex (and perhaps contested) topic area.

Please comment on item, including wording (optional):

RAMESES Delphi - Round 3

Results section - Meta-Narrative Synthesis

The following question covers a potential Item for inclusion in the Results section of the RAMESES publication standards.

Please click on the NEXT button below to proceed.

RAMESES Delphi - Round 3

Item 13: Document flow diagram - Meta-Narrative Synthesis

Item 13: Document flow diagram

Perhaps within a flow diagram, provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin. Example templates (which are likely to need modification to suit the data) are provided in SOURCES X.

ORIGINAL TEXT IN ROUND 2
Ideally within a flow diagram, provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin. A template (which may need further modification to suit the data) is given in Figure X.

RATINGS FROM ROUND 2

Relevance
Response rate (%): 21/33 (94)
Mode: 7
Median: 7
Inter-quartile range: 5 to 7

Content:
Response rate (%): 31/33 (82)
Mode: 7
Median: 6
Inter-quartile range: 4.5 to 7

Please rate this Item for:

	1 = Strongly Disagree	2	3	4	5	6	7 = Strongly Agree
Relevance - (Item inclusion)	◯	◯	◯	◯	◯	◯	◯
Content - (Item wording)	◯	◯	◯	◯	◯	◯	◯

Explanation:
A flow diagram provides an accessible summary of the sequence of steps and gives and indication of the volume of data included and excluded at each step.

Please comment on item, including wording (optional):

Appendix 7 Notes on teleconference with a review team the project team provided methodological support to

Teleconference with Durham Health Services and Delivery Research review team

Date
30 October 2012.

Time
10.00 a.m. to 11.10 a.m.

Participants
Geoff.

Jan.

Madeline.

Neill.

Purpose of meeting was to get feedback from review team in Durham about:

1. process of learning about realist reviews
2. concepts that were easy/harder to grasp in realist review.

General comments
Felt that the review method had been helpful as it enables reviewers to learn more about a topic than might get from a Cochrane review.

BUT required review team to be:

- engaged
- prepared to unlearn and relearn new things (may act to inhibit uptake of method?)
- be comfortable at the beginning to not know where you are heading
- read and reread the literature, engage with it deeper than just (for example) skim reading – 'takes time to make connections . . .'
- greater clarity comes from immersing yourself into the literature and this then helps with knowing where to head/go/change – '. . . saw things you never saw in other research methods . . .'
- was much harder work than any other review they had done. ALSO to make progress had to have LOTS of meetings. Review was more time consuming and labour intensive than they had anticipated.

Specific challenges

Consensus that worked examples were the most helpful way to learn.

There was praise for commitment, willingness and clarity of training from GWo from all of Durham team.

Suggestions for areas that need specific attention:

Clarifying terms – C, M, O programme theory and middle-ranged theory, relationship of an Intervention to CMO.

Focusing review – team felt that they had a huge topic to cover in a short period of time and so may not have done the subject matter 'justice'.

'Blueprint'/template – some members of the team felt that having a template of what to do might help. But there was also an appreciation that realist review was a review method that was iterative.

To help some learn, a 'quick start' style of guide covering the main concepts might be helpful.

Searching and inclusion – when does the systematic searching stop and realist searching start? Issue was more about what studies/documents to include. The review team understood the concept of relevance, but found that they could only resolve this with lots of discussion.

Realist logic – team members were worried that they might not have got realist logic. Getting feedback from trainer helped.

Analysis and synthesis

- Having a worked example of this that traces the 'journey' from a piece of data → inference → theory would help.
- Explaining the need that C, M and O may change over time and depending on which outcome is important – again worked example would help.
- Explaining the need to change level of abstraction of analysis – go deep and then back to more abstract – again worked example would help.

Reporting

Huge tension here between wanting to report all that they found and also to provide a document that they think might be relevant to policy and decision-makers, especially 'coal face' managers. Helped by context and content expertise and thinking like a manager.

Review team in the peer review of their report was asked by one peer reviewer to provide minute details of each CMO. Team agreed on the need for transparency, but felt that worked examples of how to 'walk this balance' of how much detail would help.

Comment about published reviews from team – in wanting to learn about realist review they turned to the published literature and found that it was full of examples that confused as opposed to helped. GWo explained regarding issue of 'fake handbags'. (Reviews which claimed to be realist reviews, but were in fact not.)

Team suggested that it is useful to point readers towards good examples, perhaps by focusing on the positives – e.g. this is a really good example of . . .

GWo 30 October 2012.

Appendix 8 Notes from the realist review training workshop held at Queen Mary University of London in March 2011

Advancing Realist Research Conference

Date

25 March 2011.

Venue

G O Jones Room, Queen Mary University of London, London, UK.

Participants

Participant	Affiliation
Rob Anderson	Peninsula Medical School, UK
David Baker	Dartmouth College, USA
Andrew Booth	University of Sheffield, UK
Madeline Carter	University of Durham, UK
Steve Dewar	Marie Curie Cancer Care, UK
Marjolein Dieleman	Royal Tropical Institute, the Netherlands
Carole Doherty	University of Surrey, UK
Tim Dornan	Maastricht University, the Netherlands
Ruth Garside	Peninsula Medical School, UK
Barend Gerretsen	Royal Tropical Institute, the Netherlands
Trish Greenhalgh	Queen Mary University of London, UK
Sacha Harris	Imperial College, UK
Andrea Herepath	Cardiff University, UK
Roger Kneebone	Imperial College, UK
Patricia Lanter	Dartmouth College, UK
Bruno Marchal	Institute of Tropical Medicine, Belgium
Ana Manzano-Santaella	University of Leeds, UK
Douglas Noble	Queen Mary University of London, UK
Ray Pawson	University of Leeds, UK
Mark Pearson	Peninsula Medical School, UK
Birte Snilstveit	3ie, UK
Charitini Stavropoulou	University of Surrey, UK
Katherine Stevenson	Jönköping University, Sweden
Neill Thompson	University of Durham, UK
Hugh Waddington	3ie, UK
Rebecca Walwyn	University of Leeds, UK

Participant	Affiliation
Gill Westhorp	Community Matters, Australia
Geoff Wong	University College London, UK

Feedback from sessions

Methods 2: introduction to RAMESES (Realist and Meta-narrative Evidence Synthesis: Evolving Standards)

Participants were presented information on the RAMESES project and asked what they would like from it.

Guidance/standards:

- Protocols for RS needed – consensus on this
- Explain how RS fits in with other review methods.
- Guidance needs to establish what counts as INTERNAL and EXTERNAL validity.
- Standards set should be broad enough to be suitable for 'all' purposes of RS – possibly principles based and not too 'rigid'.
- Reviewers using RS should understand:
    - realist ontology.
    - realist theory of causation.
- Guidance/standards must be useful to FUNDERS/REVIEWERS/RESEARCHERS.

Methodological:

- In general the HOW TO do X is a big problem in RS – tools needed.
- Glossary of terms/concepts:
    - Mechanism.
    - Programme theory.
    - Middle-range theory.
    - Context.
    - Programme/intervention.
    - Policy.
- Relationship between RS concepts.
- How to ensure transparency in a RS.
- How to write a RS protocol.
- When should RS be used?
- What can it be used for (e.g. just to understand policy or in other circumstances)?
- How to focus a RS so that it is 'do-able'.
- How to select/develop programme theory.
- How do you know what studies to include?
- How to pull out context from included studies.
- How do you analyse CMOs.

SYNTHESIS: group discussion

In small groups, participants were asked to try to map out the relationship between:

- programme theory
- mechanism
- context.

A summary of the main points

- These concepts were hard to define and distinguish and it was not clear to the participants if definitions would be relative or absolute. How they related to each other was also not clear to the participants. In addition these concepts (e.g. programme theory/logic models) were not unique to RS and this added to confusion over definitions.
- Some mentioned that they might be better off thought if as 'sensitising principles' and that precise definitions may not be either necessary or achievable. It was raised that some may need a precise definition in order to be able to use the concept – might this apply more to novices?
- The different way of thinking (about the world) needed to undertake a RS might mean some will struggle.
- Specific points discussed in the session:
    - Theories that are important are the ones that have bearing on the question of causality.
    - Context:
        - Context pre-exists the intervention. It can have two 'states' – at the beginning it is everything that has a bearing on X – at the end it is everything that actually did have a bearing on X. May also be thought of as 'context to describe' and 'context to explain'.
        - Contexts are defined in relation to a particular mechanism and conceptualising it this was helps in working out middle-range theory.
    - Mechanisms:
        - Mechanism is what is going through a person's head.
        - Mechanisms operate at different levels (e.g. individual psychology, group dynamics etc.) Mechanisms should be anchored to the outcome (same unit of analysis).
        - Mechanisms need to be distinguished from interventional modality/strategy.
- IN terms of taking things forward, ideas include:
    - Seminar series (e.g. phone, Adobe Connect mediated).
    - Reading list/materials.
    - From RAMESES try to put all RS reviews online as examples 'next to' standards – people can decide for themselves.
    - Bearing in mind that different people at different stages need different things!

Appendix 9 Realist synthesis: Realist And Meta-narrative Evidence Syntheses – Evolving Standards (RAMESES) training materials

REALIST SYNTHESIS

RAMESES TRAINING MATERIALS

July 2013

Geoff Wong, Gill Westhorp, Ray Pawson and Trish Greenhalgh

Authors' affiliations

Geoff Wong[1], Gill Westhrop[2], Ray Pawson[3] and Trish Greenhalgh[1]

[1] Centre for Primary Care and Public Health, Queen Mary University of London, 58 Turner Street, London E1 2AB, UK

[2] Community Matters, P.O. Box 443, Mount Torrens, SA 5244, Australia

[3] Department of Social Research Methodology, University of Leeds, Leeds LS2 9JT, UK

Corresponding author

Dr Geoff Wong
Senior Lecturer in Primary Care
Blizard Institute
Barts and The London School of Medicine and Dentistry
Queen Mary, University of London
58 Turner Street
London E1 2AB
(Tel): +44 (0)20 7882 2483
(Fax): +44 (0)20 7882 2200
(Email): grckwong@gmail.com

Competing Interests

None declared.

Acknowledgements

This project was funded by the National Institute for Health Research Health Services and Delivery Research Program (NIHR HS&DR) - project number 10/1008/07. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR program, NIHR, NHS or the Department of Health.

Section 1. A guide to the training materials

This training package has been developed to provide practical guidance to reviewers who want to undertake a realist synthesis (or realist review – the terms are synonymous). There has been, over recent years, a growing demand for training but so far, no 'how to' manuals exist. Development of the training materials was funded as part of the RAMESES project (http://www.ramesesproject.org).

The package comprises eight sections. Section 1 is this introduction to the training materials.

Section 2 provides a brief overview of realism, a glossary of realist terminology, and some information about further references and resources.

Sections 3 to 7 focus on five areas that have been a frequent source of difficulties and misconceptions for realist reviewers. The topics are:

- Section 3: Focussing reviews
- Section 4: Program theory
- Section 5: Developing a search strategy
- Section 6: Selection and appraisal of documents
- Section 7: Applying realist principles in analysis
- Section 8: Further reading and resources

The quality standards for realist synthesis developed through the RAMESES project can be found online at: http://www.ramesesproject.org/index.php?pr=Project_outputs#method. Note that the quality standards are not the same as the publication standards: they are an additional resource. These are designed to be used as a self-assessment tool by reviewers (and in particular those new to realist review) – a way to check whether the work done is consistent with the publication standards, and to work towards excellence. They may also assist peer reviewers of protocols and realist syntheses to provide structured and consistent feedback to review teams.

The five focus areas (Sections 3 to 7 of this training package) are not the only areas that realist reviewers find challenging. The RAMESES research team selected these as priorities based on our past experiences in practice and training, an on-line discussion list operated as part of the RAMESES project, the literature, and our work in preparing the publication standards for realist syntheses (http://www.biomedcentral.com/1741-7015/11/21).

Realist reviewers all have different needs. Some are about to embark on a review but others are already in the process of doing one; some have years of experience in realist research and others are novices. Catering to such diverse needs is a big task. Also, how a realist synthesis is undertaken depends on issues such as the research question, the resource available, the nature and amount of relevant literature, funder's expectations, and end users' needs. Consequently, it is impossible to be prescriptive about 'what must be done'. The training materials serve more as guidance than as 'must-do's'. In true realist fashion, we expect that these resources will operate differently for different groups: as a structured introduction for newcomers and an aide-memoire for old hands. We also anticipate that they will evolve as others question, challenge and seek to improve them.

Our approach is to set out the main principles for each of these challenging areas and provide a series of resources for each one. Each section provides:

- Learning objectives for the topic

- An explanation about why the topic matters – why it is important to get right
- 'Quality standards' for the topic.
- Examples drawn from published reviews.
  In each case, there is at least one example from the published literature of how the topic area has been tackled successfully, and at least one worked example (also drawn from the published literature) of how the topic in a review might be improved. Our case examples are necessarily brief and so learners may need to read in full the document each example is drawn from to more readily appreciate the comments we have made.
- A learning activity that provides an opportunity to practice a key skill or work with a key idea.
- A set of focused questions to help reviewers to reflect on their own review project and how to achieve the standards in their particular project

Further reading for all the sections may be found at the end of these training and learning materials in section 8.

The examples we have selected are not featured to 'name and shame', or 'name and fame', particular authors. Rather, our purpose is to clarify and explain how the challenges can be tackled by giving 'real life' examples and by suggesting potential solutions. Our goal is to improve the overall standard of realist syntheses.

If you would like to help us further improve these training materials, please contact us either by:
Email Geoff wong – grckwong@gmail.com
Or
Via the RAMESES JISCM@il email list
www.jiscmail.ac.uk/RAMESES

Section 2. Background materials

2.1 What is realism?

> *"Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the philosophy of science and social science."* (1)

'Realism' refers to a philosophy of science. It sits, broadly speaking, between positivism ('there is a real world which we can see and understand directly through observation') and constructivism ('given that all we can know has been interpreted through human senses and the human brain, we cannot know for sure what the nature of reality is'). Realism agrees that there is a real world and that our knowledge of it is processed through human senses, brains, language and culture. However, realism also argues that we can improve our understandings of reality because the 'real world' constrains the interpretations we can reasonably make of it. While our knowledge will always be partial and imperfect, it can accrue over time. Below, we introduce key ideas in realist philosophy, how they apply to social programs and what they imply for the role of researchers and reviewers.

Mechanism

Realism can help us understand the social world. Used in this way, it acknowledges the existence of an external social reality and the influence of that reality on human behaviour. To understand the relationship between context and outcome, realism uses the concept of 'mechanism'.

There are many definitions and conceptualisations of mechanisms (even within realism) (see Section 2.3, Glossary). In realist philosophy, mechanisms are causal forces or powers. They cause things to happen, something realist have termed – generative causation (see Figure 1). Mechanisms in social science are comparable but not identical to mechanisms in natural science (e.g. the mechanism of gravity accounts for why a dropped object falls to the ground). Social mechanisms may usefully be defined as '... underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest.' Here 'entities' may refer to things such as norms or belief systems, 'processes' are sequences where later events depend on earlier ones, and social structures may refer to things such as gender, class, or cultural patterns of relationships. Like the mechanisms in natural sciences, they possess a number of features: they are not 'visible', but must be inferred from the observable data; they are context sensitive, and they generate outcomes (2).

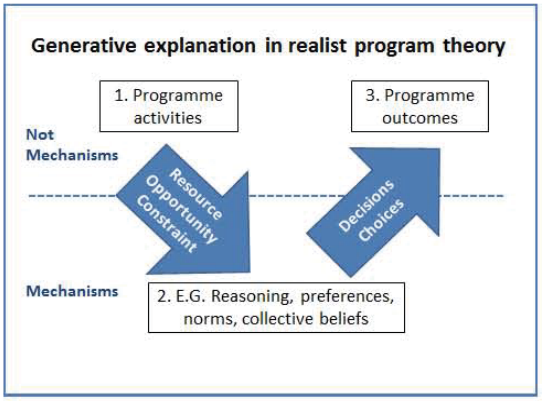


Figure 1: A pictorial representation of mechanisms

Social programs or interventions work by changing the decision-making of subjects. (We use the term 'subjects' here as shorthand for all those who may be directly affected by an intervention and whose decision-making does or could affect outcomes. In Pawson and Tilley's Close Circuit TV-in-car-parks example, 'subjects' included potential offenders, car owners, car park security staff, police, and passers-by. In human services programs, it usually refers to participants. However, mechanisms at earlier stages of implementation might involve funding providers, agency managers, service providers and so on.) The program or intervention changes the resources or opportunities available to subjects and, in that sense, changes the context for those subjects. The new context then triggers new mechanisms.

Using this logic, potential program mechanisms can be identified by asking what it is about a program that generates change. An intervention itself does not directly cause outcomes; it is the participants' reaction to the opportunities provided by the program that triggers the change. A realist approach therefore looks for interactions among the opportunities or resources provided by the intervention and the reasoning or responses of the participants.

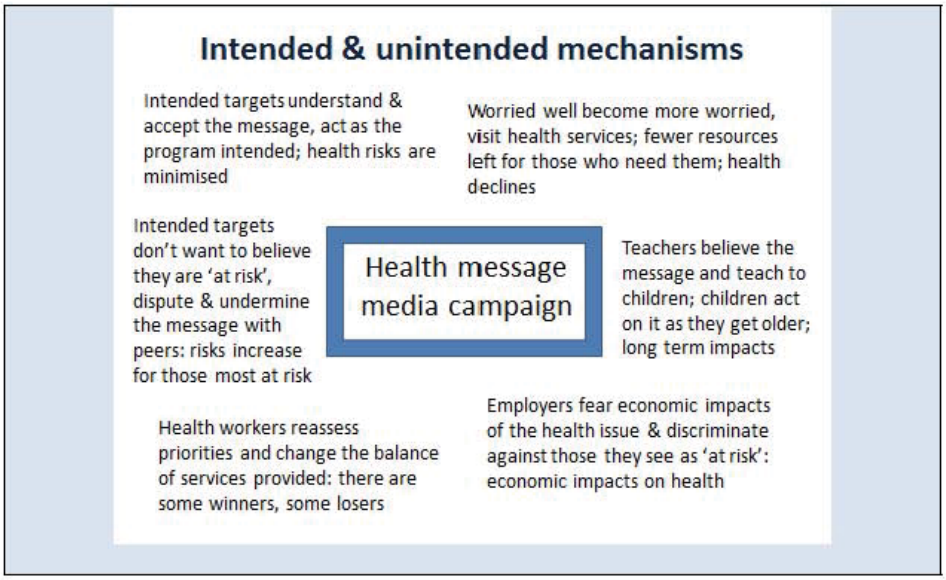
One route to identifying program mechanism is to reconstruct, in imagination, the reasoning of participants or stakeholders. When asked how the intervention was influential, a subject might reply, 'It made me ponder A, see alternative B, grasp opportunity C'. Mechanisms may also generate negative effects. So other participants may say, 'I've tried D previously, I'm bored with E, I prefer to do F'. Starting in this way generally reveals to researchers that they are dealing with a potentially a large number of mechanisms. The role of the realist researcher therefore necessarily involves identifying 'main mechanisms' – those that are common and significant enough to contribute to the pattern of outcomes of the intervention.

An important principle of realism is that the 'causes' of outcomes are not simple, linear or deterministic. Programs often work through multiple mechanisms. Some mechanisms are obvious and correspond to those intended by the program's designers, some are less obvious, and some are not anticipated by the designers. A mechanism is not inherent to the intervention, but is a function of the participants and the context. Consequently, the same intervention can trigger different mechanisms for different participants, even within one location. Programs run across very different social contexts are quite likely to generate different patterns of outcomes in those different contexts (see Box 1).

Box 1: An example of the intended and unintended mechanisms

Intended and Unintended Mechanisms

Consider the example of a health promotion media campaign (safe sex, safe needle usage, more fruit and vegetables, more exercise …). It is likely that the campaign intends to reinforce healthy behaviour by the majority of the population, but change the behaviours of those 'at risk'. However, the campaign is likely to trigger multiple mechanisms concurrently, for a number of different groups. Some of those mechanisms support the intended outcomes of the campaign, some undermine it, and some generate unintended outcomes. Some will operate over the short-term, some over the long term. See the results of a quick thought experiment in the diagram. Note that in each case we have identified (in very brief form) the 'reasoning' of the subgroup, the action that they take as a result, and the outcomes those actions may generate.



It is also important to note that different kinds of mechanisms operate at different levels of reality and through different kinds of programs (see Figure 2). Consider, for example, the different kinds of interventions that could be tried in an attempt to reduce domestic violence. Family therapy and Cognitive Behaviour Therapy (CBT) are different kinds of therapeutic approaches and work through different mechanisms. Family therapy (at least according to its program theory) works by changing family dynamics, which changes family members' experiences in the family, which changes their responses to experiences and thus their behaviour. CBT works, so the theory goes, by retraining the cortical and limbic systems in the brain, and does so regardless of whether the program focus is reducing family violence or overcoming addiction. Drug therapy works by changing chemical processes within the brain, while education and training programs work by changing knowledge, skills and/or attitudes. Legislation may work through deterrence or – where prison terms result – by temporary disablement (i.e. the offender cannot offend against the family members while imprisoned –although their violence may simply be directed to others (a mechanism known as displacement). Community development may work by changing community norms about acceptable and unacceptable behaviour or by changing bridging social capital, such that victims can access support to escape from violence. Note in each case that we have identified a program type or strategy, and that the mechanisms through which they work operate at a different level of reality than the actions of the programs themselves.

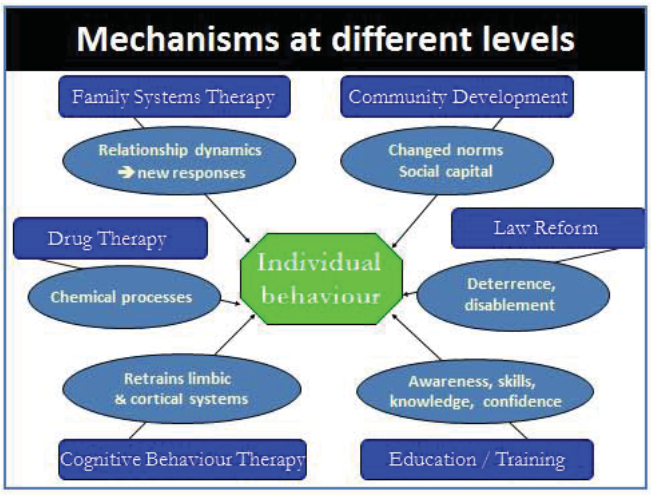


Figure 2: Mechanisms operate at different 'levels'

In summary, generative mechanisms:

- cannot be seen or measured directly (because they happen in people's heads, or at different levels of reality than the one being observed);
- are context-sensitive;
- are multiple (hence, when researched, they need to be unpicked, defined and prioritised);
- are best expressed at a somewhat abstracted level, so that they are not tied unnecessarily to particular people, places or things.

Context and Mechanism

Researchers will be familiar with the observation that varying outcomes occur in different contexts. The realist explanation for this variability revolves around mechanisms and their interactions with other mechanisms and context. Although the endless permutations and combinations of interactions might be expected to produce no observable patterns, the fact is that patterns do occur. Realism suggests that this is because similar mechanisms are being triggered in some contexts, producing the similar bits in the pattern; and different mechanisms are being triggered in other contexts, producing a different part of the pattern.

The term 'context' may refer to broad social or geographical features (for example the country in which an intervention operates and its cultures); to features affecting the implementation of programs (for example whether the program occurs in a jail, a hospital or health service, whether there is adequate funding, the qualifications of staff). It could also relate to the make-up of the participants on a program or the different population profiles of locations in receipt of an intervention. It could also relate to the conditions in which subjects seek to enact their choices (graduates of a vocational training program will find it easier to get work in a context of high employment; recipients of a housing subsidy will find it harder to use that subsidy in a context of housing shortages.) 'Context' in short can take on a multitude of forms. The realist use of context,

however, is not just a matter of listing the infinite potential 'surrounds' to an intervention. What matters is developing an understanding of how a *particular* context acts on a *specific* program mechanism to produce outcomes – how it modifies the effectiveness of an intervention. We now modify our earlier diagram (Figure 1) to represent this interplay and the patterns of outcomes that can be expected to result (see Figure 3).

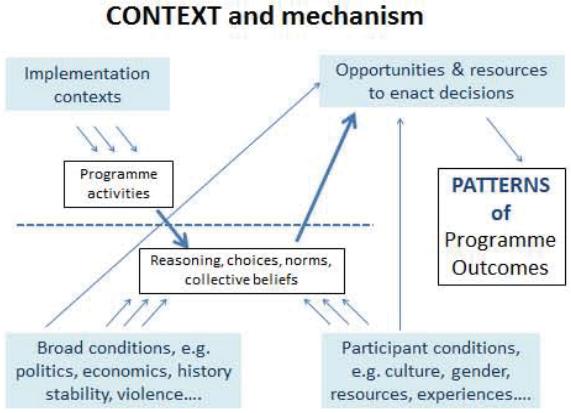


Figure 3: The interplay between context and mechanism

In summary, realism holds that mechanisms matter because they generate outcomes, and that context matters because it changes (sometimes very dramatically) the processes by which an intervention produces an outcome. Both context and mechanism must therefore be systematically researched along with intervention and outcome. By implication, research or evaluation designs that strip away or 'control for' context with a view to exposing the 'pure' effect of the intervention limit our ability to understand how, when and for whom the intervention will be effective.

2.2 Theory and Realist Synthesis

Realist review was developed as a theory based approach to synthesising existing evidence. It is, therefore, part of the school of 'theory-led' or 'theory-driven' research. 'Theories' rather than 'programs' are thus the basic unit of analysis. This move represents a considerable intellectual leap that is often misunderstood. Research begins with program theory and ends, if it has been successful, with a revised, more nuanced and more powerful program theory. Being disembodied from any specific application in this manner, research findings have the potential to be applied on any subsequent occasion that such a theory comes into application. This borrowing of policy ideas happens all the time. Some realists are fond of saying that there is nothing new under the sun when it comes to developing the ideas for interventions.

There are a number of reasons why theory is important. As we have already noted, theory is implicit in all programs. Using an abstracted program theory for the review helps reviewers to move beyond the minutiae of particular programs to focus on the major ideas within them. Once developed, reviewers can also compare the program theories (explicit or implicit) of the various interventions they examining to more general substantive theory, which can help in understanding the differences

between interventions and perhaps, therefore, the differences in how they work and the outcomes they generate.

Other reasons for using a theory-based approach include:

- There are always more questions that you could ask in a review than you will have the capacity to answer. Using theory helps to focus the review and to decide which questions to ask.
- There is always more literature that you could examine than you will have the resources to do. Theory helps to determine which literature is most relevant.
- The information gathered in a review always has to be interpreted and theory provides a guide for interpretation.
- The information gathered in a review is usually complex and messy: theory provides a basis for abstraction and for understanding 'the patterns in the data'
- Given that evaluations of the 'same' program (or kinds of programs) almost always show different results in different situations (or for different groups), theory provides a basis for explanation of the patterning of outcomes.
- Attributing outcomes to programs is complex. In primary evaluation, theory provides a basis for causal attribution. In reviews, theory provides a framework with which to assess the plausibility of attributions made by original authors.

However, the word 'theory' has many meanings. As Pawson and Tilley noted in their book Realistic Evaluation (3), is can refer to everything from grand over-arching theories such as Marxism to specific hypotheses that are tested in a laboratory experiment. In theory-based research and evaluation, there are four kinds of theory that matter (see Figure 4).

One is the underlying philosophy (realist philosophy). Realism takes particular positions both in relation to the nature of reality (that is, ontology) and the nature of knowledge (epistemology). This is realist philosophy (also sometimes called **realist theory**).

The second is **methodological (i.e. research and evaluation) theory**, or the implications for research and evaluation methodology that realist philosophy implies. One most often hears of 'realist evaluation' and 'realist synthesis' (or review), but other forms of realist research are also possible (see Section 8. Further reading and resources).

The third is **program theory**. This is the theory about what a program or intervention is expected to do and in some cases, the theory about how it is expected to work. Realist program theory goes a little further and includes descriptions of contexts, mechanisms and outcomes.

The fourth is '**formal theory**' or '**substantive theory**'. This is theory within particular domains – sociology, economics, psychology, education, health and so on. Examples include game theory in economics, constructivist learning theory in education, attachment theory in human development and so on. Sometimes substantive theories are used to design programs. They may also be used to inform program theory that is developed 'after the event' – for example, when evaluators develop program theories for programs that are already underway, or when reviewers develop the 'initial rough theory' for their review. Substantive theories are often used to help make sense of CMO patterns – to contribute to the 'synthesis' stage of a realist synthesis.

It is not necessary to have a strong command of all these types of theory before beginning a realist review. However it is worth remembering that the philosophical underpinnings of the various studies in a review may be different. It is still possible to undertake a realist review, regardless of the philosophical assumptions built into the program or the research methodologies used within the documents that are reviewed.

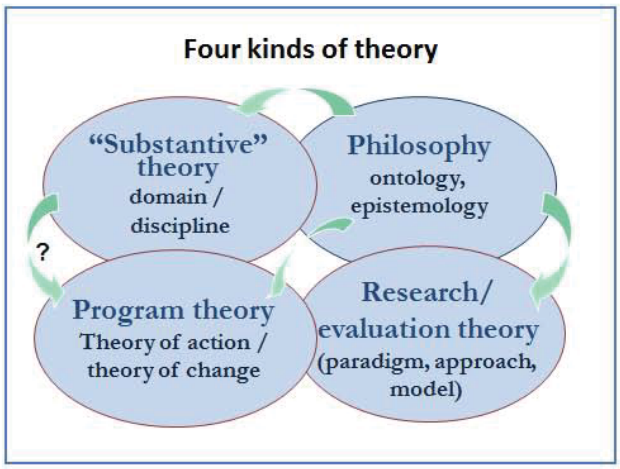


Figure 4: Main types of theory relevant to realist reviews

There are three other uses of the term 'theory' that are important in realist synthesis. These are 'initial rough theory', 'refined theory' and middle range theory.

The term **'initial rough theory'** refers to the initial sketch of a theory that is used to guide a realist synthesis. This is often a program theory. However, some questions for realist synthesis do not refer to a particular program. Here, the 'initial rough theory' sketches 'whatever it is that the question is investigating' (what would be called 'the evaluand' in evaluation) and how it is expected to work.

Initial rough theories may or may not be constructed in realist terms. Program theories provided by commissioners of reviews, for example, are not usually framed that way. To support the process of undertaking a review, it is at least useful to construct an initial rough theory of action (What is supposed to happen?) and an initial rough theory of change (Why is that supposed to work?). If it is possible to construct the theory of change in realist terms, so much the better.

A **'refined theory'** is the product of a realist review. In the process of conducting a review, some aspects of the initial rough theory may have been proved wrong. Others may have been supported with strong evidence. Many (perhaps most) will have been refined to some extent. Refining a theory might mean becoming clearer about the contexts in which, or population groups for which, an intervention will or will not work. It might mean developing a more sophisticated understanding of how particular mechanisms work. It might mean refining understandings of outcomes ('to what extent', 'in what respects', 'over what timeframe'). The product may be a set of CMOs – statements about the contexts in which particular mechanisms generate particular outcomes. The final stage of theory refinement is to develop an explanation of why the CMO patterns look the way they do.

Perhaps a useful analogy to use for the theory building and testing process of realist reviews is that of being a police detective. Both work 'backwards' and 'outwards' (by this we mean casting the net wide) from the outcome of interest. For the detective the outcome is the crime – say murder most dreadful. The detective starts to form ideas about how and why it might have happened. The former may seem obvious – a stab wound is seen, but that may not be the cause of death – and so the detective can speculate from the clues available at the crime scene, but cannot be sure until there is a post mortem (and perhaps not even then). The analogy here is clear – realist reviewers can speculate based on clues from (for example) the articles they read, but need to continue to seek data to confirm their speculations. The detective will search for more clues, interview suspects, obtain forensic evidence (such as looking for fingerprints on the murder weapon), all in a bid to confirm, refute or refine any initial speculations as to 'who dunnit'. This is analogous to the realist reviewer seeking more data – be it from formal searching or though lay or content experts. Both may cast the net wide in search of data to test speculations. Neither knows for sure what they are looking for, but follow leads or clues that show promise. Finally, neither knows for sure the 'truth' but instead builds up an evidence base to support their theories – a case that is judged on coherence and plausibility.

The term '**middle range' theory** refers to the level of abstraction at which useful theory for realist work is written: detailed enough and 'close enough to the data' that testable hypotheses can be derived from it, but abstracted enough to apply to other situations as well. This is a theory that lies;

> *"...between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change...*
> *It is intermediate to general theories of social systems which are too remote from particular classes of social behavior, organization and change to account for what is observed and to those detailed orderly descriptions of particulars that are not generalized at all. Middle-range theory involves abstraction, of course, but they are close enough to observed data to be incorporated in propositions that permit empirical testing.[our emphases]"* (4)

Middle-range theory can be thought of as a 'knowledge repository'. It holds promise for evaluation and systematic review methodology because it offers an approach to linking findings from program to program and from policy to policy. Interventions are normally targeted at specific outcomes in specific populations but the problems they deal with often have a common genesis. For instance, very many programs designed to change behaviour wrestle with the problem of persuading inveterate 'outsiders' (be they drug users, educational underachievers, excessive eaters, etc). to become reformed 'insiders'. We know that prizing people out of such out-groups is difficult and know that targeting 'marginal members' may be a more realistic aim. Accordingly, the opportunity for learning in realist synthesis is thus the middle-range task of trying to figure out what constitutes marginal membership of such groups. There is a healthy sociological and evaluation literature to draw on in accomplishing such a task and its accomplishment offers significant potential learning for the *next* behaviour change program.

Realist reviews may draw on existing middle range theories to develop the 'initial rough theory' that they test. The outcomes of a realist review are also ideally framed as middle range theory – that is, theory that can usefully be applied to a family of interventions, or to a problem that manifests in a number of domains. Program mechanisms are also usually described at a middle range level.

There is more detail about program theory in realist review in Section 4 below.

2.3 A Glossary of Terms

There are multiple definitions and descriptions of almost all the terms used in realist research and evaluation. Those that are provided here draw from a particular school that draws on the work of authors such as Popper, Campbell, Bhaskar, Sayer, and Archer (5). Our interpretations for the purpose of realist evaluation and realist synthesis draw heavily on publications by Pawson and Tilley (3), Pawson (1) and Pawson (5).

Context

"Context often pertains to the "backdrop" of programs and research. ... As these conditions change over time, the context may reflect aspects of those changes while the program is implemented. Examples of context include cultural norms and history of the community in which a program is implemented, the nature and scope of existing social networks, or built program infrastructure. ... They can also be trust-building processes, geographic location effects, funding sources, opportunities, or constraints. Context can thus be broadly understood as any condition that triggers and/or modifies the behavior of a mechanism." (6)

Context-mechanism-outcome (CMO) configurations

A CMO configuration is a statement, diagram or drawing that spells out **the relationship between** particular features of context, particular mechanisms and particular outcomes. In a sentence, they take the form of "In 'X' context, 'Y' mechanism generates 'Z' outcome." To more fully appreciate CMO configurations readers may wish to familiarise themselves with the concept of mechanisms (as defined below in this section).

"CMO configuring is a heuristic used to generate causative explanations pertaining to the data. The process draws out and reflects on the relationship of context, mechanism, and outcome of interest in a particular program. A CMO configuration may pertain to either the whole program or only certain aspects. One CMO may be embedded in another or configured in a series (in which the outcome of one CMO becomes the context for the next in the chain of implementation steps). Configuring CMOs is a basis for generating and/or refining the theory that becomes the final product of the review." (6)

More details about CMO configurations may be found on pages 21 to 27 in 'The Science of Evaluation: A Realist Manifesto' (5).

Demi-regularity

"Demi-regularity means semi-predictable patterns ... The term was coined by Lawson (1997), who argued that human choice or agency manifests in a semi-predictable manner - "semi" because variations in patterns of behavior can be attributed partly to contextual differences from one setting to another." (6)

Lawson's interest was with long-term social and economic change. He was concerned to distance realism from the idea that there are 'laws' of social evolution *and* from the alternative that there is nothing but patternless fluctuation. This perspective sits comfortably with the realist notion of program effectiveness. What should not be anticipated is the discovery of intervention panaceas, nor will outcomes be complexly haphazard. There will be some patterning. We should be able to discern broad lessons on for whom, in what circumstances and in what respects an intervention is more likely to succeed and these 'demi-regs' are the basis for decision support in the policy making process.

Hypothesis

"A hypothesis is a logical supposition, a reasonable guess, an educated conjecture. It provides a tentative explanation for a phenomenon under investigation." (7).

Hypotheses can be developed and used at many levels in realist research – for example, hypotheses about the main ideas in program theory, about mechanisms, about the aspects of context that will influence whether and how mechanisms work.

Mechanism

There are many definitions of mechanism. What they all have in common is that mechanisms generate outcomes. Examples include:

"Mechanisms are the agents of change. They describe how the resources embedded in a program influence the reasoning and ultimately the behaviour of program subjects." (p13)(5)

"...mechanisms are underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest. There are three essential clues located in a "realist" reading of mechanisms.
These are that:
1. Mechanisms are usually hidden;
2. Mechanisms are sensitive to variations in context; and
3. Mechanisms generate outcomes." (2)

In their book 'Realistic Evaluation' Pawson and Tilley (3) provide explanations for a number of aspects of 'mechanism':

- a generative process which creates or constitutes a regularity (p 67);
- which is located at a different 'layer' of social reality than the regularity it explains (p 67);
- an "underlying" process, which cannot usually be directly observed (p 65);
- operating at both micro (individual) and macro (social/structural) levels (p 65);
- involving both people's choices (agency) and "the capacities they derive from group membership" (structure) (p 66);
- demonstrating "how program outcomes follow from the stakeholder's choices (reasoning) and their capacity (resources) to put these into practice" (p 66);
- "propositions about what it is within the program which triggers a reaction from its subjects" (p 66)
- a hypothesis about how programs work which always works as "a 'weaving process' which binds resources and reasoning together" (p 66);
- "not variables or correlates which associate with one another; rather we are trying to explain how the association *itself* comes about" (p 67): "the mechanism is responsible for the relationship itself" (p 68).
- "not expressible as properties of the individual" (p 68)
- "A mechanism is thus not a variable but an *account* of the make-up, behaviour, and inter-relationships of those processes which are responsible for the regularity. A mechanism is thus a theory... which spells out the potential of human resources and reasoning." (p 68).

Middle-range theory

A theory that is specific enough to generate hypotheses (for example in the form of propositions) to be tested in a particular case, or to help explain findings in a particular case, but general enough to apply across a number of cases or a number of domains.

Stratified reality

Realists believe that the world we live in is layered or stratified. All around us real social structures and systems exist and operate independently of our conception of them. Events can be seen, but the mechanisms that cause them are not readily observable because they exist at a different layer or strata. Understanding them requires theory and abstraction. This way of thinking about the world and how it operates implies that reality is stratified.

"The stratification of reality in the philosophical ontology of CR [critical realism] has two dimensions. The first is the ... central distinction between the events that we can experience and describe, and the hidden, but nonetheless real, mechanisms behind them. The second dimension is that reality is assumed to consist of hierarchically ordered levels where a lower level creates the conditions for, but does not determine, the higher level. The distinction between the levels lies not in the entities, but in the generative mechanisms that operate at each level. It is not possible to reduce the causes of what occurs to one level to those of another level (whether lower or higher), because at each level something qualitatively new emerges ... . These levels and their causes form an open, interactive world of things and contingent tendencies, which, according to CR, constitutes the proper object of scientific investigation." (8)

The most basic distinction within the notion of a stratified reality is that between 'structure' and 'agency'. Programs are attempts to induce social change and it is important to understand how these different strata play a part in producing social transformation. A grand attempt to understand the anatomy of societal change can be found in 'realist' methodology, most especially in the works of Archer (9). Her theory of 'morphogenesis' attempts to answer the age old sociological chicken-or-egg question about what comes first in propelling social change – is it 'structure' or is it 'agency'? Put simply, her answer is 'chickenegg'. People's immediate actions are shaped within social structures in which they sit – communities, organisations, legal systems, power relationships, etc. However, in a longer time frame, these structures themselves change as a result of the activities and choices of the historically situated individuals who make them up. Social change, in short, occurs through a never ending cycle: 'structural conditioning' shapes 'social interaction' which in turn shapes 'structural elaboration', which then provides 'structural conditioning' and so on, and so on. Realists thus suppose that change is something that no one steers. It happens perpetually and of its own accord whenever people and groups reflect on their own position.

What happens, as with policies and programs, when someone or some institution tries to steer change? The same dynamic persists. The intervention sits alongside other structures (organisations, communities) that clamour for the agents' attention. Some agents will choose to respond to the program and in so doing will subtly change the nature of the intervention and its place in social structure. Realist evaluation and realist synthesis attempt to retain this idea of the stratified and temporal unfolding of change. Programs and their effects cannot be considered in isolation from the rest of society. They are shaped by and shape history.

Substantive theory

'Substantive theories' are existing theories within particular disciplines. They may be used to help understand interventions. For example, in the social sciences theories may deal with topics such as

'cognitive development', 'deviance control', 'incentivisation' or any of the wider ambitions of interventions.

Theory

There are multiple definitions for the word 'theory'. One simple definition is that, "A theory is an attempt to organize the facts – some 'proven', some more conjectural – within a domain of inquiry into a structurally coherent system." (10) For a discussion of the different types and roles of theory in realist synthesis, see section 2.2 above.

Section 3. Focussing reviews

Learning objectives for this section

- Explain the importance of focussing a review
- Know what constitutes good practice when focussing reviews
- Describe the steps that may help to focus your review

3.1 Why focus reviews?

A realist research question contains some or all of the elements of 'What works, how, why, for whom, to what extent and in what circumstances, in what respect and over what duration?'. Realist analysis is then used to answer the question. 'Realist analysis' requires the application of a realist philosophical 'lens' to data; it seeks to analyze data using realist concepts. Specifically, realism adheres to a generative explanation for causation - that is, an outcome (O) of interest was generated by relevant mechanism(s) (M) being triggered in context (C). It is this process that above all else distinguishes a realist review from all other review types.

As with any other review method, realism has to wrestle with complexity. The problems that interventions hope to deal with are complex and multifaceted. Programs themselves are complex and adaptive. The situations in which programs are inserted are complex and changing. It is impossible for any review method, realist or otherwise, to be comprehensive in covering all the contributory processes and contingencies. This brute fact sometimes runs against the grain for reviewers who bring to the exercise the idea that they are providing an 'overview' or that they operate with a census of 'all' relevant primary studies. Even if one is undertaking a commissioned review and that commission points the reviewer to a well circumscribed family of interventions of type 'X', the questions that could be asked are still infinite. It is always necessary to 'focus'. It follows that no one should be unduly defensive about the fact that a realist review will provide partial knowledge. The key is to prioritise and make the chosen lines of investigation absolutely clear.

Because a realist synthesis will generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it may be necessary to 'contain' a review. Many different aspects of a realist review might need to be focussed. Examples of *how* a review might be narrowed include:

- the question(s) to be answered (refining from broader to narrower)
- the aspect of program theory to be investigated
    - a sub-set of programs within a program family (e.g. routine screening rather than all health screening)
- scale of review (e.g. focus on particular countries in international development reviews, or cultures, or timeframes),
- the extent to which the review aims to be comprehensive
    - rapid review – using realist analytic processes within a more limited literature set
    - systematic review – aiming to include all evidence on the topic

Focussing may also take place at different time points in the review process. Not all eventualities can be anticipated at the start of a review and the theory 'axe' may be wielded too hastily. There is a normal process in any inquiry whereby new, unforeseen aspects are uncovered whilst one 'reads into' a project. The proper rhythm of inquiry is thus for its potential scope to widen before a well-

informed choice can be made on how and when to narrow it. Examples of time points *when* focussing may be needed include:

- when negotiating the research project or funding contract;
- while writing and negotiating the research protocol (where required for funding projects)
- when an advisory group is established
- when content experts are consulted
- when it becomes clear how much evidence is available for particular aspects of the question;
- when evidence suggests new pathways that could be explored

It is entirely legitimate for the synthesis' objectives, question and/or the breadth and depth of the review to evolve or be refined as the review progresses.

3.2 Quality standards for focussing reviews

When focussing reviews, we recommend the standards in Table 1 should be used.

Table 1: Quality standards for focussing the review

	Inadequate	Adequate	Good	Excellent
The review question is sufficiently and appropriately focussed.	The review question is too broad to be answerable within the time and resources allocated. There is no evidence that progressive focussing occurred as the review was undertaken.	Attempts are made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints. Attempts are documented so that they can be described in publications as appropriate.	Adequate plus: The focussing process is iterative. Commissioners of the review are involved in decision-making about focussing. Decisions made about which avenues are pursued and which are left open for further inquiry are recorded and made available to users of the review.	Good plus: The review team draws on external stakeholder expertise to drive the focussing process in order to achieve maximal end-user relevance.

3.3 Examples from the literature

Case study 1

An example of (initially inadequate) focusing can be found in an early review undertaken by one of us: 'Internet-based medical education: a realist review of what works, for whom and in what circumstances' by Wong et al. (11). In this realist review, the stated focus of the review was:

> *"...to [a] explain what sort of Internet-based medical education 'works', for whom and in what circumstances, ..."*

Whilst this review objective encompasses the expected elements of a realist question, it is rather broad – and potentially too broad to be answerable, even in the context of this review, which was as a doctoral thesis. A lesson here is that even in the relatively long time frames allowed for a thesis (in comparison, for example, to the tighter timescales of commissioned reviews), the temptation can be to be too ambitious in the scope.

Focussing in this review only occurred, much later on during data analysis and was driven more by chance than planning:

"As the review progressed we became aware of ... the emergence of two prominent demi-regularities [that] prompted us to narrow our review focus to the two candidate theories discussed below."

Whilst such a strategy does mean that the review is focussed on what was present within the included studies, there are potential problems that might impede progress. For example:

- it is theoretically possible that there are no identifiable demi-regularities in the included studies
- there may be many demi-regularities and it is not clear which are more prominent and should be pursued (in preference to others)
- the focus chosen by concentrating on "prominent demi-regularities" may not be important to the potential users of a review's results.

With the example above, if we now 'grade' it against the quality standards set out in Table 1, the focussing strategy used can at best be described as 'Adequate'. Adequate because progressive focussing has occurred and it is reported. One aspect of this review that at least tries to take into account the needs of other stakeholders can be found thus:

> *"Several previous systematic reviews and two meta-analyses have compared the efficacy and utility of Internet-based education with conventional teaching methods or no teaching [References x2]. Two main questions face researchers in this field: efficacy (can Internet-based medical education work, and if so what is the 'effect size' compared to conventional teaching?) and effectiveness (under what real-world circumstances does it actually work, and how might its impact and cost-effectiveness be maximised?)."*

Use is made of what is already known about the review's topic area to provide a very rough direction of focus. More iterative focussing, (for example) by consulting other educators on what they perceive their 'burning' unanswered questions to be might have helped improve the review further.

Case study 2

Other reviewers have approached the issue of focussing in different ways. In 'Implementing successful intimate partner violence screening programs in health care settings: Evidence generated from a realist-informed systematic review', O'Campo et al. have set out to uncover, "...why and how universal IPV [intimate partner violence] screening programs in health care settings are effective." (12). Two strategies are used to focus the review:

- the authors construct a program theory for IPV (Figure 1, page 856). This is then used to explain their initial focus – only on the initial steps of the IPV program theory, namely "...screening and risk assessment and identification of IPV Victims."
- the literature is consulted to check where most interest lies in IPV screening programs and this is identified as being in routine screening.

Case study 3

In Jagosh et al.'s review 'Assessing the outcomes of participatory research: protocol for identifying, selecting, appraising and synthesizing the literature for realist review' (13), the review team involves knowledge-user partners very early on to assist in formulating their review question and focus:

> *"Through the initial funding application process, the research questions were developed by the core group and sent to the [knowledge-user] partners to further define the aim of the proposed review according to their experiences and the priorities of their organizations"*

In summary, reviewers should try to iteratively refine the focus of their reviews. This process should be planned from the outset and negotiated with funders, so that they are aware of, and able to participate in, key decisions that will structure the outcomes of the review.

3.4 Learning activity

This activity is designed to provide practice in refining and focussing a realist review question.

Imagine that you have been asked to undertake a realist review to inform a new government policy to promote healthy eating in adults. Your colleagues propose that the review question should be: "What interventions promote healthy eating in adults?"

As a learning activity, you might like to try the following tasks:

- firstly, rewrite this question in realist terms;
- secondly, propose a first *refinement* to *focus* this question more narrowly; and
- thirdly, make notes about how you might go about further refining the question at particular later stages of the review.

3.5 Reflection activity

How a review is focussed will depend on a range of issues (for example, your research question, resource allocation, funder's expectations, end users' needs and so on). It is therefore impossible to be prescriptive and restrictive on what must be done. However, this does not mean there are not strategies that may be employed to start this process. To assist on this front, we have developed a list of questions a reviewer / review team might like to ask themselves: These questions are based on the quality standards in Table 1 and are listed in Box 2. We suggest that a reviewer might like to go through the questions in Box 2 to work out if the questions are relevant to their review and then how each question might be addressed.

Box 2: Questions to assist the focussing process in realist syntheses

- Can you complete your review within the time and resources allocated?
- Have you discussed the need to focus your review with (where relevant):
  - your supervisor?
  - within your review team?
  - your funding body / commissioners of the review?
  - potential users of your review?
- What processes will you develop and put in place to focus your review? For example:
  - 'What' will you focus?
  - 'When' will you do your focussing?

Section 4. Program theory

Learning objectives for this section

- Explain the importance of program theory
- Describe how program theories are developed
- Set out the steps needed to develop and use program theory
- Distinguish what constitutes good practice in using program theory

4.1 Developing and refining realist program theory

"Intervention are theories" (1)

Realist synthesis has most often been used to make sense of complex interventions. These interventions or programs often have multiple components (which interact in non-linear ways), outcomes (some intended and some not) and long pathways to the desired outcome(s). The term 'program theory' refers to an abstracted description and/or diagram that lays out what a program (or family of programs or intervention) comprises and how it is expected to work (also see section 2.2 and 2.3).

Program theory serves two main functions in a realist synthesis. The first is to 'sketch the terrain' that will be investigated, and in the process to assist in refining the elements and scope for the review. The second is to provide a structure for review findings.

There are many forms of program theory (see Funnell and Rogers (14)), but the general idea is to identify and map out:

- the key components (functions, strategies or activities) of the program;
- the outcomes is the program intended to generate;
- the components that contribute to particular outcomes. In some programs (but not all) it's useful to develop a rough sequence in which things need to happen, or a rough hierarchy of outcomes, in order to develop a sense of how the program is expected to work.

This can be done in a workshop with stakeholders, by reviewing program documentation (policy documents, funding applications, program descriptions and so on) or by reviewing a small selection of literature about the program type.

In multi-faceted (or multi-component) programs, where each element may trigger several mechanisms and various mechanisms will be affected by different contextual factors, the potential extent and complexity of the program theory can make the task seem overwhelming – or at least, it can sometimes be unclear where to start.

This is one of the stages where the previous topic (Section 3: Focussing reviews) is important. There are a number of ways to go about focussing the theory development stage.

One is to concentrate on specific outcomes of interest and then work backwards and 'outwards' from the outcome of interest to construct an initial rough theory.

Another is to spend a limited amount of time (perhaps a few hours) developing a very rough sketch of the program and then ask: Is there a particular component of this program on which everything else hinges? One aspect of the program that, if it fails, might bring the whole program down?

For example, programs which work by engaging a smaller group to pass on information or expertise to a wider group (peer education programs, training teachers to work differently with students, training managers to work differently in their staff teams) all share a common potential 'breaking point'. If the first group does not successfully engage the second group, the program will not 'work'. If a 'breaking point' exists, that might make an appropriate focus for thorough development of program theory.

A third way is to identify the areas of a very rough program sketch are already well understood, and to pick an area of the program theory that is not yet well understood, and focus attention there.

Once the focus is selected, the rough theory needs to be developed. This is not an easy task and alas, mostly goes unreported in publications. The realist reviewer above all else needs to be curious and critical. A useful heuristic is keep asking 'Why?' and 'How?' questions and to only stop when a sufficiently coherent and plausible set of theories has emerged.

For instance, when investigating a policy document, one will often be confronted with a rather glossy set of expectations that if 'X' is put in place 'Y' will follow. The document may be read in several ways:

1) Uncritically, much in the way that the author intends you to read it, where the relationship between X and Y seems obvious, automatic and sensible.
2) Cynically, where you delve immediately for the political ideology that gave birth to the program and the benefits to power-holders that may follow from it.
3) Critically, where you interrogate what the policy architect has said (or failed to say) about how X will relate to Y. You should always put some basic questions to the absent author – 'How does the program work? Why is it that you suppose X should bring about Y?'

Option 3 above describes the realist mind-set. You may be lucky. You may immediately find the odd explanation or two. The document might go on to suggest what it is about the proposed intervention that will make a difference to individual behaviour or community life. The author might well have made a credible link between the resources on offer in the intervention and the reasoning that follows in the minds of program subjects. Rarely will the notion of 'mechanism' be employed but, nevertheless, some nascent program theories may be evident.

However, if you keep on reading, a wider set of justifications for the program will unfold. They may well involve alternative and competing explanations for mounting the program. If you keep pursuing the 'why?' question a broader set of explanations will build. If you read other materials (those produced a little further down the policy chain), these program theories will become more detailed. Program practitioners are often deeply sensitive to the key realist question – what exactly is it about their intervention that makes a difference to participants in their program. Dig deeply here. After a while your reading will more than likely take you to oppositional accounts. For every policy protagonist there is an antagonist. Again, avoid purely ideological critiques (option 2) in favour of relentless chasing of the 'why?' question (option 3). The literature will often suggest reasons why the program resources might be misinterpreted, reinterpreted or simply ignored by the some members of target community. All programs have unintended consequences and here are the program theories that might begin to account for them.

A critical reading should continue along these lines. With increasing familiarity with the literature you (the reviewer) will be able to see similarities and differences in the emerging program theories. Realists are fond of saying that there is nothing new under the sun in the world of intervention theory and you may well be able to embed your emerging set of program theories into previous policy analysis. Academics are great classifiers and typologists and you may be able use their ideas to bring some order to the chatter of program theories you have uncovered. So begins the journey of program theory development.

The second main function of program theory is to provide a structure for review findings. Realist reviews bring together diverse sorts of evidence from diverse sorts of research. The 'nuggets' of evidence need to be aligned with the particular elements of the program theory to which they are relevant. The evidence will support some parts of the theory, refine other parts, and refute some parts all together.

Once the evidence is aligned, it is necessary to then 'step back' and synthesise it – to produce a refined theory that provides the portable lessons for translation to other circumstances.

The ideal theory resulting from a completed realist synthesis would consist of:

1) An outline of the contexts in which, populations for which, and main mechanisms by which, particular outcomes are achieved (that is, a description of the CMO configurations identified through the review);
2) One or more middle-range theoretical explanation of how and why particular mechanisms generate certain outcomes within certain contexts; and
3) A middle-range theoretical explanation of the *pattern* of outcomes found – why the pattern of CMOs looks the way it does. This usually draws on formal theory in the domain in which the review is being carried out.

The whole process of moving from initial rough theory to refine theory is summarised in Figure 5.

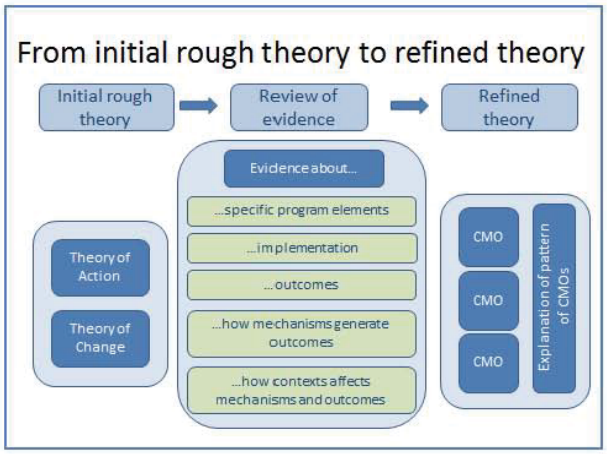


Figure 5: The process and sources of evidence used to refine theory

4.2 Quality standards for program theories

A program theory has an important role in a realist review and efforts to construct one should begin early in a review. Invariably any initial attempts at such a task will be tentative and will need to be progressively refined as the review progresses and understanding about the topic under study grows. For this topic area, we suggest that quality is defined as set out in Table 2.

Table 2: Quality standards for constructing and refining a realist program theory

	Inadequate	Adequate	Good	Excellent
An initial realist program theory is identified and developed.	A realist program theory is not offered or A program theory is offered but is not converted to a realist program theory at any stage of the review.	An initial program theory is identified and described in realist terms (that is, in terms of the relationship between contexts, mechanisms and outcomes). The refined theory is consistent with the evidence provided.	Adequate plus: An initial realist program theory is identified and described at the outset. The theory is refined iteratively as the review progresses.	Good plus: The relationship between the program theory and relevant substantive theory is identified. Implications of the final theory for practice, and for refinements to substantive theory where appropriate, are described. The final realist program theory comprises multiple context-mechanism-outcome configurations (describing the ways different mechanisms fire in different contexts to generate different outcomes) and an explanation of the pattern of CMOs.

4.3 Examples from the literature

Case study 1

The implications of not developing a program theory as a 'road map' are illustrated in the review by Daykin et al. (15). In this review, the reviewers are clearly dealing with a heterogeneous complex intervention – Patient Advice and Liaison Services (PALS) in England, United Kingdom – which has multiple components, long implementation chains and multiple intended outcomes. The question used to guide their review was:

> "*What context and mechanism factors have been identified as leading to favourable/unfavourable outcomes in patient and public involvement in NHS services?*"

Perhaps a first difficulty, in relation to program theory, lies with the structure of this question itself. It asks for context and mechanism factors to be identified, but it does not explicitly ask for an explanation of how contextual factors affect the operation of mechanisms[1].

[1] There may also be a clue in the term 'mechanism factors'. In realist theory, mechanisms are not 'factors' in the traditional sense of 'mediating and moderating factors', but causal processes.

Daykin et al. searched the literature on PALS and related interventions comprehensively and systematically and were able to identify a number of 'themes' in relation to each intervention type. For example, for PALS specific literature, they identified four themes: Resources for PALS; Accessibility of PALS; Priorities for PALS users; and Organisational issues (pages 61 to 62). Here perhaps lies a second issue in relation to program theory. It is not clear whether the 'themes' are intended to be contextual factors or causal processes, and so it is not clear what role they play in a realist explanation of outcomes.

There are instances where relationships between these elements are clear. For example, 'Accessibility of PALS' is reported to contribute to "take up" (an intermediate outcome) and one element of accessibility is 'physical location' (a contextual factor). However, 'take-up' may also be affected by a range of other factors, and those relationships are not clarified. Our suggestion is that developing a program theory could have assisted the team to explain the relationships and influences of between the various components of the program and the 'themes' and between those themes and the pattern of program outcomes in different contexts. It may also have assisted greater integration of the review's findings across PALS and other related programs. This is one of the goals of using program theory, rather than focusing on programs per se: it allows synthesis of learning across similar programs.

Case study 2

McMahon and Ward's realist review, 'HIV among immigrants living in high-income countries: a realist review of evidence to guide targeted approaches to behavioural HIV prevention', sought to develop a program theory from the outset (16). Their review's initial goal was to examine and make sense of the adaptations that are made to interventions trying to change behaviour of immigrants living in high-income countries so as to reduce HIV transmission. We have used this example to illustrate:

- use of the literature to inform the initial stages of theory building
- theory refinement – making it more specific and testable
- testing theory

This review also demonstrates the idea raised in 'focusing the review' above – that any single review will only be able to examine part of an entire program theory in any detail.

Scoping literature and initial theory development

When scoping the literature, the authors found that the predominant thinking behind adapting interventions related to making the interventions "culturally appropriate". Their initial attempts to develop a program theory therefore sought to explain and make sense of adaptations to enhance cultural appropriateness and the responses of individuals to those adaptations.

This provides a clear example of the process of focussing reviews, undertaken during literature scoping. As a result, other kinds of adaptations not related cultural appropriateness were deliberately excluded.

The authors used a "known set" of articles (which "... were comprised of studies known to the lead author and studies found in preliminary searches of databases in the initial stages of the review") to construct their initial, generic theory. The initial theory (see Figure 6) outlined the relationship between what was done (Intervention Adaptation Activity), the Theorised Mechanism of Adaptation Activity and the Anticipated Response to Adaptation Activity (outcome), and included 'counter-mechanisms' (what the authors have called Potential Resistance to Adaptation Activity). Whilst this initial program theory is portrayed in a linear fashion, they acknowledge that:

"It is important to note that in reality these 'chains' can operate in non-linear and unpredictable ways depending on the context [Reference x1]. Here, for simplicity, the implementation 'chain' is presented in a linear 'path' with the participant response and participant resistance represented as outcomes that point in different directions."

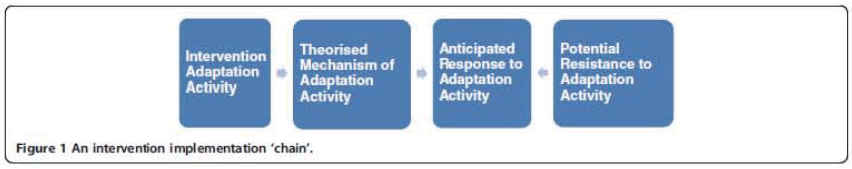


Figure 6: An intervention implementation 'chain' sourced from McMahon and Ward (16)

Theory refinement

The first step in theory refinement was to identify what sort of adaptations were made to interventions to make them culturally appropriate. McMahon and Ward grouped the adaptations they identified under the following headings: 'staffing', 'language', 'content', 'ethnic diversity', 'settings', 'community consultation' and 'priority setting'. For each adaptation group, they then inferred possible mechanisms – their initial list of inferred mechanisms being – 'authenticity', 'understanding', 'consonance', 'specificity', 'embeddedness', 'endorsement' and 'framing'. In effect, this process made the theory of cultural appropriateness more specific and testable (hence more middle-ranged). The authors provide a detailed account of how they went about constructing their initial generic program theory in their main article as well as in an Additional File (http://www.systematicreviewsjournal.com/content/supplementary/2046-4053-1-56-s3.pdf).

Testing theory

McMahon and Ward's review explicitly set out to test and refine their theory. A program theory was specified for each inferred mechanism, using the same structure as the initial generic theory. For example, the adaptation activity of "integrating cultural values and elements into the intervention context" works through the inferred mechanism of 'consonance'. The outcome of interest is that "... immigrants understand intervention in a symbolic sense – 'shared values'" and the important influencing context for 'consonance' is titled "balance 'old country' and 'new country values' (see Figure 7). Within each of their included studies, data (confirming and disconfirming) were sought to identify whether the inferred mechanism was in operation and to understand its relationships to context and outcome. The authors provide a detailed discussion of how they have tested and refined their program theory for four (out of the seven) inferred mechanisms in their paper.

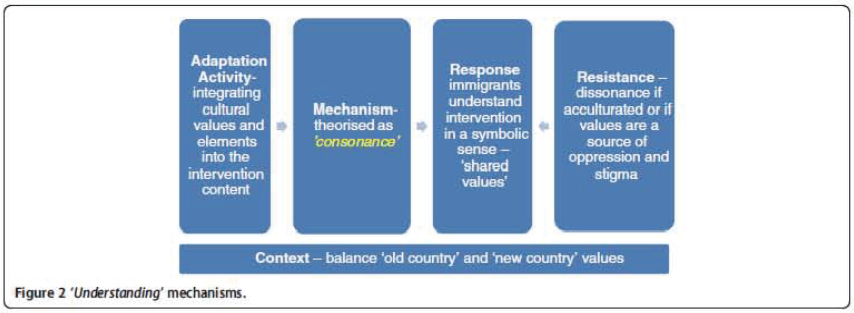


Figure 7: 'Understanding' mechanisms sourced from McMahon and Ward (16)

4.4 Learning activity

This activity is designed to provide practice in developing initial program theory. Note that we have constructed the activity to operate 'backwards from intended outcomes'. It is equally possible to construct a theory 'forwards from intended activities'.

Choose a program or intervention in an area of interest to you. The program should be large enough and well-established enough to have generated both policy and research documentation.

Collect three documents about that program.

1. A formal policy or program description produced by the central agency for the program.
2. Implementation: either instructions or guidelines about how to implement the program at the local level or a piece written by a practitioner about their experiences in implementation of the program.
3. Research or evaluation report about the program.

Begin by reading the policy or program document. Note each of the outcomes that the program is **intended** to achieve. For this activity, constructing a diagram might be helpful.

(For an additional challenge, consider whether there is an 'order' to these outcomes. Do some need to be achieved earlier for later ones to be achieved? Are some to be achieved for individuals and others for communities? Using short names for each of the outcomes, organise the outcomes into a format that makes 'logical sense' to you. If you are constructing a diagram, put these at the 'end' of the diagram).

Next, select **one** of the outcomes that the program is expected to achieve. Ask yourself: how or why is this outcome expected to be achieved? Check sections or sentences that describe the rationale for the choice of strategy. Note too that the choice of strategy itself contains a clue (strategies or activities usually reflect the perceived 'solution' to the perceived nature of the 'problem'). If intended causal processes are not immediately apparent, additional questions might include: Who is expected to do what differently, in order for this outcome to be achieved? What different choices or decisions would they need to make, in order to do that? What will the program do or provide to assist them to do that?

Make notes or give titles to each of the main ideas about how or why the outcome is expected to be achieved (there's often more than one). If you are constructing a diagram, add each one separately. Draw arrows between the elements to show how they relate to each other and how they contribute to the outcome you are considering.

Pause and consider the diagram so far. There may be a number of elements comprising one chain that contributes to the outcome. See if you can give that chain overall a name – a noun or short phrase that describes the overall causal process at work (e.g. 'deterrence' in crime prevention; 'changing norms' in community development; 'the Hawthorne effect' in research). At one level of abstraction this chain constitutes a mechanism.

In order to describe the mechanism in more depth, ask yourself: how or why does each link in the chain work? (This has previously been described as working out 'what lies beneath the arrow'). These causal processes 'below' each link are also mechanisms, but viewed at a greater level of detail. (One of the characteristics of realist explanation is that one can 'zoom in' and 'zoom out' to different levels of abstraction, using the same basic explanatory framework). Continue asking 'how' and 'why' questions until a sufficiently coherent and plausible explanation has been developed.

Now read the other two documents you collected. Do they mention the outcome that you were considering? What do they tell you about it? More importantly, do they suggest changes to how do they change the description of how the outcome is achieved? Amend your sketch or notes to reflect these changes.

As you read the two additional documents, you might also start to collect clues about the contexts in which the intended processes do and do not work as intended. These too can be named and added to the sketch of program theory. This is the beginning of the process of theory refinement. In a review, evidence about each aspect of the initial theory is collected and the theory is gradually refined in the light of that evidence.

4.5 Reflection activity

To assist program theory development, we have put together a list of questions that review teams might like to ask themselves. These questions are based on the quality standards in Table 2 and are listed in Box 3. We suggest that reviewers might like to go through the questions to work out whether the questions are relevant to their review and then how each question might be addressed.

Box 3: Questions to assist constructing and refining a realist program theory

- Do you need to construct a realist program theory for each outcome of interest? If not, why not?
- What sources and resources (e.g. other researchers, experts, service users) will you draw on to help you develop your realist program theory?
- What processes will you develop and put in place to:
  - develop *and*
  - iteratively refine your program theory / theories?
- Are there existing substantive theories that are will help to inform your program theory / theories?
- What assumptions are built into the program theory?:
  - What assumptions are we (the reviewers) making?
  - What assumptions are there in the data?
  - Which ones do we need to challenge and why?
- What data, from where, might help to test and refine the theory?

Section 5. Developing a search strategy

Learning objectives for this section

- Explain the importance of developing a search strategy that meets your review questions' needs
- Define what constitutes good practice for developing searches for realist syntheses
- Develop and use a search strategy for realist syntheses

5.1 Search strategies suitable for realist syntheses

What constitutes 'the right evidence' is different in a realist synthesis than it is in other form of review. Data that may usefully contribute to a realist synthesis are:

- not decided by research type (e.g. randomised controlled trial (RCT)) but by relevance to the review question;
- not restricted to research into or evaluations of programs per se, but related to the program theory that underpins the program;
- not necessarily about the whole research question, but relevant to a sub-section of it;
- not necessarily drawn from a whole text/document, but from a sub-section of it relevant to a particular aspect of the review question;
- able to shed light on any aspect of C, M or O for any element of the theory;
- different for theory building (which does not need to be as rigorous) as opposed to theory testing (which needs to be sufficiently rigorous to support the conclusion being drawn on for the review).

There are two related processes when trying to find the 'right' evidence for realist reviews. One process is searching (this section) and the other is about making judgments on whether or not a data should be included (Section 6).

While initial searches are undertaken to develop theory and need not be as systematic, the searching processes used to test theory should be more systematic and transparent.

Searching for theory testing should be guided by the objectives and focus of the synthesis and revised iteratively in the light of emerging data. Data relevant to a realist synthesis may lie in a broad range of sources that may cross traditional disciplinary, program, and sector boundaries. The search phase is thus likely to involve searching for different sorts of data, or studies from different domains, with which to test different aspects of the provisional theory.

Search methods using forward and backward citation tracking may be particularly valuable in finding the documents necessary to refine and test provisional theories. Realist syntheses do not exclude sources solely on the basis of their study design; hence, 'methodological filters' (e.g. to identify randomised controlled trials) add little to the search and run the risk of excluding relevant papers (see Section 6 for an explanation about inclusion process and how this impacts on searching).

Searching is likely to need to be iterative because, as the synthesis progresses, new or refined elements of theory may be required to explain particular findings, or to examine specific aspects of particular processes. As new elements of theory are included, searches for evidence to support, refute, or refine those elements may be required. Imagine, for example, that the review of interventions to improve healthy eating by adults described in learning activity in section 3.4 above revealed that different cultural groups respond in different ways to a particular intervention. It may be necessary to undertake additional searches to gather more information on aspects of those cultures in order to theorise and explain this pattern of responses. A single pre-defined search is unlikely to be sufficient and may suggest insufficient reflection on emerging findings. The important judgement centres on whether or not the searches carried out in a review were likely to have located the sources needed for further theory development and/or testing.

A final point about search strategies for realist reviews addresses the balance between a search process being comprehensive versus theoretical saturation. Comprehensive searches set out to find (as much as is practically possible) each and every document on the topic of interest. For example, searches in Cochrane systematic reviews are designed to be comprehensive. In a realist synthesis, the ultimate product is explanatory theory. Providing the necessary evidence to demonstrate that a theory is coherent and plausible does not necessarily require the unearthing of every document about that theory. Thus in a realist synthesis, searching can be stopped if saturation is reached – that is a judgement can be made to stop searching if sufficient evidence is found such that it is reasonable to claim that the theory is coherent and plausible.

The extent to which theoretical saturation and comprehensiveness are goals in a particular review will be influenced by the specific question the review is addressing. A review that seeks to understand and build theory about the effects of context in relation to a specific program will need to be more rigorous about ensuring that research and evaluation documents about that specific program are included, so that important contexts are not overlooked. This is an example of a particular type of 'comprehensiveness'. A review that seeks to understand the operations of mechanisms at a broader level (such as Wong et al.'s review of smoking in cars, below in section 5.3) will prioritise theoretical saturation over comprehensiveness.

Many review teams have significant searching expertise. The key is to ensure that such expertise is adapted to the requirements for a realist review.

5.2 Quality standards for search strategy

For this topic, we suggest that quality is defined as set out in Table 3.

Table 3: Quality standards for developing a search strategy

	Inadequate	Adequate	Good	Excellent
The search process is such that it would identify data to enable the review team to develop, refine and test program theory or theories	The search is incapable of supporting a rigorous realist review. Common errors include: • The search is driven by a methodological hierarchy of evidence (e.g. privileging RCTs) rather than the need to identify data to develop, refine or test program theory/ies • The search process is not informed by the objectives and focus of the review • The database(s) selected are narrow in the subject matter that they contain (e.g. limited to specific topics rather than extending to social science, psychology etc.) • Searching is undertaken once only at the outset of the review and there is no iterative component	Searches are driven by the objectives and focus of the review. The search strategy is piloted and refined to check that it is fit for purpose. Documents are sought from a wide range of sources which are likely to contain relevant data for theory development, refinement and testing. There is no restriction on the study or documentation type that is searched for	Adequate plus: further searches are undertaken in light of greater understanding of the topic area. These searches are designed to find additional data that would enable further theory development, refinement or testing.	Good plus: the searching deliberately seeks out data from situations outside the program under study where it can be reasonably inferred that the same mechanisms(s) might be in operation.

5.3 Examples from the literature

Case study 1

Connelly et al.'s synthesis set out to "...present practice-relevant guidance on interventions to reduce at least one measure of adiposity in child populations that do or do not contain overweight or obese children."(17). They undertook a comprehensive search, drawing on the literature to guide

the development of their searching. They also searched more than one database and the reference lists of included trials.

> *"We used the search strategy and inclusion criteria described by Summerbell et al. [Reference x1] applied to the following electronic databases: Medline, Embase, Cinhal, PsycINFO up to 30 April 2006. This identified randomized, controlled trials or controlled trials of interventions to prevent overweight or obesity in populations that included non-overweight children with or without overweight or obese children. Trials had to include an outcome that measured an index of adiposity ... We also searched the reference lists of included trials and published reviews for potentially relevant studies."*

Only one search was undertaken, only randomised controlled trials were included and the authors did not perform any preliminary searching to develop a program theory. The databases the authors searched were likely to contain additional relevant data, as the topic areas for their review are in health and psychology. Improvements the authors could have made to their search strategy to make it more likely to find data that would inform a realist synthesis include:

- Not using a methodological filter to decide which trials to include (i.e. consider including trials other then randomised controlled trials);
- Developing a program theory for interventions to reduce adiposity in children and basing their search strategy on finding the data needed to test and refine their program theory;
- Searching more than once, guided by the need to seek more data to test and refine that theory.

Case study 2

Morgan's review set out to, "...identify some of the underpinning factors that promote the development of evidence based health policy."(18). Only one health related database, PubMed, was searched. Data sources that contain data on health policy are likely to be found outside of the 'health' databases. A potential improvement to this search strategy would be to search in more than one database that is likely to contain relevant data.

Case study 3

In Jackson et al.'s review of the program 'Moving To Opportunity' (MTO), their iterative search strategy included three databases; four search engines; the reference lists of included articles. They also consulted external expertise. All of this was driven by the needs of the review:

> *"A "snowball" approach was used in which one reference led to others. Other evaluations were revealed through correspondence with Dr. Jeffrey Kling, one of the principal MTO researchers. ... Additional literature was also accessed to help understand key concepts and issues raised through the review including housing theories, studies of poverty, housing and health, and social determinants of health."*(19)

Case study 4

In common with other review methods, piloting of a search is an important process and should be undertaken where possible. The purpose of any piloting is to check that the search has the ability to find documents that have already been identified. The process of using a 'known set' of documents to develop and refine a search is clearly described in McMahon and Ward's review (see the 'Systematic searching for primary studies' section of their paper) and can be a valuable way of testing if a search strategy is able to find relevant documents (16).

Case study 5

Searches can sometimes reveal a lack of evaluation studies on a topic area. In the review undertaken by Wong et al., the goal was to develop a, "...framework of threats to the program theory of public health legislation..." and then to test this framework using the case example of banning smoking in private vehicles carrying children. A threat they wanted to understand better was that of enforcement of such legislation. However, they report that they were:

> *"... unable to find any formal studies that evaluated the enforcement of smoking bans in vehicles carrying children, we deliberately chose to seek out studies which examined the closely related topics of the enforcement of cellular phone use and child restraints in vehicles. Our logic for searching in these areas were that they involved enforcement of 'in vehicle' behaviours that were potentially equally hard to enforce and (in the case of child restraints) involved the safeguarding of children."*(20)

The rationale behind such a search was based on the realist principle of causal mechanisms. The assumption was that it may be reasonable to extrapolate from studies of cellular phone use and child restraints in vehicles because similar mechanisms *may* be in operation. The Wong et al. paper provides only a brief explanation of whether a search directed by mechanism can yield helpful transferrable lessons. A more detailed analysis is provided in Pawson et al.'s paper, starting at 'Is the Law Enforceable?' on pages 536 to 541 (21).

Learning and reflection activities

The learning and reflection activities for this section have been combined with those of section 6 (see sections 6.4 and 6.5).

Once the documents have been 'screened in', assessment of the data within the document begins. In any document, there may be several pieces of data that serve different purposes, such as helping to build one theory, refining another theory and so on. An appraisal of the contribution of any section of data within a document should be made on the same two criteria of *relevance* and *rigour*.

6.2 Quality standards for selecting and appraising documents

For this topic area, we would expect quality to be defined as set out in Table 4.

Table 4: Quality standards for selection and appraisal of documents.

	Inadequate	Adequate	Good	Excellent
The selection and appraisal process ensures that sources relevant to the review containing material of sufficient rigour to be included are identified. In particular, the sources identified allow the reviewers to make sense of the topic area; to develop, refine and test theories; and to support inferences about mechanisms	The selection and appraisal process does not support a rigorous and complete realist review. For example: • Selection is overly driven by methodological hierarchies (e.g. the restriction of the sources to RCTs to the exclusion of other forms of evidence) • Sources are appraised using a technical checklist for a particular method (e.g. assessment of quality for an RCT) rather than by making a defensible judgement on the relevance and rigour of the source • Selection and appraisal processes are overly restrictive and exclude materials that may be useful for a realist analysis • Selection and appraisal processes are not sensitive enough to exclude irrelevant materials	Selection of a document for inclusion into the review is based on what it can contribute to the process of theory development, refinement and/or testing (i.e. relevance). Appraisals of rigour judge the plausibility and coherence of the method used to generate data.	Adequate plus: During the appraisal process limitations of the method used to generate data are identified and taken into consideration during analysis and synthesis.	Good plus: Selection and appraisal demonstrate sophisticated judgements of relevance and rigour within the domain.

6.3 Examples from the literature

Case study 1

The use of checklists to judge the quality of an entire document can mean that documents may be excluded before they are even analysed for relevance and rigour. An example of this practice can be seen in Carr et al.'s report (22). Their health technology assessment review set out "... to identify, describe, classify and analyse the range of models developed to date for delivering health-related lifestyle advice (HRLA), or training, for effectiveness, mechanism of effect, cost-effectiveness, equity and acceptability in improving the health and wellbeing of individuals and communities, with particular reference to the reduction of inequalities in the UK." They undertook an exemplary and comprehensive search that identified a large number of potential documents for inclusion on this complex topic.

As can be seen from the figure 8 below, from 269 potentially eligible studies, 243 were excluded after a range of quality assessment checklists were used (these were for quantitative studies, the *Quality Assessment Tool for Quantitative Studies* and the Critical Appraisal Skills Program (CASP) checklist for qualitative research).

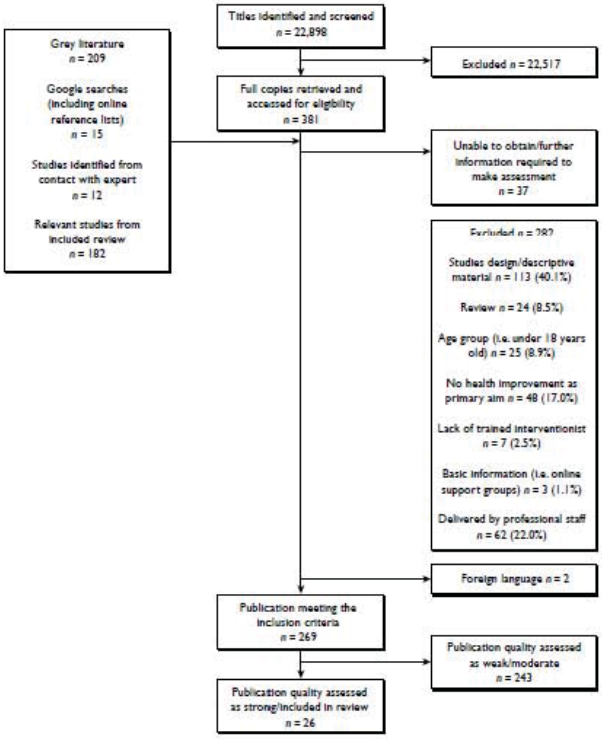


Figure 8: Study selection process figure sourced from Carr et al.

The quality checklists meant that of the 26 included studies, 23 were RCTs and there was one ethnographic study, one process evaluation and one controlled before and after study. The conclusions the authors were able to reach were mixed and inconclusive. The complex nature of the family of interventions the authors set out to study would of course produce a significant challenge to any review team. However we suspect that the overly restrictive inclusion process also contributed to limiting the data that could have been available for the realist review component of this review.

Case study 2

Selecting documents based on methodological hierarchies (in particular restricting sources to RCTs) restricts the range of evidence available, to the detriment of how informative the findings can be.

Kane et al. specifically wanted to explore what kind of a contribution randomised controlled trials (RCTs) can make to a realist review.

> *"Since randomised control trials (RCTs) have high internal validity, in this paper we review RCTs of interventions involving CHWs for improving child health in LMIC from a realist perspective with the aim to see if the RCTs can yield insight into the working of the CHWs [community health workers]."*(23)

Their conclusion is self-explanatory and clearly highlights the limitations they found in only including RCTs in a realist synthesis:

> *"The RCTs under review offered a fair amount of information about the interventions, only some information about context - allowing us to formulate only generic hypotheses. Disentangling context from intervention elements was a daunting task, particularly when doing this across RCTs. ...*

Authors seldom described or discussed the mechanisms that explained their study outcomes. We realise that the RCT design, the exacting reporting requirements and word limits of journals, restrict authors from sharing all their operational experiences. In addition RCTs tend to report average effects and not differential effects of interventions, and less so of the context and rarely of the mechanisms triggered by their interactions. This makes the RCTs less useful for answering the questions regarding how interventions work. These generic hypotheses seem to be recurring in the literature, however they have not been explicitly tested across contexts."

Case study 3

The article, 'Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy' provides an opportunity to look in more detail at how the selection and appraisal of documents works in a realist review (21). This paper provides a more methodological analysis of the paper by Wong et al. mentioned above (see Case study 5 section 5.3). The purpose of their review was to develop a, "...framework of threats to the program theory of public health legislation..." and then to test this framework using the case example of banning smoking in private vehicles carrying children.

Their program theory was expressed in a series of questions (see Box 2 page 520 in the 'Known Knowns...' article).

Their second question and its sub-questions were:

> *"2. Is there likely to be public support for such a law?*
> *2.1. What is the overall magnitude of support for such a law?*
> *2.2. What are the levels of support among smokers?*
> *2.3. What is the motivation behind public support?*
> *2.4. Does endorsement depend on the extent and success of previous smoking bans?"*

We can see that the data needed to address each of these questions is very different, because of the nature of the question asked. For question 2.1, data from surveys was used.

However, whilst relevant, such data are not without their problems - the issue of rigour being raised here. Pawson et al. offer warnings about the challenges of survey data:

> *"There are two familiar problems with such materials—attitudinal responses on health matters can be unreliable and the data, perforce, provide only a snapshot of*

> *opinion at particular time and place. Survey responses can never be taken entirely at face value. Well-known technical problems exist due to the slipperiness of question wording. … questions carry subtle differences of emphasis that might shape the willingness to support a ban. Public compassion might well differ for "children," "children under ten" and "preschool' children, not to mention the "elderly," "pregnant women," "nonsmokers," and so on. Probably, even more of a threat in the present case is the "social desirability effect." Respondents, naturally enough, prefer to be on the side of the angels and thus often "fake good" when confronted by a stranger asking questions about sensitive topics [Reference x1]. Put in a nutshell, the problem is that smoking addicts, who suffer routine stigma on top of slow poisoning, may well choose to dissemble."* (p528)

Thus while survey data may be relevant, caution needs to be exercised as to their credibility and trustworthiness. The implication is that inferences made on the basis of evidence have to take into account the rigour by which the data were generated. In the example we have used, the authors use the survey data not as definitive data that settles any argument, but more as part of an explanation building process:

> *"A sizable number of studies have shown significant levels of support for a ban in smoking in cars carrying children. …. In this case, the solidity of smokers' support is attested in further evidence on the grounds for that support, namely, their beliefs about the vulnerability of children, their sentiments of regret about taking up smoking and their acknowledgment that public sympathy for the smoking habit has declined under incremental legislation."* (p531)

6.4 Learning Activity for Sections 5 and 6

*This activity is designed to provide practice in developing a search strategy **and** selecting and appraising documents for a realist review.*

Searching
Return to the rough program theory that you developed in learning activity 4.4 (or any other program theory that is of interest and available to you).

Choose one 'strand' (or sub-section) of that program theory.

Write a list of search terms that you could use to construct a search to test that strand of theory. Note that these terms will relate to the elements within the strand of theory, not just the name or type of program that the review deals with.

Selection for relevance
Make a list of the kinds of evidence that might be used to test that strand of theory. Might it be evidence from opinion surveys? Ethnographic studies? Program evaluations? Census data? Administrative data from programs? Align each type of evidence with the particular element of the strand of program theory for which it would be appropriate.

Assessment of rigour

For each of the types of evidence you could use, make a few brief notes about the issues that might affect data quality or rigour that you would need to take into account during appraisal of the documents.

6.5 Reflection Activity for Sections 5 and 6

To assist you in developing a suitable search strategy and in selecting and appraising documents appropriately, we have developed a list of questions a reviewer / review team might like to ask themselves. These questions are based on the quality standards in Tables 3 and 4 and are listed in Box 4. We suggest that a reviewer might like to go through the questions in Box 4 to work out if the questions are relevant to their review and then how each question might be addressed.

Box 4: Questions to assist developing a search strategy and selection and appraisal of documents

Developing a search Strategy

- How will you ensure that your search process is such that it would identify data to enable the review team to develop, refine and test program theory or theories?
- Is the necessary searching expertise available to you? If not, what will you do to remedy this?
- Will your search be piloted and refined?
- Will further searching be undertaken if more data are needed?

Selection and appraisal of documents

- Are *relevance* and *rigour* being used to guide the selection and appraisal process? If not, why not?

Section 7. Applying realist principles in analysis

Learning objectives for this section

- Apply realist principles in analysis
- Deal with differential patterns of outcomes in analysis
- Define what constitutes good practice in applying realist philosophy and realist logic in analysis

7.1 The implications of realist philosophy for analysis

Realism is not one unified philosophy of science, but most schools within it subscribe to a number of core assumptions. These include generative causation (that is, mechanisms generate outcomes), the contingency of mechanisms (that is, mechanisms are context sensitive), the existence of many mechanisms operating at all levels of reality (which may or may not act and may or may not interact depending on the context) and a stratified reality.

The notion of 'stratified reality' has two meanings in realist philosophy. The first is that all systems have sub-systems and all systems are part of larger systems. That is, all systems are embedded within other systems. Causation operates both upwards and downwards through these systems. The second meaning of 'stratified reality' lies in Roy Bhaskar's philosophical constructs of three domains in reality: the real (the domain in which mechanisms exist, whether or not they are observed and whether or not they are currently operating), the actual (events) and the empirical (what is observed by humans) (24). This second meaning informs realist understanding of how causation works and distinguishes between what exists and what we can know about it. The first meaning – open, embedded, interactive systems – is central to the process of analysis in actual cases.

These assumptions constitute a realist philosophical 'lens' and lie at the heart of a realist analysis. It is this that distinguishes a realist review from all other review types.

A realist philosophical lens has distinct implications for the nature and process of analysis in a realist review. These include:

- the purpose of a review
- the basic explanatory structure used within a realist review
- the analytic tasks that follow from that explanatory structure
- the logic of comparison within realist analysis
- the relationship of evidence to theory
- the conceptual tools used to deal with evidence, and in particular, apparently conflicting or contradictory evidence

Some of these have been discussed earlier in these materials but are summarised briefly here in order to demonstrate 'the logic of analysis' overall.

Purpose

The purpose of a realist review is explanatory, not simply descriptive. It seeks to explain how and why programs generate different outcomes in different contexts. The aim is to assist policy makers and practitioners to make decisions about whether to use particular kinds of programs to achieve

their intended goals in particular contexts and how to adapt them to those contexts to increase their chances of success.

Explanatory Structure

The basic explanatory structure in realism is generative causation - that is, an outcome (O) of interest was generated by relevant mechanism(s) (M) being triggered in specific context(s) (C). The implication for a realist review is that realist program theory takes this structure. It comprises sets of context-mechanism-outcome configurations (CMOCs). Each CMO can be 'read as a sentence' (i.e. 'In this context, that mechanism generates this outcome'.)

Outcomes can lie at different levels of systems, at different points along an implementation chain, and at different points in time (that is, they can be sequenced, with earlier outcomes necessary before later outcomes can be achieved). Wherever an outcome lies, it will be the result of one or more mechanisms operating within particular contexts.

Analytic tasks

The basic analytic task in a realist review is to **find and align the evidence** to demonstrate that particular mechanisms generate particular outcomes and to demonstrate which aspects of context matter. Working from the basic analytic structure described above, it follows that relevant mechanisms cannot be identified without reference to outcomes (mechanisms are what cause outcomes) and that relevant aspects of context cannot be identified without reference to mechanisms. An 'ideal' realist synthesis provides evidence for outcomes, evidence to support the existence of the hypothesised mechanisms, evidence that those mechanisms cause those outcomes, evidence that features of context exist and evidence that those features of context affect whether and which mechanisms fire.

It is because these different kinds of evidence are likely to be found in different kinds of sources that realist review casts its evidentiary net as widely as it does.

The logic of comparison in realist synthesis

In realist evaluation and realist review, the basic logic of comparison is "intra-program, inter-context comparison, on the basis of the program theory". In realist evaluation, the term 'intra-program' refers to the program being evaluated. In realist review, it refers to making comparisons within the pool of literature that has been included as relevant to the intervention(s) under study.

'Inter-context' means making comparing across contexts – be that population groups, cultures, geographic locations or organisational settings. The features of context that might be important are initially hypothesised to be important on the basis of program theory. Comparisons are then made on the basis of those features of context. Are the outcomes indeed different in these different sub-groups? Box 5 offers and example of these types of comparisons.

Box 5: An example of intra-program, inter-context comparisons based on program theory

Intra-program, inter-context comparisons based on program theory: An example

Perhaps a program aims to increase employment outcomes for unemployed people. It is hypothesised to work by building bridging capital between unemployed people and employers. The relevant social theory is social capital theory; the intended mechanism is bridging capital. It might therefore be hypothesised that the program is most likely to be effective in urban population centres, where there are large numbers of employers and unemployed people who are not already linked. In country towns and villages, already strong existing social networks may mean that there is little chance for the program to build social capital and therefore little change in employment outcomes should be expected. In really remote regions where there are few opportunities to strengthen linkages, no change in employment outcomes would be expected. Based on this program theory, comparisons would be made across urban, rural and remote settings.

Evidence for increased bridging capital might include strengthened networks between unemployed people and employers, perhaps collected using social network analysis. Evidence for outcomes would relate to employment of program participants. Evidence that bridging capital operates as a mechanism in employment programs might be found in research about the ways in which people find jobs. Evidence that the bridging capital mechanism fired in this program and accounted for employment outcomes might include the proportion of newly-employed participants whose employment was with employers 'reached' through network-building activities. It might also come from interviews with employers, program participants and/or program staff. Both changes in the strength of networks and employment outcomes would be compared across urban, rural and remote settings to test the program theory about contextual differences affecting the bridging capital mechanism and the outcomes achieved.

The realist review might therefore ask: How much of this evidence is available, and in what sorts of research or evaluation might it be found? Is it possible to disaggregate program outcomes for urban, rural and remote settings? Are there case studies that examine in depth how the program worked in different settings?

The relationship of evidence to theory

Realists do not believe that there is such a thing as final truth, knowledge or 'Truth with a capital T'. All knowledge is partial; all theories remain theories that can be refined or disproved as new evidence comes to light.

In a realist review, the task is to align the 'nuggets' of evidence drawn from different sources against the element of program theory to which they refer, and then to ask: What does this evidence suggest about this aspect of our theory? Does it support it? Does it disprove it? Does it suggest an amendment to it?

Amendments to the theory will take the form of a new CMO configuration. Perhaps one or more of the hypothesised CMOs are removed altogether. Perhaps it has been demonstrated that a particular mechanism requires an alignment of two or more features of context. Perhaps the target group(s) for which an intervention is effective have been refined.

This stage of the work provides both summary and analysis, but it does not yet provide a full *synthesis* of the findings. The final step of the process involves one further level of abstraction – making sense of the pattern of findings. This is most commonly done using existing formal theory in the field in which the analysis is undertaken (e.g. some form of learning theory in education, sociological theory, economic theory, political science, organisational theory...). For an example, see Pawson's use of Merton's theory of reference group behaviour (in Chapter 7) to explain the range of outcomes in interventions that aim to 'name' and shame' (1).

A full realist analysis addresses both these levels and attempts to make sense of the relationship between these two levels. Syntheses that address only one level may also be considered realist syntheses, assuming that they apply and demonstrate application of a realist philosophy of science. The level(s) of analysis chosen will depend on the review's focus. The theories used may have been developed and/or refined from the data and/or be refinement of existing substantive theory.

Conceptual tools

Realist review cannot rely solely on quantitative analytic techniques to achieve its ends (although it can use them where the available data allows). Its purpose is explanatory and it needs to be able to explain the conflicting patterns of outcomes that are almost invariably found in reviews of social programs. Pawson's book 'Evidence Based Policy' (1) suggested what some of these tools might be:

- **juxtaposing** ("for instance, when one study provides the process data to make sense of the outcome pattern noted in another")
- **reconciling** (identifying differences which explain apparently contradictory sets of findings)
- **adjudicating** between studies (based on the quality of research);
- **consolidating** (building 'multi-faceted explanations of success')
- **situating** ('this mechanism in context A, that one in context B")

The key analytic process in realist review involves iterative testing and refinement of theoretically based explanations, using empirical findings in data sources. Reviewers may draw on any appropriate analytic techniques to undertake this testing. When reporting a review, explanation and justification for the choice of techniques should be provided.

7.2 Quality standards for understanding and applying the underpinning principles of realist reviews

Table 5 below sets out our recommended quality standards for understanding and applying the principles of realist review overall.

Table 5: Quality standards for understanding and applying the underpinning principles of realist reviews.

	Inadequate	Adequate	Good	Excellent
The review demonstrates understanding and application of realist philosophy and realist logic which underpins a realist analysis.	Significant misunderstandings of realist philosophy and/or logic of analysis are evident. Common examples include: • program/intervention activities or strategies are confused with mechanisms • no attempts are made to uncover mechanisms • outcomes are assumed to be caused by the program/intervention • relationship(s) between an outcome, its causal mechanism(s) and context(s) are not explained • some theory is provided but this is not explicitly linked to outcome(s)	Some misunderstandings of realist philosophy and/or logic of analysis exist, but the overall approach is consistent enough that a recognisably realist analysis results from the process.	The review's assumptions and analytic approach are consistent with a realist philosophy at all stages of the review. Where necessary a realist program theory is developed and tested.	Good plus: Review methods, strategies or innovations used to address problems or difficulties within the review are consistent with a realist philosophy of science.

7.3 Examples from the literature

Case study 1

One of the most difficult challenges in conducting a realist review is to be able to make a clear distinction between what is an intervention and a mechanism. Through the realist lens, interventions do not cause outcomes, mechanisms do. An example of this confusion may be found in the review by Daykin et al.:

> *"Mechanisms for effective PPI [patient and public involvement] are likely to include adequate provision of resources and information systems to monitor the impact of PALS [Patient Advisory and Liaison Services]. In addition, they may include appropriate models of PPI as well as support for deliberative processes and methods of addressing discursive strategies of engagement with service users' agendas."* (15)

Here, "adequate provision of resources" and "information systems" are (wrongly in our view) described as mechanisms. In realist logic, the provision of something is 'something an intervention does'. Provision of a resources does not cause change, but the reasoning in response to the resource provided does – the reasoning in response to a resource being one of the commonly used definitions of a program mechanism (3) (and see section 2.3 A Glossary of Terms).

Case study 2

Jackson et al.'s realist review on Moving To Opportunity illustrates how interventions, programs and strategies might be found within documents in a realist review (19). Our post-hoc analysis of this review (for educational purposes) can be found in Box 6. The square brackets indicate where we have inserted what we have identified as the mechanism in operation for the outcome of interest.

Box 6: A post-hoc analysis of Jackson et al.'s MTO review to illustrate realist concepts

A post-hoc analysis of Jackson et al.'s MTO review to illustrate realist concepts

Moving to Opportunity (MTO) was a program where the *"... central goal ... was to move families living in public housing in high-poverty neighborhoods to public or private housing in lower-poverty neighborhoods in order to provide 'better opportunities' (e.g., employment, education and housing) for families."* (p962)

In this program the main program strategy was to relocate families. However the program's designers also felt that additional strategies were need to enable relocation to occur, for example:

"In each city partnerships were formed with local public housing authorities which administered the rental assistance, and one or more local, non-profit counseling organizations which provided counseling on how to find rental units and work with landlords when appropriate." (p962)

MTO set out to change multiple outcomes for impoverished families. In Jackson et al.'s review, they focussed on *"... improvement in mental health for adult women, children and adolescent girls"* and provide *"... some insights about the mechanisms and contexts through which the intervention appears to have impacted mental health."* (p963)

The way in which the MTO program strategy changed context on one level is clear to see - it moved families from one context to another, a high-poverty neighbourhood to one with lower-poverty. However, it was not the move (the program strategy) that necessarily caused the mental health of families to improve. As Jackson et al. explain:

"... moving from high-poverty neighborhoods 'worked' for many adults, adolescent girls, and children in terms of mental health outcomes. Statistically significant improvements in mental health for adults (mainly women), female youth, and girls appear to have been related to moving to a better physical and social environment and especially reduced levels of violence in new neighborhoods. Indeed, moving appeared to create an immediate 'resolution' to the stress [mechanism] *that many adult women, female youth, and children were experiencing because of the violence within their previous neighborhoods. Children and female youth who moved appear to have been less afraid to leave their homes and/or spend time outside* [mechanism – reduced fear of violence]*, and were thus able to participate in a broader social life including after- school activities."* (p967)

Conversely for adolescent males, their mental health was not on the whole changed for the better:

"Moving to a new neighborhood was reportedly not as positive for many adolescent males as for females. The lack of significant positive changes in mental health for adolescent males in the MTO program may have been related to the fact that the fear of violence [mechanism] *was less significant for this sub-population, and not enough of a 'push' factor to warrant leaving valued relationships to people or place. At least some male youth may have moved involuntarily, resulting in no improvement in mental health in some evaluations, and decreased mental health in others. These outcomes may be related to the lack of social integration* [mechanism] *into new neighborhoods. Many adolescent males appear to have kept ties with their old neighborhoods. At least two factors are potentially at play: a desire to maintain existing relationships and contacts* [mechanism] *with familiar places in old neighborhoods and a response to feelings and experiences of discrimination* [mechanism]*, including racial discrimination, in new neighborhoods."* (p967)

As might be expected for such a complex intervention as MTO, other 'competing' mechanisms were at play and the mere act of moving (changing one context) resulted in the triggering of desirable mechanisms for some, and other less desirable mechanisms with unanticipated outcomes for others.

Case study 3

A common problem occurs when reviews stop at the level of description of the data or thematic analysis, rather than moving on to offer causal explanation. This does not mean that the findings of such reviews are not valuable, but it does imply that they cannot call upon the realist warrant of their findings being transferable because the same mechanism(s) may be in operation.

McLean et al.'s review "... sought to identify which TBI [traumatic brain injury] behavioral interventions may be most successful within the context of a nursing home." (25). Their realist review provided detailed summaries of the data on two themes - 'caregiving context of a nursing home' (with six subthemes) and 'behavioural interventions for individuals with TBI' (with five subthemes). From their data analysis and synthesis, a matrix was produced that matched the "fit" between 'Nursing home contextual factors' and 'TBI behavioral interventions' (see Table 1 on page 20).

The final model presented by the authors was their 'Practice model of behavioral caregiving' (see figure 1 on page 21). They summarise their model as follows:

> *"There are 3 key factors relating to context, social interaction, collaboration, and everyday activities and routines, and 3 key components of behavioral intervention, antecedent strategies, meaningful activity, and pharmacotherapy. These 6 elements are not isolated components of behavioral caregiving. They are connected to each other and to the 4 stages of the care planning process and the central players, the care aide and resident. This analysis led to a Model of Behavioral Caregiving for TBI (see Figure 1)."*

This analysis is not realist (in the sense used in these training materials) for two reasons. Firstly, whilst context is reported, it is not clear how and which outcome(s) they affect. Secondly, the analysis does not identify and test and mechanisms.

Case study 4

When analysing data using a realist lens, it can be a challenge identifying whether data should be classified and conceptualised as context, mechanism or outcome. Jagosh et al.'s review, 'Uncovering the Benefits of Participatory Research: Implications of a Realist Review for Health Research and Practice', demonstrates the value of clearly naming each element and the role it plays in the explanatory process (6).

Their findings indicated that the outcomes of using a participatory research approach could be explained by a theory of partnership synergy. Synergy was defined as "...combining the perspectives, resources, and skills of a group of people to "create something new and valuable together—a whole that is greater than the sum of its individual parts."

The authors provide numerous examples in their article. Here is one example, which also clearly labels context, mechanisms and outcome in their explanation of the data sample they provided:

> *"In Messengers for Health, recruitment to the advisory board was accelerated by the good reputation and connectedness of an initial community partner:*
>
> > *"The initial partnership in Messengers for Health proved critical in gaining the trust of extended community partners because A.K.H.G.M. [an initial community partner . . . and a parent who lost a child to cancer] is a member of the tribe, is fluent in her language, and is a well-respected individual in the community. At an interview training session one year into funding, community women stated that they were interested in the project because this person was involved. (Reference x1)"*
>
> *Although community members had reason to mistrust outside researchers (context), they felt willing to participate because they trusted (mechanism) the judgment of a well-respected and long-standing community member who was already involved. Trust, respect, and consequent synergy were established from this initial partnership, propelling subsequent stages of program planning (outcome)."*

Jagosh et al.'s review also illustrates another interesting methodological point. They demonstrate that the outcome of one CMO configuration can become the context for another outcome.

> *"Our realist analysis provided evidence that synergy has the potential to build over time when the partnership's activities repeatedly produce successful outcomes. This evidence was synthesized by identifying the outcome of one CMO configuration as forming part of the context in the next phase of research along a chain of planning and implementation stages—what we call a "$C^1M^1O^1$-C^2" pattern, in which outcome^1 becomes a contributor to context^2. This demonstrates how partnerships alter elements of context over time, leading to enhanced outputs and outcomes. ...*
>
> *Barriers to conducting a randomized community trial included community resistance and the demands placed on them given the complex and structured research protocol (context^1). A decision was made at the outset to hire only African-Americans familiar with the community as project staff (context^1). Because of their prior history in the community, the project staff were glad to assist community members beyond the scope of the study (mechanism^1). This led to the staff's greater investment in the project (outcome^1 → context^2). The staff's deepening investment increased the community members' trust in the project (mechanism^2), resulting in closer interactions between the project staff and the community members (outcome^2 → context^3). Because of the greater sense of trust and safety (mechanism^3) due to the previously described trust-building processes, some participants revealed their desire to enroll in the project even though their children had not participated in the school-based asthma program (outcome^3). This led to new methods of recruitment being developed (outcome^3 → context^4), and new recruitment methods led to higher than expected enrollment (context^4). This added to the project stakeholders' desire to overcome attrition obstacles (mechanism^4). As a result, a new capacity to retain*

participants and prevent attrition in a complex clinical trial was created in a mobile population by addressing problems as they arose and through the project stakeholders' increasing sense of motivation, trust, and co-ownership of the project (outcome[4])." (p329-330)

7.4 Learning activities

These activities are designed to provide practice in identifying realist concepts and in realist analysis.

Learning activity 1

Choose any published realist review. Check that you understand the question that it aims to answer. Then turn to the summary of findings section.

To what extent is it possible to identify what has been classified as outcomes, what as context, and what as mechanism? If the authors have not labelled the elements of their findings in this way, are you able to do so from the text?

Now go back to the evidence within the review. Is evidence clearly aligned against a program theory? Is the theory clearly refined in the light of the evidence?

Imagine that you had to provide feedback to the reviewers. What recommendations about improving the 'realist nature' of their review might you provide?

Learning activity 2

This activity is designed to be undertaken by two or more members of a review team. (If you are not undertaking a realist review currently, you can still do this task by choosing a topic of interest to you and working with a colleague).

Select any three articles that will be included in your review. Independently from your colleague, read each of them, highlighting outcomes, mechanisms (or clues about potential mechanisms) and features of context that appear to affect outcomes.

Write or sketch a summary of your analysis based on these three articles, in realist terms.

Now meet with your colleague and compare findings. How are your interpretations of the data similar? How are they different? How would you go about seeking additional data to resolve any differences in interpretation that you might have?

7.5 Reflection activity

To assist you in applying realist principles in analysis, we have developed a list of questions a reviewer / review team might like to ask themselves. These questions are based on the quality standards in Table 5 and are listed in Box 6. We suggest that a reviewer might like to go through the questions in Box 6 to work out if the questions are relevant to their review and then how each question might be addressed.

Box 6: Questions to assist understanding and applying the underpinning principles of realist reviews

- Does the review team understanding underpinning principles of realist reviews?
- Does the review team know how to apply the underpinning principles of realist reviews in their analysis of their data?
    - If 'no' to either question, what steps are you taking to ensure you have sufficient methodological expertise? For example:
        - Recruiting realist review expertise
        - Organising training
        - Organising ongoing methodological support
- What opportunities have been built into the review process to enable the review team to discuss, analyse and/or synthesise the data together?

Section 8. Further reading and resources

VARITIES OF REALIST RESEARCH

Maxwell J. A Realist Approach for Qualitative Research. London: Sage; 2012.

McKelvey B. Toward a Campbellian Realist Organization Science. In: Baum J, McKelvey B, editors. Variations in Organization Science: In Honor of Donald T. Campbell.Thousand Oaks: Sage; 1999. p. 383-411.

Olsen W. Realist Methodology. London: Sage; 2010.

Sayer A. Method in Social Science: A Realist Approach. 2nd ed. Abingdon, Oxon: Routledge; 1992.

MECHANISMS

Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. American Journal of Evaluation 2010;31(3):363-81.

Gerring J. Causal mechanisms: Yes, But...
http://people.bu. edu/jgerring/documents/CausalMechanisms_Extended pdf 2013
[accessed 2013 Jul 7]

Hedström P, Ylikoski P. Causal Mechanisms in the Social Sciences. Annual Review of Sociology 2010;36:49-67.

THEORY

Friedman V. Designed Blindness: An Action Science Perspective on Program Theory Evaluation. American Journal of Evaluation 2001;22(2):161-81.

Pawson R. Middle Range Theory and Program Theory Evaluation: From Provenance to Practice. In: Vaessen J, Leeuw F, editors. Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences.New Brunswick, New Jersey: Transaction Publishers; 2010. p. 171-202.

Rogers P, Petrosino A, Huebner T, Hacsi T. Program theory evaluation: Practice, promise, and problems. New Directions for Evaluation 2004;2000(87):5-13.

Weiss C. Which links in which theories shall we evaluate? New Directions for Evaluation 2013;2000(87):35-45.

REALIST REVIEW

The main text for realist synthesis is **Pawson, R: Evidence-Based Policy: A Realist Perspective (Sage, 2006)**

For underlying principles and assumptions: see Chapter 2
For Pawson's critique of existing review methods: see Chapter 3

For the stages in conducting a review: see Chapter 4

For an overview: 'Realist Synthesis: an introduction' – available on line and free at:
http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf

For the origins and influences on Pawson's interpretation of realism, see Pawson, R. (2013) The Science of Evaluation: A Realist Manifesto (Part 1). Part 2 of this book provides Pawson's explanation of complexity from a realist perspective.

Rycroft-Malone R, McCormack B, Hutchinson A, DeCorby K, Bucknall T, Kent B, et al. Realist synthesis: illustrating the method for implementation research. Implementation Science 2012;7:33.

WEBSITES

The RAMESES Project
http://ramesesproject.org/

The realist hive
http://blogs.exeter.ac.uk/realisthive/

Theory-driven inquiry for health systems research
http://www.itg.be/internet/ds/tde/index.html

Realist synthesis: The website
http://www.leeds.ac.uk/sociology/realistsynthesis/

Section 9. References

(1) Pawson R. Evidence-based policy: A realist perspective. London: Sage; 2006.

(2) Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. American Journal of Evaluation 2010;31(3):363-81.

(3) Pawson R, Tilley N. Realistic evaluation. London: Sage; 1997.

(4) Merton R. On Theoretical Sociology. Five Essays, Old and New. New York: The Free Press; 1967.

(5) Pawson R. The Science of Evaluation: A Realist Manifesto. London: Sage; 2013.

(6) Jagosh J, Macaulay AC, Salsberg J, Bush PL, Henderson J, Sirett E, et al. Uncovering the Benefits of Participatory Research: Implications of a Realist Review for Health Research and Practice. Milbank Quarterly 2012;90(2):311-46.

(7) Leedy P, Ormrod J. Practical research: Planning and design. 7 ed. Thousand Oaks: Sage; 2001.

(8) Wikigren M. Critical realism as a philosophy and social theory in information science? Journal of Documentation 2005;61(1):11-22.

(9) Archer M. Realist Social Theory. Cambridge: Cambridge University Press; 1995.

(10) Klee R. Introduction to the philosophy of science. Cutting nature at its seams. New York: Oxford University Press; 1997.

(11) Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. BMC Med Educ 2010;10:12.

(12) O'Campo P, Kirst M, Tsamis C, Chambers C, Ahmad F. Implementing successful intimate partner violence screening programs in health care settings: Evidence generated from a realist-informed systematic review. Social Science & Medicine 2011;72:855-66.

(13) Jagosh J, Pluye P, Macaulay A, Salsberg J, Henderson J, Sirett E, et al. Assessing the outcomes of participatory research: protocol for identifying, selecting, appraising and synthesizing the literature for realist review. Implementation Science 2011;6(24).

(14) Funnell S, Rogers P. Purposeful Program Theory: Effective use of theories of change and logic models. San Francisco, CA: Jossey-Bass; 2011.

(15) Daykin N, Evans D, Petsoulas C, Sayers A. Evaluating the impact of patient and public involvement initiatives on UK health services: a systematic review. Evidence & Policy 2007;3(1):47-65.

(16) McMahon T, Ward P. HIV among immigrants living in high-income countries: a realist review of evidence to guide targeted approaches to behavioural HIV prevention. Systematic Reviews 2012;1:56.

(17) Connelly J, Duaso M, Butler G. A systematic review of controlled trials of interventions to prevent childhood obesity and overweight: A realistic synthesis of the evidence. Public Health 2007;121:510-7.

(18) Morgan G. Evidence-based health policy: A preliminary systematic review. Health Educ J 2010;69(1):43-7.

(19) Jackson L, Langille L, Lyons R, Hughes J, Martin D, Winstanley V. Does moving from a high-poverty to lower-poverty neighborhood improve mental health? A realist review of 'Moving to Opportunity'. Health & Place 2009;15:961-70.

(20) Banning smoking in vehicles with children: A realist review. University of East Anglia, UK: SAPC; 2010.

(21) Pawson R, Wong G, Owen L. Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy. American Journal of Evaluation 2011;32(4):518-46.

(22) Carr S, Lhussier M, Forster N, Geddes L, Deane K, Pennington M, et al. An evidence synthesis of qualitative and quantitative research on component intervention techniques, effectiveness, cost-effectiveness, equity and acceptability of different versions of health-related lifestyle advisor role in improving health. Health Technol Assess 2011;15(9):1-284.

(23) Kane S, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M. A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. BMC Health Services Research 2010;10(286).

(24) Bhaskar R. A Realist Theory of Science. London: Verso; 1978.

(25) McLean AM, Koppang J. Behavioral caregiving for adults with traumatic brain injury living in nursing homes: developing a practice model. J Theory Construct Test 2010;14(1):17-22.

Appendix 10 Training materials for meta-narrative reviews

Training materials for meta-narrative reviews

Prof Trisha Greenhalgh
Dr Geoff Wong

Global Health Innovation and Policy Unit
Centre for Primary Care and Public Health
Blizard Institute
Queen Mary, University of London
Yvonne Carter Building
58 Turner Street
London E1 2AB

Emails:
p.greenhalgh@qmul.ac.uk
grckwong@gmail.com

Acknowledgements

This project was funded by the National Institute for Health Research Health Services and Delivery Research Programme (NIHR HS&DR) - project number 10/1008/07.
The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR program, NIHR, NHS or the Department of Health.

Version 1
Last updated: 7 July 2013

1. Introduction

This document has been developed to provide practical methodological advice to reviewers who want to undertake a meta-narrative review (or synthesis – the terms are synonymous). We wrote this document for several reasons. As researchers in the field, we have noted rising demand for training in meta-narrative reviews, but as yet no 'how to' methodological manuals exist. When we and our colleagues have provided training in meta-narrative reviews, recurrent questions and training needs arise. We have been funded to develop training materials for meta-narrative reviews as part of the RAMESES project (http://www.biomedcentral.com/1471-2288/11/115). Finally, whilst developing the RAMESES publication standards for meta-narrative reviews (http://www.biomedcentral.com/1741-7015/11/21) and running the RAMESES JISCMail (http://www.jiscmail.ac.uk/RAMESES), our understanding of the training needs of our fellow meta-narrative review researchers has grown.

2. A brief overview of meta-narrative reviews

Meta-narrative review is a relatively new method of systematic review, designed for topics which have been *differently conceptualised and studied* by different groups of researchers. Here's an example. Many groups have studied the building of dams in India. Some have conceptualised this dam-building as engineering; others as colonialism; others as a threat (or promise) to the local eco-system; others as inspiration for literature and drama, and so on. If we were to summarise this topic area in a way that was faithful to what each different group set out to do, we would have to start by asking how each of them approached the topic, what aspect of 'dams in India' they chose to study and how. In order to understand the many approaches, we would have to consciously and reflexively *step out of* our own world-view, learn some new vocabulary and methods, and try to view the topic of 'dams in India' through multiple different sets of eyes. When we had begun to understand the different perspectives, we could summarise them in an over-arching narrative, highlighting what the different research teams might learn from one another's approaches.

Some reviewers might be interested only in summarising the findings of randomised controlled trials of 'dam present' versus 'dam absent' on a predefined outcome, and if that was the focus of the review, a Cochrane review with statistical meta-analysis would be the gold standard approach. The meta-narrative approach is intended for those reviews where the underlying research goal is to identify and explore the *diversity* of research approaches to a topic.

The methodology of meta-narrative review was developed by Trish Greenhalgh and her team in 2004 when reviewing the literature on diffusion of service-level innovations in healthcare (1). A methods paper was published in Social Science and Medicine in early 2005 (2). The inspiration for this method was Kuhn's 1962 book *The Structure of Scientific Revolutions*, which argued that science progresses in paradigms (i.e. particular ways of viewing the world, including assumptions about how the world works) and that one scientific paradigm gives way to another as

scientific progress renders yesterday's assumptions and practices obsolete.[REF] Newton's theories and methods, for example, became less and less able to answer the emerging questions of particle physics, leading Einstein to develop his theory of relativity. Meta-narrative review looks historically at how particular research or epistemic traditions have unfolded over time and shaped the kind of questions being asked and the methods used to answer them. A research tradition is a series of linked studies, each building on what has gone before and taking place within a coherent paradigm (that is, within a shared set of assumptions and preferred methodological approach shared by a group of scientists).

Further reading
Researchers who are interested in finding out more about the meta-narrative review method should consult, 'Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review' by Greenhalgh et al (2) and the RAMESES publication standards for meta-narrative reviews (3).

3. Training materials for meta-narrative reviews

In this section, we will focus on the specific areas in undertaking a meta-narrative review which we have noted have been the source of frequently encountered difficulties and misconceptions for meta-narrative reviewers. These are not the only ones that meta-narrative reviewers will find challenging, but we have identified these topics as particularly troubling through our past experiences in practice and training, the RAMESES JISCMail list, the literature and in preparing the RAMESES Publication standards for meta-narrative reviews.

We appreciate that the needs of each meta-narrative reviewer, from novice to relative expert, will be different. We felt that the greatest developmental need was in setting out what the main principles were for each of the challenging areas, oriented more towards the less experienced reviewer. We have done this by providing 'Quality standards' for each area we covered. We have used examples of published reviews from the literature to show how these standards have or have not been met – with a focus on illustrating the importance of ensuring the principles in the quality standards are met. Whilst learning needs differ, quality standards apply to meta-narrative reviewers of all levels. For each topic area, we have provided a series of questions to help novice reviewers to reflect on (and hopefully learn) how they might meet each of the quality standards set out for each topic. For the more experience reviewer, we hope that the questions will still be of some use as an aide memoire or perhaps for use as training materials for fellow review team members?

Topics covered in this document include:

- Understanding and applying the underpinning principles of meta-narrative reviews
- Focussing reviews
- Finding the most relevant evidence

With each of the above topic areas, we will provide:

- Objectives
- An explanation on why the topic area is important to get right
- What would constitute high 'quality' for this topic area
- A worked example (drawn from the published literature) of how the topic area in a review might be improved.
- Example(s) from the published literature of how the topic area has been tackled successfully.
- Reflection activities

How a meta-narrative review is undertaken will vary greatly depending, for example, on the research question, resource available, funder's expectations, end users' needs and so on. As such it is impossible to be prescriptive and restrictive on what must be done. Our training materials should be thought of more as guidance than 'must-dos'. This is an important difference from Cochrane reviews, which tend to be undertaken according to very strict and standardised protocols.

Additional detail on the quality standards on each topic area can be found online at: http://www.ramesesproject.org/index.php?pr=Project_outputs#method

We draw our examples from published meta-narrative reviews, some of which are cited to illustrate our claim that the review did *not* meet the quality standard we propose. We appreciate that the authors of such examples may feel that we are being unfairly critical of their work. We wish to stress that meta-narrative review is an evolving field of secondary research and that since quality standards were not available when those reviews were undertaken, it is hardly surprising that different authors used different approaches. However, the methodology of meta-narrative review is now maturing and it is important to point out that not all early examples followed what were subsequently established by the RAMESES project as the key standards.

4. Understanding and applying the underpinning principles of meta-narrative reviews

4.1 Objectives

For this topic, we hope that when you have finished reading about it you will:

- Understand what the underpinning principles are of meta-narrative reviews
- Have read about examples of how meta-narrative reviews have been developed
- Know what constitutes good practice when developing meta-narratives
- Be aware of the steps you may need to take to ensure you apply the underpinning principles of meta-narrative reviews

4.2 The need to understand and apply the underpinning principles of meta-narrative reviews

Meta-narrative review (which is rooted in a constructivist philosophy of science) was inspired by the work of Thomas Kuhn, who observed that science progresses in paradigms. Meta-narrative reviews often look historically at how particular research traditions or epistemic traditions have unfolded over time and shaped the 'normal science' of a topic area.

The review seeks first to identify and understand as many as possible of the potentially important different research traditions that have a bearing on the topic. In the synthesis phase, by means of an over-arching narrative, the findings from these different traditions are compared and contrasted to build a rich picture of the topic area from multiple perspectives. The goal of meta-narrative review is sense-making of a complex (and perhaps contested) topic area. During analysis and synthesis, six guiding principles (pragmatism, pluralism, historicity, contestation, reflexivity and peer review) should be used and these are described in more detail below:

- Principle of pragmatism: what to include is not self-evident. The reviewer must be guided by what will be most useful to the intended audience(s), for example, what is likely to promote sense making.
- Principle of pluralism: the topic should be illuminated from multiple angles and perspectives, using the established quality criteria appropriate to each. For example, reviewers should avoid beginning with a single 'preferred' perspective or methodological hierarchy and proceed to judge work in other traditions using these external benchmarks. Research that lacks rigor must be rejected, but the grounds for rejection should be intrinsic to the relevant tradition, not imposed on it.
- Principle of historicity: research traditions are often best described as they unfolded over time, highlighting significant individual scientists, events and discoveries which shaped the tradition.
- Principle of contestation: 'conflicting data' from different research traditions should be examined to generate higher-order insights (for example, about how different research teams framed the issue differently or made different assumptions about the nature of reality).
- Principle of reflexivity: throughout the review, reviewers must continually reflect, individually and as a team, on the emerging findings.
- Principle of peer review: emerging findings should be presented to an external audience and their feedback used to guide further reflection and analysis.

4.3 Quality standards for understanding and applying the underpinning principles of meta-narrative reviews

For this topic area, we would expect quality to be defined as set out in Table 1.

Table 1: Quality standards for understanding and applying the underpinning principles of meta-narrative reviews

	Inadequate	Adequate	Good	Excellent
The review demonstrates understanding and application of the purpose and principles underpinning a meta-narrative review.	Significant misunderstandings of purpose and principles underpinning a meta-narrative review. Common examples include: • Analysing only one paradigm / epistemic tradition • No application of the six underlying principles	Some misunderstandings of purpose and principles underpinning a meta-narrative review, but the overall approach is consistent enough that a recognisable set of distinct meta-narratives together with a higher-order synthesis of the findings from this process.	The review's assumptions and analytic approach are consistent with the purpose and underpinning principles of a meta-narrative review. In particular, the philosophical position is explicitly constructivist. A sufficient range of paradigms/epistemic traditions has been included to make sense of an unfolding and complex topic area from multiple perspectives and to use contrasts between these as higher-order data.	Good plus: Review methods, strategies or innovations used to address problems or difficulties within the review are philosophically coherent and make a clear and illuminative contribution to the knowledge base on the topic area.

4.4 Examples from the literature

Pragmatism

When applying the principle of pragmatism the reviewer must be guided by what will be most useful to the intended audience(s), for example, what is likely to promote sense making. This principle applies through out a meta-narrative review, from the focusing through to scoping the literature and then to analysis and, if needed, driving the need for further searching. As a 'rule of thumb' the goal is to make sense of the data and any leads or 'trails' that emerge during a meta-review's processes should be pursued.

An example of pragmatism (at the focusing and scoping stages of a meta-narrative review) can be seen in Collins et al.'s review (see section 5.4 as well) (4). Their review had the, *"... objective ... to monitor thematic trends in this knowledge base over time, and to track scholarly prescriptions for municipal government intervention on local health inequities."* Initially the reviewers were uncertain as to which bodies of evidence would need to be included in their review. Through scoping of the literature, they decided that four bodies of evidence would most likely contain the data they needed and so decided that it made sense to focus on these four bodies of literature.

Reflexivity and peer review

During their review, review teams need to continually reflect, individually and as a team, on the emerging findings. Addis et al.'s undertook a review, *"to provide baseline knowledge of the health, social care and housing needs of older LGBT people that could be used to inform policy and define research questions."*
They acknowledge the need for reflexivity through the use of the principle of peer review, as there are:
"… dangers of reviewers 'flying solo' in the literature that is poorly organised and presented and is not amenable to appraisal using standard tools. We sought to use additional measures to help protect against bias and the high level of agreement between researchers may appear to indicate that our conclusions were sound. However, high rates of agreement might simply indicate that we brought similar biases to understanding the relevance of the material and drawing conclusions from it. We therefore engaged the wider research team and policy leads in a process of testing the findings against their expectations and experience."(5)

Peer review is also used by other reviewers. Peer review is the requirement to present emerging findings to an external audience and their feedback used to guide further reflection and analysis. Along with Addis et al. above, Kitson et al. invited researchers from other research traditions and a patient group to, *"… share experiences …"* and, *"… to plan further work."*(6)

Pluralism

A key principle in meta-narrative reviews is to develop an account of the topic area that is illuminated from multiple angles and perspectives. A meta-narrative review must analyse more than one paradigm and produce a recognisable set of distinct meta-narratives together with a higher-order synthesis of these results. Recognised problems with some published meta-narrative reviews are:

- they analyse sources from only one paradigm, as is the case in Kitson et al.'s review, where despite many other features of good practice in meta-narrative review, only a nursing perspective is taken (6)
- the analysis and synthesis lacks a meta-narrative dimension, as can be seen in Addis et al.'s review, where the results are reported as a thematic narrative summary but not teased out into separate research (or epistemic) traditions which are then compared and contrasted (5).

Collins et al. in their meta-narrative review scoped the literature and judged that to make sense of the literature at least four perspectives needed to be examined in detail:

"…[a] substantial proportion of the health inequities knowledge base present lifestyle- and healthcare- (referred to in this article as 'behavioural' and 'biomedical', respectively) oriented perspectives regarding solutions to health inequities. Meanwhile, the high number of abstracts with social and physical environment SDOH [social determinants of health] profiles likely reflects the fact that the 'local' or 'municipal' level was one of four overarching search themes employed in the search strategy."(4)

Historicity

In a meta-narrative review, research traditions are often best described as they unfolded over time, highlighting significant individual scientists, events and discoveries which shaped the tradition. Collins et al. took this approach and reported that:

"The changes in publication activity in the four bodies of literature are displayed in Figure 3. … Changes in the SDOH [social determinants of health] profile of the article abstracts are displayed in Figure 4, using five-year increments to simplify the analyses. … Taken together, these findings suggest that broader, more critical perspectives on health inequities were prominent in the early stages of development of the knowledge base, but that over time these perspectives gave way to a focus on 'behavioural' and 'biomedical' explanations for, and solutions to, health inequities."(4)

Contestation

During a meta-narrative review, 'conflicting data' from different research traditions should be examined to generate higher-order insights (for example, about how different research teams framed the issue differently or made different assumptions about the nature of reality). In the illustrative text below from the review by Collins et al., they point out that there was a geographical difference in how researchers envisaged the role of municipalities, which has implications on how research from different parts of the world needed to be interpreted differently:

"The seven categories of roles were emphasized to varying extents across the different geographical regions of origin. In abstracts of Canadian, European, and Australian & New Zealand origin, the most commonly prescribed role was to 'join or build on existing local health networks'. Canadian abstracts also emphasized the need for greater 'intra-municipal capacity building' to tackle local health issues. 'Improving the social, economic, and built environments' was the most commonly prescribed role among abstracts of a global/transcontinental origin, and of a Mexican, South & Central American origin, while abstracts of American origin stressed the need for municipalities to 'conduct health impacts assessments, and assess local needs'. The varying emphases placed on potential roles likely speak to the diverse jurisdictional responsibilities of municipal governments across and within countries, as well as the unique and highly specific health and social issues facing municipal governments within these countries. Accordingly, these differences signal the need for researchers to interpret these findings with caution by considering the applicability of these 'roles' within the context of a given municipal government's jurisdictional powers, functions, and public policy priorities."(4)

4.5 Reflection activity for understanding and applying the underpinning principles of meta-narrative reviews

It is essential that before and during a meta-narrative review, review teams ensure that they understand and apply the underlying principles of meta-narrative reviews. Box 1 contains questions that we hope will help a review team to undertake a rigorous meta-narrative review.

Box 1: Questions to assist the focussing process in meta-narrative reviews

- Does the review team understanding underpinning principles of meta-narrative reviews? Do they, for example, accept the Kuhnian notion of paradigm and recognise that the task is to surface, summarise and contrast different paradigms? Are they familiar with the difference between a 'technical' and an 'interpretive' approach to systematic review and with the six principles of pragmatism, pluralism, historicity, contestation, reflexivity and peer review? Could they defend the need for an interpretive approach and for following all six of the key principles?
- Does the review team know how to apply the underpinning principles of realist reviews during their meta-narrative reviews?
    - If 'no' to either question above, what steps are you taking to ensure you have sufficient methodological expertise? For example:
        - Recruiting meta-narrative review expertise
        - Organising training
        - Organising ongoing methodological support
- What opportunities have been built into the review process to enable the review team to:
    - reflect on, analyse and/or synthesise the data together?
    - enable peer review?

5. Focussing reviews

5.1 Objectives

For this topic, we hope that when you have finished reading about it you will:

- Understand the importance of the need to focus a meta-narrative review
- Have read about examples of how reviews have been focussed
- Know what constitutes good practice when focussing reviews
- Be aware of the steps you may need to take to focus your review

5.2 The need to focus reviews

A meta-narrative review asks some or all of the following questions: (1) Which research (or epistemic) traditions have considered this broad topic area?; (2) How has each tradition conceptualized the topic?; (3) What theoretical approaches and methods did they use?; (4) What are the main empirical findings?; and (5) What insights can be drawn by combining and comparing findings from different traditions?'

Because a meta-narrative review may generate a large number of avenues that might be explored and explained, and because resources and timescale are invariably finite, it is almost always necessary to 'contain' a review. Many different aspects (the 'what') of a meta-narrative review might need to be focussed, examples include:

- the question(s) to be answered (refining from broader to narrower)
- scale of review (e.g. focus on particular countries in international development reviews, or cultures, or timeframes),
- the extent to which the review aims to be comprehensive
    - rapid review – building and making sense of meta-narratives within a more limited literature set
    - systematic review – aiming to include all evidence on the topic

Focussing may also take place at different time points in the review process ('when'), and different aspects may be focussed at different times during the course of the review. Examples of time points when focussing may be needed include:

- when negotiating the research project or funding contract;
- while writing and negotiating the research protocol (where required for funding projects)
- when an Advisory Group is established and begins to meet
- when content experts are consulted
- when it becomes clear how much evidence is available for particular aspects of the question;
- when evidence suggests new pathways that could be explored

Focusing needs to be considered from the start and reviewed, perhaps iteratively, as it progresses. It is legitimate (indeed, expected) for the objectives, question and/or the breadth and depth of the review to evolve or be refined as the review progresses.

5.3 Quality standards for focussing reviews

For this topic area, we would expect quality to be defined as set out in Table 2.

Table 2: Quality standards for focussing the review

	Inadequate	Adequate	Good	Excellent
The review question is sufficiently and appropriately focussed.	The review question is too broad to be answerable within the time and resources allocated. There is no evidence that progressive focussing occurred as the review was undertaken (indeed, the authors may inappropriately consider that the research question must be established at the outset and not change further).	Attempts were made by the review team to progressively focus the review topic in a way that takes account of the priorities of the review and the realities of time and resource constraints.	Adequate plus: There is evidence that the focussing process was iterative and reflexive. Commissioners of the review were involved in decision-making about focussing. Decisions made about which avenues were pursued and which left open for further inquiry are clearly documented and made available to users of the review.	Good plus: The review team drew systematically on external stakeholder expertise to drive the focussing process in order to achieve maximal end-user relevance.

5.4 Examples from the literature

Focusing has been tackled in different ways by different researchers.

Kitson et al.'s meta-narrative review on 'Defining the fundamentals of care' which set out, *"... to try to establish what is considered to be the fundamental aspects of patient care and what research evidence there was in the literature that could inform nursing practice."* To help focus their review the review team drew on external stakeholder expertise to drive the focussing process in order to achieve maximal end-user relevance.

"... planning phase for the initiative commenced in 2008 with the inaugural meeting of the Oxford International Learning Collaborative (ILC). The purpose of this group has been on building research capacity in AHSCs [Academic Health Science Centres] around key areas of nursing interventions—called the fundamentals of care. The group has international membership and is diverse in its background and experience although the majority of members are from the nursing profession. We are adding to the diversity of this original group by inviting members of the Cochrane Nursing Care Field (CNCF) to be involved in the process and facilitate a joint seminar with a patient group in Oxford so they can share experiences with these aspects of care (http://www.healthtalkonline.org) and to plan further work." (6)

Collins et al. used a different strategy to help them narrow down the bodies of work that would form the focus of their meta-narrative review (4). Their review had the, *"... objective ... to monitor thematic trends in this knowledge base over time, and to track scholarly prescriptions for municipal government intervention on local health inequities."* To help them understand the literature they needed to focus on in their

review, they undertook and reported in their paper an overview of the current issues in local health inequalities.

"Four bodies of literature on health inequities - 'health promotion' (HP), 'Healthy Cities' (HC), 'population health' (PH), and 'urban health' (UH) - were examined for the meta-narrative mapping analysis. These four literature bodies were chosen because, as discussed earlier, they have made the most significant scholarly contributions to understanding patterns of health inequities, and identifying and describing interventions to reduce health inequities."

5.5 Reflection activity

Box 2: Questions to assist the focussing process in meta-narrative reviews

- Can you complete your review within the time and resources allocated?
- Have you discussed the need to focus your review with (where relevant):
  - your supervisor?
  - within your review team?
  - your funding body / commissioners of the review?
  - potential users of your review?
- What processes will you develop and put in place to focus your review? For example:
  - 'What' will you focus?
  - 'When' will you do your focussing?

Note also that the task addressed in next section (scoping the literature to find the most relevant evidence) is closely linked in practice to the task of focusing (as the review is progressively scoped, it is also progressively focussed), hence these two aspects of the review should be considered together in practice even though we have separated them out analytically in this document.

6. Finding the most relevant evidence

Three specific process will help the meta-narrative reviewer find the most relevant evidence:

- Scoping the literature
- Developing and pursuing a search strategy
- Selecting and appraising the documents

These will be covered in turn below.

6.1 Scoping the literature

6.1.1 Objectives

For this topic, we hope that when you have finished reading about it you will:

- Understand the importance of the need to scope the literature
- Have read about examples of how scoping has been undertaken
- Know what constitutes good practice when scoping the literature
- Be aware of the steps you may need to take to scope the literature

6.1.2 The need for scoping the literature

An important process in a meta-narrative review is to identify a sufficiently diverse range of sources to build as comprehensive a map as possible of research undertaken on the topic. This step identifies in broad terms the different research traditions, situated in different literatures, which have addressed the topic of interest. Initial attempts (which may be iterative) to make sense of a topic area may involve not just informal 'browsing' of the literature but also consulting with experts and stakeholders. As noted above, the scoping process takes place in parallel with, and feeds into, the focussing of the review – though these processes may feel as if they are pulling in different directions ('scoping' tends to reveal numerous new avenues that seem to need exploring whereas 'focussing' tends to be a process of deciding *not* to pursue certain avenues).

6.1.3 Quality standards for scoping the literature

For this topic area, we would expect quality to be defined as set out in Table 3.

Table 3: Quality standards for scoping the literature

	Inadequate	Adequate	Good	Excellent
The scoping of the literature has been sufficiently and appropriately undertaken	The scoping of the literature has been limited and cursory (e.g. only a single source is used – perhaps the Medline database – and/or the review has inappropriately concentrated on a single research tradition – for example 'evidence based medicine')	Attempts made to utilise a broad range of relevant sources and to build as comprehensive a map as possible of the research traditions on the topic.	Adequate plus: A coherent and through search strategy, deliberately including exploratory methods such as browsing and modified in the light of emerging findings, is used to identify research traditions.	Good plus: Systematic use is made of experts and stakeholders in identifying research traditions.

6.1.4 Examples from the literature

A common strategy used to help work out what the scope of the literature is in meta-narrative review is to undertake informal searches, consult experts in the field and/or to track citations from the reference lists of relevant documents.

Collins et al. (see Section 5.4 above) predominately used the literature to help them both focus their review and identify the four bodies of literature on health inequalities which they thought would help them to make more sense of their topic (4). Some review teams included content experts; others combined an exploratory literature search and internal expertise (7). An alternative strategy has been to consult with external subject specific experts and to recruit such individuals into the review team (5). Greenhalgh et al. recruited Kyriakidou only after identifying the need for an organisational psychologist (1).

6.2 Developing and pursuing a search strategy

6.2.1 Objectives

For this topic, we hope that when you have finished reading about it you will:

- Understand the importance of developing a search strategy that meets your review questions' needs and is faithful to the methodology of meta-narrative review.
- Have read about examples of how searches have been developed for meta-narrative reviews
- Know what constitutes good practice when developing searches for meta-narrative reviews
- Be aware of the steps you may need to take to develop and use a search strategy for meta-narrative reviews

6.2.2 The need for search strategies suitable for meta-narrative reviews

There are two phases of searching in meta-narrative reviews. Initially informal, iterative and exploratory searching is undertaken to build a broad overview of the different research traditions, situated in different literatures, which have addressed the topic of interest (as discussed in Section 6.1 Scoping the literature).

After scoping, the research or epistemic traditions identified from the literature need to be mapped and the more formal searching takes place. Searching in a meta-narrative review is guided by the objectives and focus of the review, and revised iteratively in the light of emerging data. Searching is directed at finding sufficient data to develop and make more sense of the relevant research traditions that have been identified from the scoping search phase. The data may lie in a broad range of sources that cross traditional disciplinary, programme and sector boundaries. This stage is likely to involve searching for different kinds of data in different ways.

6.2.3 Quality standards for search strategies

For this topic area, we would expect quality to be defined as set out in Table 4.

Table 4: Quality standards for search strategy

	Inadequate	Adequate	Good	Excellent
The search process is such that it would identify data to enable the review team to develop and refine the map of seminal papers and primary research studies.	The search is incapable of supporting the development of a rigorous meta-narrative review. Errors may include: • The search is driven by a methodological hierarchy of evidence (e.g. privileging RCTs) rather than the need to identify the range of research paradigms (concepts, theories, methods and instruments) that have been brought to bear on a topic • The search process is not informed by the objectives and focus of the review • The database(s) selected are narrow in the subject matter that they contain (e.g. limited to biomedical topics and approaches rather than extending to social science, psychology etc.) • Searching is undertaken once only at the outset of the review and there is no iterative component	Searches are driven by the objectives and focus of the review and are piloted and refined to check that they are fit for purpose. Documents are sought from wide range of sources likely to contain relevant data on research traditions. There is no predefined restriction on the study or documentation type that is searched for	Adequate plus: further searches are undertaken in light of greater understanding of the topic area, particularly through the use of citation-tracking of seminal papers. These searches are designed to find additional data that would allow greater sense to be made of component research traditions and/or draw higher order insights from contrasts between traditions.	Good plus: The search reflects a high degree of scholarly insight into the key research traditions of the review.

6.2.4 Examples from the literature

Many of the published meta-narrative reviews had searches which were driven by the objectives and focus of the review, sought documents from wide range of sources which are likely to contain relevant data on research (or epistemic) traditions

and had no a priori restriction on the study or documentation type that is searched for.

The tension for reviewers was between specifying too many search terms and hence narrowing down too much and too early what might be found and using broad search terms and accepting that sensitivity and specificity would be limited (8). The disadvantage of the lower sensitivity and specificity found using conventional controlled terms or key word searching is that too many documents are found and an inordinate amount of reviewer resources would be needed to sift through them. This issue was clearly identified in Contandriopolous et al.'s meta-narrative review, "… focused on the *collective level of analysis* in order to understand *deliberate interventions* aimed at influencing behaviors or opinions though the *communication of information*." (9)

They realised that the broad focus of their topic area made it, *"… challenging to identify a coherent and precise set of keywords for the search process."* Moreover their knowledge of the field had identified another review in a closely related topic that had relied on a keyword approach. This review, *"… enabled the identification of 169 relevant documents out of 4,250 hits (before triaging on the basis of strength of evidence). We anticipated that in our case, a similar strategy would yield even more chaff and less wheat because the disciplinary traditions targeted are broader and each relies on distinct vocabulary and conceptualizations. …we relied instead on a non-keyword-based reviewing process that we dubbed double-sided systematic snowball.*"

They go on to provide details on how they applied this method of searching:
"Our starting point was to identify, through team consensus, some seminal papers (n = 33) that were considered to have shaped the evolution of the field. We started by identifying a heuristic list of seven 'traditions' … Each tradition was exemplified by one or more publications. …
We then used the ISI Web of Science Citation Index to identify all documents (n = 4,201) that cited those seminal papers. The snowball process here was prospective, since it exclusively targeted documents published after the selected seminal paper. We then triaged the results using the titles and (if present) the abstracts ... This process identified 189 documents that we then retrieved and read for further selection according to the same criteria. At the end of this prospective snowballing, we selected 102 documents for detailed analysis. Next we used the bibliographies of those 102 documents as a basis for retrospective systematic snowball sampling."

The search strategy used in Contandriopolous et al.'s meta-narrative was based on that developed and used by Greenhalgh et al. (1;2). It illustrates that a different way of searching may be needed in meta-narrative reviews that is not only more likely to find relevant data, but also is possibly a more efficient use of time and resources.

6.3 Selecting and appraising the documents

6.3.1 Objectives

For this topic, we hope that when you have finished reading about it you will:

- Understand the how documents are selected and appraised for meta-narrative reviews
- Have read about examples of how selection and appraisal have been undertaken in meta-narrative reviews
- Know what constitutes good practice when selecting and appraising documents for meta-narrative reviews
- Be aware of the steps you may need to take when selecting and appraising documents for use in meta-narrative reviews

6.3.2 The need for selecting and appraising documents

Meta-narrative review is not a technical process (that is, it is not simply a matter of checking and categorising pieces of data against a checklist or set of criteria). Rather, it is an interpretive process of *sense-making of the literature*, selecting and combining data from primary sources to produce an account of how a research tradition unfolded and why, and then (in the synthesis phase) comparing and contrasting findings from these different traditions to build a rich picture of the topic area from multiple perspectives. This process requires a series of judgements about the unfolding of research in particular traditions, and about the relevance and robustness of particular data within that tradition.

Meta-narrative review takes its quality criteria from the traditions included in the review. Studies in these separate traditions should be appraised using the quality criteria that a competent peer-reviewer *within that tradition* would be required to use, as judged by scholars in that field.

6.3.3 Quality standards for selecting and appraising documents

For this topic area, we would expect quality to be defined as set out in Table 5.

Table 5: Quality standards for search strategy

	Inadequate	Adequate	Good	Excellent
The selection and appraisal process ensures that sources relevant to the review containing material likely to help identify, develop and refine understanding of research traditions are included.	The selection and appraisal process does not support a rigorous and complete meta-narrative review. For example: • Selection is overly driven by methodological hierarchies (in particular the restriction of the sources to RCTs to the exclusion of other forms of evidence) • Sources are appraised using a technical checklist focused on methodological procedure rather than by making a defensible judgement on the contribution that a source might make. • Selection and appraisal processes are overly restrictive and exclude materials that may help sense-making of research traditions. • Selection and appraisal processes are not sensitive enough to exclude irrelevant materials	Selection of a document for inclusion in the review is based on what it can contribute to making sense of research traditions. All the key high-quality sources identified are included in the review and the poor-quality ones accurately excluded.	Adequate plus: During the appraisal process studies in the separate traditions are appraised competently using quality criteria that scholars within that tradition would recognise.	Good plus: The judgements made when appraising papers are a model of good scholarship in each of the included traditions.

6.3.4 Examples from the literature

Two separate processes take place in this stage. One is to decide if a document should be included into a meta-narrative review and the other is to appraise the included document using quality criteria acceptable to that tradition.

As mentioned in Table 5 above, inclusion should be based on relevance – how can a document contribute to sense-making? In published reviews, this is operationalised by reviewers by using broad inclusion criteria. For example, in Collins et al.'s review, they inclusion criteria were as follows:

"Abstracts had to mention, in some capacity [our emphasis], differences in health outcomes or well-being, and/or the SDOH [social determinants of health]. Abstracts that discussed policy implications were also of distinct interest for review, but this was not an explicit inclusion criterion. Abstracts that described health differences in a strictly clinical scope were excluded, as were abstracts that referred to inequalities or disparities in a different context (e.g., measurement disparities). Highly technical pieces that discussed new clinical technologies, or issues related to healthcare systems and/or delivery, were excluded. Abstracts were also excluded if they contained the words "National Population Health Survey" or "Ottawa Charter for Health Promotion", but lacked any other information relevant to the review." (4)

A point worth noting from the example above is that the reviewers, quite rightly, did not exclude any documents based on any methodological hierarchy or technical checklist.

Once documents have been selected for inclusion, 'quality' appraisal takes place and how this is operationalised in a review is best illustrated in the methodological paper to Greenhalgh et al.'s review, 'Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations". For each of the eight research traditions, different quality criteria were used to judge the data within each tradition.

"We judged primary studies in any one tradition according to the quality criteria set by experts within that tradition … Reassuringly, we found that studies with comparable design tended to be judged similarly whatever the research tradition (for example, a survey of organisational attributes in the management literature would be judged by similar criteria and standards as a survey of consumer views in psychology, namely, appropriateness of sampling frame, validity of questionnaire items, completeness of response, and so on). Furthermore, whilst all traditions whose methodological toolkit included (say) the survey classified this as a high-quality method, those traditions whose toolkit did not include the survey were dismissive of any work based on this method, regardless of the research question being considered."(2)

6.4 Reflection activity for finding the right evidence

To assist reviewers in developing a suitable search strategy and in selecting and appraising documents appropriately, we have developed a list of questions a reviewer / review team might like to ask themselves: These questions are based on the quality standards in Tables 3, 4 and 5 and are listed in Box 3. We suggest that a reviewer might like to go through the questions in Box 3 to work out if the questions are relevant to their review and then how each question might be addressed.

Box 3: Questions to assist developing a search strategy and selection and appraisal of documents

Developing a search Strategy

- How will you ensure that your search process is such that it would help you identify research (or epistemic) traditions and map them?
- Is the necessary searching expertise available to you? If not, what will you do to remedy this?
- Will your search be piloted and refined?
- Will further searching be undertaken if additional sources are judged to be needed?

Selection and appraisal of documents

- Is relevance being used to guide the selection process? If not, why not?
- Are selected documents going to be quality appraised using criteria accepted within each tradition? If not, why not?

Conclusion

Meta-narrative review is a relatively new method of systematic review, as such the method is likely to develop and evolve in time with use. In this document we have focussed on the specific areas in undertaking a meta-narrative review which we have noted have been the sources of frequently encountered difficulties and misconceptions for meta-narrative reviewers. We have deliberately focussed this document towards the needs of less experienced reviewers as we felt that this group had the greatest developmental. How a meta-narrative review is undertaken will vary greatly and so it is impossible to be prescriptive and restrictive on what must be done. Instead we see our training materials more as guidance than ‘must-dos’.
We anticipate that with the growing use of meta-narrative reviews, new challenges and learning needs will emerge. We de believe that our quality standards and training materials should evolve to take into account methodological develop. We would therefore welcome and invite interested researchers to join us in updating and developing meta-narrative review methodology. Please contact us by email or via the RAMESES JISCM@ail (www.jiscmail.ac.uk/RAMESES).

Glossary

Normal Science
Normal science is a paradigm along with the practices and empirical approaches which are taken for granted by scientists within a particular tradition.

Paradigm
A paradigm is a particular way of viewing the world, including assumptions about how the world works, what are the important questions in a particular topic area, and what study designs and methods are best for adding to the knowledge base.

Research or epistemic tradition
A research tradition comprises studies building on what has gone before, each building on what has gone before, usually situated within a coherent paradigm, though an interdisciplinary tradition may bridge more than one paradigm.

An epistemic tradition is the unfolding of the underpinning set of philosophical assumptions which drive the development of theory and method; scholarship may progress via debate around these assumptions even in the absence of new empirical studies.

Reference List

(1) Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organisations: systematic literature review and recommendations for future research. Milbank Q 2004;82(4):581-629.

(2) Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. Soc Sci Med 2005 Jul;61(2):417-30.

(3) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards: meta-narrative reviews. BMC Medicine 2013;11:20.

(4) Collins P, Hayes M. The role of urban municipal governments in reducing health inequities: A meta-narrative mapping analysis. International Journal for Equity in Health 2010;9:13.

(5) Addis S, Davies M, Greene G, MacBrideStewart S, Shepherd M. The health, social care and housing needs of lesbian, gay, bisexual and transgender older people: a review of the literature. Health and Social Care in the Community 2009;17(6):647-58.

(6) Kitson A, Conroy T, Wengstrom Y, Profetto-McGrath J, Robertson-Malt S. Defining the fundamentals of care. Int J Nurs Pract 2010;16(4):423-34.

(7) Masuda J, Zupancic T, Poland B, Cole D. Environmental health and vulnerable populations in Canada: mapping an integrated equity-focused research agenda. Canadian Geographer-Geographe Canadien 2008;52(4):427-50.

(8) Gagliardi A, Brouwers M, Palda V, Lemieux-Charles L, Grimshaw J. How can we improve guideline use? A conceptual framework of implementability. Implement Sci 2011;6:26.

(9) Contandriopoulos D, Lemire M, Denis J-L, Tremblay E. Knowledge exchange processes in organizations and policy arenas: A narrative systematic review of the literature. Milbank Q 2010;88(4):444-83.

Appendix 11 List of all members of the online Delphi panels

Dave Baker, Sinai Hospital of Baltimore (Baltimore, MD, USA).

Marcello Bertotti, University of East London (London, UK).

Allan Best, InSource (Vancouver, BC, Canada).

Margaret Cargo, University of South Australia (Adelaide, SA, Australia).

Simon Carroll, University of Victoria (Victoria, BC, Canada).

Colleen Davison, Queens University, (Kingston, ON, Canada).

Marjolein Dieleman, Royal Tropical Institute (Amsterdam, the Netherlands).

Tim Dornan, Maastricht University (Maastricht, the Netherlands).

Ruth Garside, Peninsula College of Medicine and Dentistry (Exeter, UK).

Bradford Gray, Milbank Quarterly (New York, NY, USA).

Joanne Greenhalgh, University of Leeds (Leeds, UK).

Lois Jackson, Dalhousie University (Halifax, NS, Canada).

Justin Jagosh, McGill University (Montreal, QC, Canada).

Monika Kastner, University of Toronto (Toronto, ON, Canada).

James Lamerton, Sunshine Coast Division of General Practice (Cotton Tree, QLD, Australia).

Fraser MacFarlane, Queen Mary, University of London (London, UK).

Bruno Marchal, Institute of Tropical Medicine (Antwerp, Belgium).

Tracey McConnell, Queen's University (Belfast, UK).

Gemma Moss, Institute of Education (London, UK).

Douglas Noble, Queen Mary, University of London (London, UK).

Patricia O'Campo, University of Toronto (Toronto, ON, Canada).

Mark Pearson, Peninsula College of Medicine and Dentistry (Exeter, UK).

Pierre Pluye McGill University (Montreal, QC, Canada).

Henry Potts, University College London (London, UK).

Barbara Riley, University of Waterloo, (Waterloo, ON, Canada).

Glenn Robert, Kings College London (London, UK).

Jessie Saul, North American Research & Analysis, Inc. (Faribault, MN, USA).

Paul Shekelle, RAND Corporation (Santa Monica, CA, USA).

Neale Smith, University of British Columbia (Vancouver, BC, Canada).

Sanjeev Sridharan, University of Toronto (Toronto, ON, Canada).

Deborah Swinglehurst, Queen Mary, University of London (London, UK).

Nick Tilley, University College London (London, UK).

Kieran Walshe, University of Manchester (Manchester, UK).

All the RAMESES Project team members were also members of the Delphi panel.

Appendix 12 Project protocol

Greenhalgh et al. BMC Medical Research Methodology 2011, 11:115
http://www.biomedcentral.com/1471-2288/11/115

BMC Medical Research Methodology

STUDY PROTOCOL Open Access

Protocol - realist and meta-narrative evidence synthesis: Evolving Standards (RAMESES)

Trisha Greenhalgh[1*], Geoff Wong[1], Gill Westhorp[2] and Ray Pawson[3]

Abstract

Background: There is growing interest in theory-driven, qualitative and mixed-method approaches to systematic review as an alternative to (or to extend and supplement) conventional Cochrane-style reviews. These approaches offer the potential to expand the knowledge base in policy-relevant areas - for example by explaining the success, failure or mixed fortunes of complex interventions. However, the quality of such reviews can be difficult to assess. This study aims to produce methodological guidance, publication standards and training resources for those seeking to use the realist and/or meta-narrative approach to systematic review.

Methods/design: We will: [a] collate and summarise existing literature on the principles of good practice in realist and meta-narrative systematic review; [b] consider the extent to which these principles have been followed by published and in-progress reviews, thereby identifying how rigour may be lost and how existing methods could be improved; [c] using an online Delphi method with an interdisciplinary panel of experts from academia and policy, produce a draft set of methodological steps and publication standards; [d] produce training materials with learning outcomes linked to these steps; [e] pilot these standards and training materials prospectively on real reviews-in-progress, capturing methodological and other challenges as they arise; [f] synthesise expert input, evidence review and real-time problem analysis into more definitive guidance and standards; [g] disseminate outputs to audiences in academia and policy. The outputs of the study will be threefold:
1. Quality standards and methodological guidance for realist and meta-narrative reviews for use by researchers, research sponsors, students and supervisors
2. A 'RAMESES' (Realist and Meta-review Evidence Synthesis: Evolving Standards) statement (comparable to CONSORT or PRISMA) of publication standards for such reviews, published in an oper-access academic journal.
3. A training module for researchers, including learning outcomes, outline course materials and assessment criteria.

Discussion: Realist and meta-narrative review are relatively new approaches to systematic review whose overall place in the secondary research toolkit is not yet fully established. As with all secondary research methods, guidance on quality assurance and uniform reporting is an important step towards improving quality and consistency of studies.

Keywords: systematic review, realist review or synthesis, meta-narrative review

Background

Introduction

Academics and policymakers are increasingly interested in 'policy-friendly' approaches to evidence synthesis which seek to illuminate issues and understand contextual influences on whether, why and how interventions might work [1-4]. A number of different approaches have been used to try to address this goal. Qualitative and mixed-method reviews are often used to supplement, extend and in some circumstances replace Cochrane-style systematic reviews [5-11]. Theory-driven approaches to such reviews include realist and meta-narrative review. Realist review was originally developed by Pawson for complex social interventions to explore systematically how contextual factors influence the link between intervention and outcome (summed up in the question "what works, how, for whom, in what

* Correspondence: p.greenhalgh@qmul.ac.uk
[1]Healthcare Innovation and Policy Unit, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, London E1 2AB, UK
Full list of author informaton is available at the end of the article

circumstances and to what extent?") [12,13]. Greenhalgh et al. developed meta-narrative review as an adaptation of realist review, for use when a policy-related topic has been researched in different ways by multiple groups of scientists, especially when key terms have different meanings in different literatures [14].

Quality checklists and reporting standards are common (and, increasingly, expected) in health services research - see for example CONSORT for randomised controlled trials [15], AGREE for clinical guidelines [16], PRISMA for Cochrane-style systematic reviews [17] and SQUIRE for quality improvement studies [18]. They have two main purposes: they help researchers design and undertake robust studies, and they help reviewers and potential users of research outputs assess validity and reliability. This project seeks to produce a set of quality criteria and comparable reporting guidance for realist and meta-narrative reviews.

What are realist and meta-narrative reviews?

Realist and meta-narrative reviews are systematic, theory-driven interpretative techniques, which were developed to help make sense of heterogeneous evidence about complex interventions applied in diverse contexts in a way that informs policy. Interventions have been described as "theory incarnate" [19], driven by hypotheses, hunches, conjectures and aspirations about individual and social betterment. Strengthening a review process that helps to sift and sort these theories may be an important step in producing better interventions.

Realist review seeks to unpack the relationships between context, mechanism and outcomes (sometimes abbreviated as C-M-O) - i.e. how particular contexts have 'triggered' (or interfered with) mechanisms to generate the observed outcomes [4]. Its philosophical basis is realism, which assumes the existence of an external reality (a 'real world') but one that is 'filtered' (i.e. perceived, interpreted and responded to) through human senses, volitions, language and culture. Such human processing initiates a constant process of self-generated change in all social institutions, a vital process that has to be accommodated in evaluating social programmes.

In order to understand how outcomes are generated, the roles of both external reality and human understanding and response need to be incorporated. Realism does this through the concept of mechanisms, whose precise definition is contested but for which a working definition is '...underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest.' [20]. Different contexts interact with different mechanisms to make particular outcomes more or less likely - hence a realist review produces recommendations of the general format "In situations [X], complex intervention [Y], modified in this way and taking account of these contingencies, may be appropriate". Realist reviews can be undertaken in parallel with traditional Cochrane reviews (see the complementary Cochrane and realist reviews of school feeding programmes in disadvantaged children [21,22]). The Cochrane review produced an estimate of effect size whilst the realist review addressed why and how school feeding programmes 'worked', explained examples of when they did not 'work', and produced practical recommendations for policymakers.

Meta-narrative review was originally developed by Greenhalgh et al. to try to explain the apparently disparate data encountered in their review of diffusion of innovation in healthcare organisations [14,23]. Core concepts such as 'diffusion', 'innovation', 'adoption' and 'routinisation' had been conceptualised and studied very differently by researchers from a wide range of primary disciplines including psychology, sociology, economics, management and even philosophy. Whilst some studies had been framed as the implementation of a complex intervention in a social context (thus lending themselves to a realist analysis), others had not. Preliminary questions needed to be asked, such as "what exactly did these researchers mean when they used the terms 'diffusion', 'innovation' and so on?", "how did they link the different concepts in a theoretical model - either as a context-mechanism-outcome proposition or otherwise?" and "what explicit or implicit assumptions were made by different researchers about the nature of reality?".

These questions prompted the development of meta-narrative review, which sought to illuminate the different paradigmatic approaches to a complex topic area by considering how the 'same' topic had been differently conceptualised, theorised and empirically studied by different groups of researchers. Meta-narrative review is particularly suited to topics where there is dissent about the nature of what is being studied and what is the best empirical approach to studying it. For example, Best et al., in a review of knowledge translation and exchange, asked how different research teams had conceptualised the terms 'knowledge', 'translation' and 'exchange' - and what different theoretical models and empirical approaches had been built on these different conceptualisations [24]. Thus meta-narrative review potentially offers another strategy to assist policy makers to understand and interpret a conflicting body of research, and therefore to use it more effectively in their work.

The need for standards in theory-driven systematic reviews

Realist and meta-narrative approaches can capitalise on and help build common ground between social researchers and policy teams. Many researchers are attracted to these approaches because they allow

systematic exploration of how and why complex interventions work. Policymakers are attracted to them because they are potentially able to answer questions relevant to practical decisions (not merely "what is the impact of X?" but "if we invest in X, to which particular sectors should we target it, how might implementation be improved and how might we maximise its impact?")

Whilst interest in such approaches is burgeoning, it is our experience that these approaches are sometimes being applied in ways that are not always true to the core principles set out in previous methodological guidance [4,13,25,26]. Some reviews published under the 'realist' banner are not systematic, not theory-driven and/or not consistent with realist philosophy. The meta-narrative label has also been misapplied in reviews which have no systematic methodology. For these reasons, we believe that the time has come to develop formal standards and training materials.

There is a philosophical problem here, however. Realist and meta-narrative approaches are interpretive processes (that is, they are based on building plausible evidenced explanations of observed outcomes, presented predominantly in narrative form), hence they do not easily lend themselves to a formal procedure for quality checking. Indeed, we have argued previously that the core tasks in such reviews are thinking, reflecting and interpreting [4,27]. In these respects, realist and meta-narrative reviews face a problem similar to that encountered in assessing qualitative research - namely the extent to which guidelines, standards and checklists can ever capture the essence of quality. Some qualitative researchers are openly dismissive of the 'technical checklist' approach as an assurance of quality in systematic review [28]. Whilst we acknowledge such views, we believe that from a pragmatic perspective, formal quality criteria - with appropriate caveats - are likely to add to, rather than detract from, the overall quality of outputs in this field. Scientific discovery is never the mere mechanical application of set procedures [29]. Accordingly, research protocols should aim to guide rather than dictate.

The online Delphi method

This study will use the online Delphi method and in this section we introduce, explain and justify our use of this method. The essence of the Delphi technique is to engender reflection and discussion amongst a panel of experts with a view to getting as close as possible to consensus and documenting both the agreements reached and the nature and extent of residual disagreement [30]. It was used, for example, to set the original care standards which formed the basis of the Quality and Outcomes Framework for United Kingdom general practitioners [31]. Factors which have been shown to influence quality in the Delphi process include: [a] composition (expertise, diversity) of the expert panel; [b] selection of background papers and evidence to be discussed by that panel (completeness, validity, representativeness); [c] adequacy of opportunities to read and reflect (balance between accommodating experts' busy schedules and keeping to study milestones); [d] qualitative analysis of responses (depth of reflection and scholarship, articulation of key issues); [e] quantitative analysis of responses (appropriateness and accuracy of statistical analysis, clarity of presentation when this is fed back); and [f] how dissent and ambiguity are treated (e.g. avoidance of 'groupthink', openness to dissenting voices) [30,32,33].

Evidence suggests that the online medium is more likely to improve than jeopardise the quality of the consensus development process. Mail-only Delphi panels have been shown to be as reliable as face-to-face panels [34]. Asynchronous online communication has well-established benefits in promoting reflection and knowledge construction [35]. There are over 100 empirical examples of successful online Delphi studies conducted between geographically dispersed participants (see for example [33,36-40]). We have been unable to find any online Delphi study which identified the communication medium as a significant limitation. On the contrary, many authors described significant advantages of the online approach, especially when dealing with an international sample of experts. One group commented: *"Our online review process was less costly, quicker, and more flexible with regard to reviewer time commitment, because the process could accommodate their individual schedules."* [40].

Critical commentaries on the Delphi process have identified a number of issues which may prove problematic, for example *"issues surrounding problem identification, researcher skills and data presentation"* [30] or *"the definition of consensus; the issue of anonymity vs. quasi-anonymity for participants; how to estimate the time needed to collect the data, analyse each 'round', feed back results to participants, and gain their responses to this feedback; how to define and select the 'experts' who will be asked to participate; how to enhance response rates; and how many 'rounds' to conduct."* [33]. These comments suggest that it is the underlying design and rigour of the research process which is key to the quality of the study, and not the medium through which this process happens.

Methods/design

Research questions

1. What are the key steps in producing a valid and reliable systematic review using a realist or meta-narrative approach?

2. How might 'high' and 'low' quality in such reviews be defined and assessed [a] at the grant application stage; [b] during the review; [c] at publication stage and [d] by end-users of such reviews?

3. What are the key learning outcomes for a student of realist or meta-narrative review, and how might performance against these outcomes be assessed?

Study design

Literature review, iterative online Delphi panel and real-time engagement with new, ongoing reviews (Figure 1).

Study objectives

1. To collate and summarise the literature on the principles of good practice in realist and meta-narrative reviews, highlighting in particular how and why these differ from conventional forms of systematic review and from each other.

2. To consider the extent to which these principles have been followed by published and in-progress reviews, thereby identifying how rigour may be lost and how existing principles could be improved.

3. Using an online Delphi method with an interdisciplinary panel of experts from academia and policy, to produce, in draft form, an explicit and accessible set of methodological guidance and publication standards.

4. To produce training materials with learning outcomes linked to these steps and standards.

5. To pilot these standards and training materials prospectively on real reviews-in-progress, capturing methodological and other challenges as they arise.

6. To synthesise expert input, evidence review and real-time problem analysis into more definitive guidance and standards

7. To disseminate these guidance and standards to audiences in academia and policy

(1) and (2) will be achieved via a narrative review of the literature and supplemented by collating feedback from presentation(s) and workshop(s). These will feed into (3), which will be achieved via an online Delphi panel. The panel will include wide representation from researchers, students, policymakers, theorists and research sponsors. For (4), we will draw on our experience in developing and delivering relevant education modules. For (5), we will capture new realist reviews in progress as people approach us for help and guidance and seek their informed participation in piloting the new materials. (6) and (7) will be addressed by preparing academic publications, online resources and by delivering presentations and workshops.

Intended outputs

We aim to generate three main outputs:

1. Quality standards and methodological guidance for realist and meta-narrative reviews for use by researchers, research sponsors, students and supervisors

2. A 'RAMESES' statement (comparable to CONSORT or PRISMA) of publication standards for such reviews, published in an open-access academic journal.

3. A training module for researchers, including learning outcomes, outline course materials and assessment criteria.

Management and governance

The development of guidelines and guidance is a complex and contested process [41]. It is crucial to avoid the 'GOBSAT' (good old boys sat around a table) approach and ensure that [a] those who contribute to the process represent a diverse, informed and representative sample from both academia and policymaking and that [b] the process itself is systematic, auditable and justifiable. To that end, we will have a small core research team which will meet regularly to review progress, set the next work phase and produce minutes. We will report six-monthly to an advisory steering group, to whom we will present a project update and financial report.

In addition, approximately halfway through the study period, we will present our emerging findings formally to a panel of external researchers in order to collate additional feedback in a technique known as the 'fishbowl'. We will recruit a maximum variety sample of approximately 10 experts in systematic review. The main criterion for inclusion will be academic standing in the critical appraisal and evaluation of qualitative research studies and/or in evidence synthesis, including but not limited to those already familiar with realist or meta-narrative review. We will circulate materials in advance of the fishbowl workshop, including goals of the project, methodology and provisional standards and guidance. The fishbowl session will comprise presentation from the research team followed by discussion, facilitated by someone outside the core research team. The session will be recorded and minuted, and recommendations used to inform revision of the protocol as needed.

The study was deemed exempt from NHS research ethics approval (personal communication S Burke 14.2.11, East London and City Research Ethics Committee).

Details of literature search methods

Our initial exploratory searches have found that the literature in this field is currently small but is expanding rapidly, and that it is of broad scope, variable quality and inconsistently indexed. The purpose of identifying published reviews is not to complete a census of realist and meta-narrative studies. Our comprehensive search will allow us to pinpoint real examples (or publications claiming to be examples) which provide rich detail on

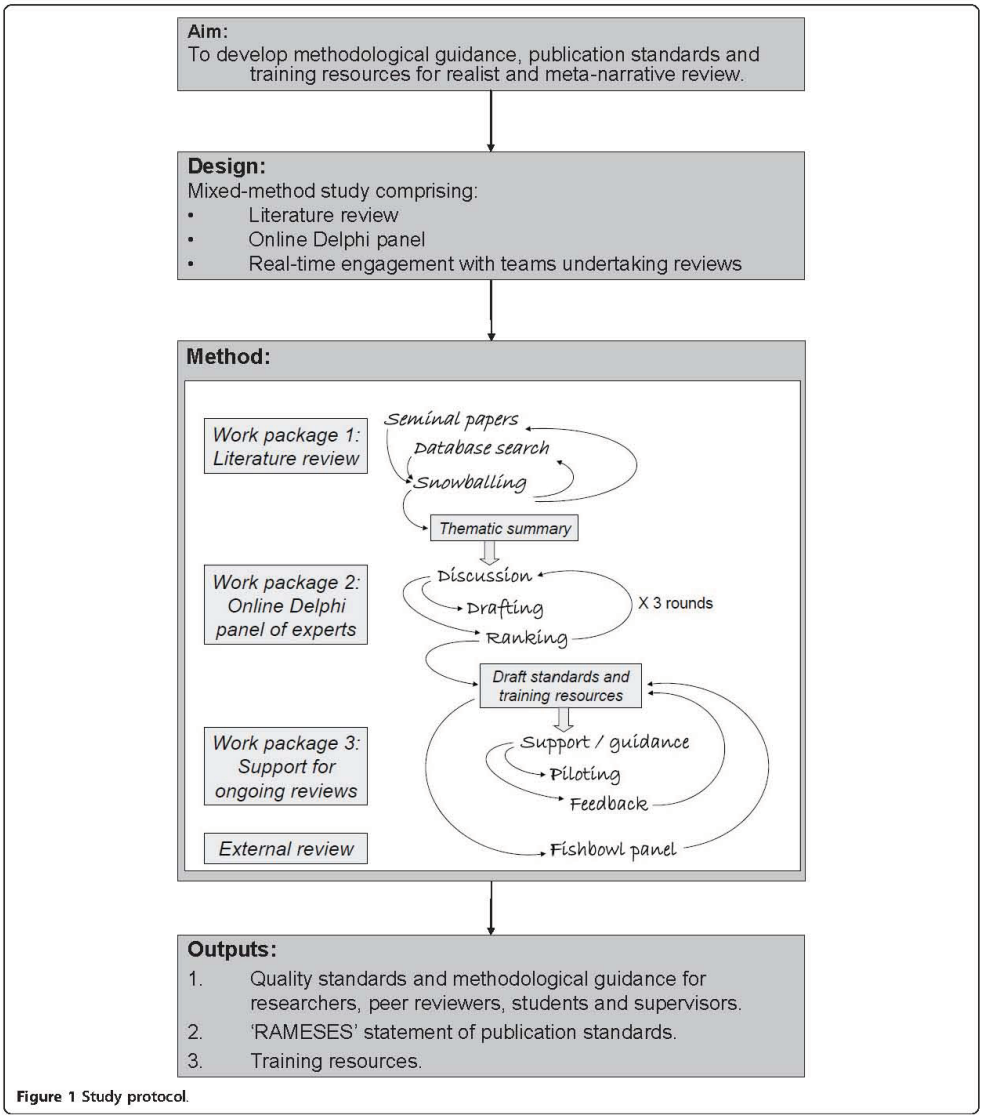


Figure 1 Study protocol.

their usage of those review activities we wish to scrutinise and formalise. To that end, and drawing on a previous study which demonstrated the effectiveness and efficiency of the methods proposed [42], and employing the skills of a specialist librarian, we will employ three approaches:

1. Identifying seminal sources known to the research team and other experts in the field (e.g. via relevant networks and email lists).

2. Snowballing both backwards (pursuing references of references) and forwards (using citation-tracking software to identify subsequent publications citing the index

paper) from seminal theoretical/methodological publications and empirical examples of realist and meta-narrative reviews. For reviews of heterogeneous bodies of evidence, snowball techniques are more effective and efficient than hand searching or using predefined search strings on electronic databases [42].

3. Database searching, especially with a view to identifying grey literature such as PhDs and unpublished reports (some will represent robust and critical applications of the methods and others will highlight 'commonly occurring mistakes and misconceptions').

In addition to identifying a broad range of examples of actual reviews, we will also capture papers describing methodological and theoretical critiques of the approaches being studied.

We will conduct a thematic analysis of this literature which will *initially* be oriented to addressing six questions, but to which we will add additional questions and topic areas (in order to better capture our analysis and understanding of the literature) as these emerge from our reading of the papers:

1. What are the strengths and weaknesses of realist and meta-narrative review from both a theoretical and a practical perspective?

2. How have these approaches actually been used? Are there areas where they appear to be particularly fit (or unfit) for purpose?

3. What, broadly, are the characteristics of high-quality (and low-quality) reviews undertaken by realist or meta-narrative methods? What can we learn from the best (and worst) examples so far?

4. What challenges have reviewers themselves identified (e.g. in the introduction or discussion sections of their papers) in applying these approaches? Are there systematic gaps between the 'theory' and the steps actually taken?

5. What is the link between realist and meta-narrative review and the policymaking process? How have published reviews been commissioned or sponsored? How have policymakers been involved in shaping the review? How have they been involved in disseminating and applying its findings? Are there models of good practice (and of approaches to avoid) for academic-policy linkage in this area?

6. How have front-line staff and service users been involved in realist and meta-narrative reviews? If the answer to this is 'usually, not much', how might they have been involved and are there examples of potentially better practice which might be taken forward?

7. How should one choose between realist, meta-narrative and other theory-driven approaches when selecting a review methodology? How might (for example) the review question, purpose and intended audience(s) influence the choice of review method?

The output of this phase will be a provisional summary organised under the above headings and highlighting for each question the key areas of knowledge, ignorance, ambiguity and uncertainty. This will be distributed to the Delphi panel as the starting-point for their guidance development work.

Details of online Delphi process

We will follow an online adaptation of the Delphi method (see above) which we have developed and used in a previous study to produce guidance on how to critically appraise research on illness narratives [38]. In that study, a key component of a successful Delphi process was recruiting a wide range of experts, policymakers, practitioners and potential users of the guidance who could approach the problem from different angles, and especially people who would respond to academic suggestions by asking "so-what" questions.

Placing the academic-policy/practice tension central to this phase of the research, we hope to construct our Delphi panel to include a majority of experienced academics (e.g. those who have published on theory and method in realist and/or meta-narrative review). We will also hope to recruit policymakers, research sponsors and representatives of third sector organisations. These individuals will be recruited by approaching relevant organisations and email lists (e.g. professional networks of systematic reviewers, C.H.A.I.N., INVOLVE), providing an outline of the study and selecting those with greatest commitment and potential to balance the sample.

We will draw on our own experience of developing standards and guidance, as well as on published papers by CONSORT, PRISMA, AGREE, SQUIRE and other teams working on comparable projects [15,17,18,43].

The Delphi panel will be conducted entirely via the Internet using a combination of email and online survey tools. It will begin with a 'brainstorm' round ('round 1') in which participants will be invited to submit personal views, exchange theoretical and empirical papers on the topic and suggest items that might could be included in the publication standards. This will be done as a warm-up exercise and panel members will be sent our own preliminary summary (see above). These early contributions, along with our summary, will be collated and summarised in a set of provisional statements, which will be listed in a table and sent to participants for ranking ('round 2'). Participants will be asked to rank each item twice on a 9-point Likert scale (1 = strongly against to 9 = strongly in favour), once for relevance (i.e. should a statement on this theme/topic be included at all in the guidance?) and once for validity (i.e. to what extent do you agree with this statement as currently worded?). Those who agree that a statement is relevant but disagree on its wording will be invited to suggest changes

to the wording. In this second round, participants will again be invited to suggest additional topic areas and items.

Each participant's responses will be collated and the numerical rankings entered onto an Excel spreadsheet. Median, inter-quartile and maximum-minimum range for each response will be calculated. Statements that score low on relevance will be omitted from subsequent rounds. Further online discussion will be invited on statements that score high on relevance but low on validity (indicating that a rephrased version of the statement is needed) and on those where there is wide disagreement about relevance or validity. Following discussion, a second list of statements will be drawn up and circulated for ranking ('round 3'). The process of collation of responses, further email discussion, and re-ranking will be repeated until maximum consensus is reached ('round 4' et seq.). In practice, very few Delphi panels, online or face to face, go beyond three rounds since participants tend to 'agree to differ' rather than move towards further consensus [38].

Residual non-consensus will be reported as such and the nature of the dissent described. Making such dissent explicit tends to expose inherent ambiguities (which may be philosophical or practical) and acknowledges that not everything can be resolved; such findings may be more use to reviewers than a firm statement which implies that all tensions have been "fixed".

Preparing teaching and learning resources

A key objective of this study is to produce publicly accessible resources to support training in realist and meta-narrative review. We anticipate that these resources will need to be adapted and perhaps supplemented for different groups of learners, and interactive learning activities added [44]. Taking account of the format and orientation of other comparable materials (e.g. courses produced by the International Cochrane and Campbell Collaborations), though not necessarily aligning with these, we will develop and pilot draft learning objectives, example course materials and teaching and learning support methods. We will draw on our previous work on course development, quality assurance and support for interactive and peer-supported learning in healthcare professionals [35,44-46]

Real-time piloting

The sponsor of this study, the National Institute for Health Research Service Delivery and Organisation (NIHR SDO) Programme, supports secondary research calls for rapid, policy-relevant reviews, some though not all of which seek to use realist or meta-narrative methods. We will work with a select sample of teams funded under such calls, as well as other teams engaged in relevant ongoing reviews (selected to balance our sample), to share emerging recommendations and gather real-time data on how feasible and appropriate these recommendations are in a range of different reviews. Over the 27-month duration of this study, we anticipate recruiting two cohorts of review teams over the course of this study: with the first cohort, we will use provisional standards, guidance and training materials based on our initial review of the literature. With the second cohort, we will pilot the standards, guidance and training materials which have been produced/refined via the Delphi process. After following two cohorts of review teams through their reviews, we will further revise the outputs as a master document before considering how to modify these for different audiences.

Training and support offered to these review teams will consist of three overlapping and complementary packages:

1. An 'all-comers' online discussion forum via Jiscm@il http://www.jiscmail.ac.uk/RAMESES for interested reviewers who are currently doing or have previously attempted a realist or meta-narrative review. This will be run via 'light-touch' facilitation in which we invite discussion on particular topics and periodically summarise themes and conclusions (a technique known in online teaching as 'weaving'). Such a format typically accommodates large numbers of participants since most people tend to 'lurk' most of the time. Such discussion groups tend to generate peer support through their informal, non-compulsory ethos and a strong sense of reciprocity (i.e. people helping one another out because they share an identity and commitment) [47] and they are often rich sources of qualitative data. We anticipate that this forum will contribute key themes to the quality and reporting standards and learning materials throughout the duration of the study.

2. Responsive support to our designated review teams. Our input to these teams will depend on their needs, interests and previous experience and hence is impossible to stipulate in detail in advance. In our previous dealings with review teams we have been called upon (for example) to assist them in distinguishing 'context' from 'mechanism' in a particular paper, extracting and formalising programme theories, distinguish middle-range theories from macro or micro theories, develop or adapt data extraction tools, advise on data extraction techniques, and train researchers in the use of qualitative software for systematic review.

3. A 'learning set' series of workshops for designated review teams. Much of the learning in such workshops is likely to come from the review teams themselves, and if participants are experienced and wish to offer teaching to others on particular relevant topics this will be encouraged. For the first workshop we will prepare a

core syllabus of basic training oriented to explicit learning outcomes, delivered as a combination of prior self-study materials and short taught sessions on the day. Even at the first workshop, however, most of the time will be spent applying the basic principles to the real worked examples of reviews being undertaken.

As explained above, the first cohort of review teams will be run as a pilot and we will explain this to the participants, thereby gaining their active engagement in improving the programme for subsequent learners.

Discussion

Realist and meta-narrative reviews are relatively new systematic review methods in health services research. They potentially offer great promise in unpacking the 'black box' of the many complex interventions that are increasingly being used to improve health and patient outcomes. As relatively experienced users of these methods, we have noted a number of common and recurrent challenges that face grant awarding bodies, peer-reviewers, reviewers and users. These centre on two closely related questions, namely how to judge if a realist or meta-narrative review, or a proposal for such a review, is of 'high quality' (including, for completed reviews, how 'credible' and 'robust' findings are) and how to undertake such reviews. Our experience to date suggests that we can go a long way towards answering these questions by giving due consideration to the theoretical and conceptual underpinnings of realist and meta-narrative reviews, outlined briefly below.

Realist review is based on a realist philosophy of science, which permeates and informs its underlying epistemological assumptions, methodology and quality considerations. Meta-narrative review takes a more constructivist philosophical position, though it is compatible with approaches which propose the existence of a social reality independent of our constructions of it. The meta-narrative approach seeks to tease out and explore the full range of philosophical positions represented in the primary literature.

One of the most common misapplications we have noted is that reviewers have not always appreciated the underlying philosophical basis of these review methods (and the implications of these for how the review should be conducted). Instead, they have based their reviews explicitly or implicitly on fundamentally different philosophical assumptions - most commonly the positivist notion that generalisable truths are best generated from controlled experiments, especially randomised trials.

Even when a realist philosophy of science has been adhered to in a realist review, reviewers - ourselves included - often struggle with recurring conceptual and methodological issues. 'Mechanisms' present a particular challenge in realist review - how to define them, where to locate them, how to identify them and how to test and refine them. Both review methods trade on the use of theoretical explanations to make sense of the observed data. Realist reviewers commonly grapple with how to define a theory (what, for example, is the difference between a 'programme theory' and a 'middle-range theory'?) and what level of abstraction is appropriate in what circumstances. On a more pragmatic level, those who seek to produce theory-driven reviews of heterogeneous topic areas wrestle with a broad range of 'how to' issues: how to define the scope of the review; how and to what extent to refine this scope as the review unfolds; what literature(s) to search and how; how to 'critically appraise' what is often a very diverse sample of primary studies; how to collate, analyse and synthesise findings; how to make recommendations that are academically defensible and useful to policymakers; and so on.

In conclusion, whilst realist and meta-narrative reviews hold much promise for developing theory and informing policy in some of the health sector's most pressing questions, misunderstandings and misapplications of these methods are common. The time is ripe to start on the iterative journey of producing guidance on quality and reporting standards as well as developing quality-assured learning resources to ensure that funding decisions, execution, reporting and use of these review methods is optimised. Acknowledging that research is never static, the RAMESES project does not seek to produce the last word on this topic but to capture current expertise and establish an agreed 'state of the science' on which future researchers will no doubt build.

The Delphi panel will commence in September 2011 and we anticipate that a paper describing the guidance will be submitted by September 2012. The online discussion forum is open to anyone with an interest in realist and meta-narrative reviews and may be found at http://www.jiscmail.ac.uk/RAMESES

Acknowledgements and funding

This work was funded by a research grant from the NHS SDO Programme (10/1008/07). We thank the participating review teams for their contribution to this work and members of the external steering group.

Author details

[1]Healthcare Innovation and Policy Unit, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, London E1 2AB, UK. [2]Community Matters, PO Box 443, Mt Torrens, South Australia, SA 5244, Australia. [3]School of Sociology and Social Policy, University of Leeds, Leeds, LS2 9JT, UK.

Authors' contributions

TG conceptualised the study with input from GWo and RP. TG wrote the first draft and GWo, GWe and RP critically contributed to and refined this manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 5 March 2011 Accepted: 16 August 2011
Published: 16 August 2011

References

1. Berwick DM: **The science of improvement.** *JAMA* 2008, **299**:1182-1184.
2. Lavis JN: **How can we support the use of systematic reviews in policymaking?** *PLoS Med* 2009, **6**:e1000141.
3. Mays N, Pope C, Popay J: **Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field.** *J Health Serv Res Policy* 2005, **10**(Suppl 1):6-20.
4. Pawson R, Greenhalgh T, Harvey G, Walshe K: **Realist review–a new method of systematic review designed for complex policy interventions.** *J Health Serv Res Policy* 2005, **10**(Suppl 1):21-34.
5. Popay J, Rogers A, Wiliams G: **Rationale and standards for the systematic review of qualitative literature in health services research.** *Qualitative Health Research* 1998, **8**:341-351.
6. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A: **Synthesising qualitative and quantitative evidence: a review of possible methods.** *J Health Serv Res Policy* 2005, **10**:45-53.
7. Tricco AC, Tetzlaff J, Moher D: **The art and science of knowledge synthesis.** *J Clin Epidemiol* 2010.
8. Lucas PJ, Baird J, Arai L, Law C, Roberts HM: **Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews.** *BMC Med Res Methodol* 2007, **7**:4.
9. Oxman AD, Schunemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006, **4**:20.
10. Thomas J, Harden A: **Methods for the thematic synthesis of qualitative research in systematic reviews.** *BMC Med Res Methodol* 2008, **8**:45.
11. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, Facey K, Hailey D, Norderhaug I, Maddern G: **Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment.** *Int J Technol Assess Health Care* 2008, **24**:133-139.
12. Pawson R: **Evidence-based policy: the promise of 'realist synthesis'.** *Evaluation* 2002, **8**:340-358.
13. Pawson R: *Evidence-based policy: A realist perspective* London: Sage; 2006.
14. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R: **Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review.** *Soc Sci Med* 2005, **61**:417-430.
15. Moher D, Schulz KF, Altman DG: **The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials.** *Lancet* 2001, **357**:1191-1194.
16. AGREE collaboration: **Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project.** *Qual Saf Health Care* 2003, **12**:18-23.
17. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, Clarke M, Devereaux P, Kleijnen J, MMoher D: **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration.** *BMJ* 2009, **339**:b2700.
18. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S, the SQUIRE Development Group: **Publication Guidelines for Improvement Studies in Health Care: Evolution of the SQUIRE Project.** *Annals of Internal Medicine* 2008, **149**:670-676.
19. Weiss C: **Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families.** In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts.* Edited by: Connell J, Others. Washington: Aspen Institute; 1995:.
20. Astbury B, Leeuw F: **Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation.** *American Journal of Evaluation* 2010, **31**:363-381.
21. Greenhalgh T, Kristjansson E, Robinson V: **Realist review to understand the efficacy of school feeding programmes.** *BMJ* 2007, **335**:858-861.
22. Kristjansson E, Robinson V, Petticrew M, Macdonald B, Krasevec J, Janzen L, Greenhalgh T, Wells G, Macgowan J, Farmer A. Shea B, Mayhew A, Tugwell P: **School feeding for improving the physical and psychosocial health of disadvantaged elementary school children.** *Cochrane Database Syst Rev* 2007, CD004676.
23. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O: **Diffusion of innovations in service organisations: systematic literature review and recommendations for future research.** *Milbank Q* 2004, **82**:581-629.
24. Best A, Terpstra JL, Moor G, Riley B, Norman CD, Glasgow RE: **Building knowledge integration systems for evidence-informed decisions.** *J Health Organ Manag* 2009, **23**:627-641.
25. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R: **Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review.** *Soc Sci Med* 2005, **61**:417-430.
26. Pawson R, Greenhalgh T, Harvey G, Walshe K: *Realist synthesis - an introduction. ESRC Working Paper Series* London: ESRC; 2004.
27. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R: **Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review.** *Soc Sci Med* 2005, **61**:417-430.
28. McClure M: **'Clarity bordering on stupidity': where's the quality in systematic review?** *Journal of Education Policy* 2005, **20**:393-416.
29. Popper K: *The Logic of Scientific Discovery* Vienna, Austria: Verlag von Julius Springer; 1935.
30. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32**:1008-1015.
31. Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK: **Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice.** *Qual Health Care* 1999, **8**:6-15.
32. Hsu C-C, Sandford B: **The Delphi Technique: Making Sense of Consensus.** *Practical Assessment Research & Evaluation* 2007, **12**.
33. Keeney S, Hasson F, McKenna H: **Consulting the oracle: ten lessons from using the Delphi technique in nursing research.** *J Adv Nurs* 2006, **53**:205-212.
34. Washington DL, Bernstein SJ, Kahan JP, Leape LL, Kamberg CJ, Shekelle PG: **Reliability of clinical guideline development using mail-only versus in-person expert panels.** *Med Care* 2003, **41**:1374-1381.
35. Russell J, Elton L, Swinglehurst D, Greenhalgh T: **Using the online environment in assessment for learning: A case-based study of a web-based course in primary care.** *Assessment and Evaluation in Higher Education* 2006, **31**:465-478.
36. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, *et al*: **Developing a quality criteria framework for patient decision aids: online international Delphi consensus process.** *BMJ* 2006, **333**:417.
37. Hart LM, Bourchier SJ, Jorm AF, Kanowski LG, Kingston AH, Stanley D, Lunman D: **Development of mental health first aid guidelines for Aboriginal and Torres Strait Islander people experiencing problems with substance use: a Delphi study.** *BMC Psychiatry* 2010, **10**:78.
38. Greenhalgh T, Wengraf T: **Collecting stories: is it research? Is it good research? Preliminary guidance based on a Delphi study.** *Med Educ* 2008, **42**:242-247.
39. Pye J, Greenhalgh T: **First aid kits for recreational dive boats: a Delphi study.** *Journal of Travel Medicine* 2010.
40. Holliday C, Robotin M: **The Delphi process: a solution for reviewing novel grant applications.** *Int J Gen Med* 2010, **3**:225-230.
41. Swinglehurst D: **Evidence-based guidelines: the theory and the practice.** *Evidence Based Medicine and Public Health* 2005, **9**:308-314.
42. Greenhalgh T, Peacock R: **Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources.** *BMJ* 2005, **331**:1064-1065.
43. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID: **Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines.** *BMC Health Serv Res* 2005, **5**:18.
44. Wong G, Greenhalgh T, Pawson R: **Internet-based medical education: a realist review of what works, for whom and in what circumstances.** *BMC Med Educ* 2010, **10**:12.
45. Greenhalgh T, Toon P, Russell J, Wong G, Plumb L, Macfarlane F: **Transferability of principles of evidence based medicine to improve educational quality: systematic review and case study of an online course in primary health care.** *BMJ* 2003, **326**:142-145.
46. Wong G, Greenhalgh T, Russell J, Boynton P, Toon P: **Putting your course on the Web: lessons from a case study and systematic literature review.** *Med Educ* 2003, **37**:1020-1023.

47. Russell J, Greenhalgh T, Boynton P, Rigby M: **Soft networks for bridging the gap between research and practice: illuminative evaluation of CHAIN.** *BMJ* 2004, **328**:1174.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2288/11/115/prepub

doi:10.1186/1471-2288-11-115
Cite this article as: Greenhalgh *et al.*: Protocol - realist and meta-narrative evidence synthesis: Evolving Standards (RAMESES). *BMC Medical Research Methodology* 2011 11:115.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library