Version Control

Number	Date	Reference
2	21.03.2024	For NIHR approval

Full title of project

Developing Realist Economic Evaluation Methods (REEM) and Guidance to Evaluate the Impact, Costs and Consequences of Complex Interventions

Short Title

Realist Economic Evaluation Methods (REEM)

Study reference numbers

Funder number	NIHR135102
IRAS	335718
Open Science Framework	osf.io/9zafj

Funding acknowledgement

This study is funded by the NIHR HSDR Programme (NIHR135102). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Key study contacts

Role	Name	Institution/Organisation	Email
PI	Dr Sonia Dalkin (SD) Dr Angela Bate (AB)	Northumbria University, UK Northumbria University, UK	s.dalkin@northumbria.ac.uk angela.bate@northumbria.ac.uk
CI	Prof Rob Anderson Prof Rachel Baker Prof Cam Donaldson Dr Meghan Kumar Gill Westhorp Dr Geoff Wong Mrs Judy Wright	University of Exeter, UK Glasgow Caledonian University, UK LSHTM, UK Charles Darwin University, AU University of Oxford, UK University of Leeds, UK	r.anderson@exeter.ac.uk rachel.baker@gcu.ac.uk cam.donaldson@gcu.ac.uk meghan.kumar@lshtm.ac.uk gillian.westhorp@cdu.edu.au geoffrey.wong@phc.ox.ac.uk j.m.wright@leeds.ac.uk
S/RA	Dr Sam Redgate Dr Andrew Fletcher Dr Aikaterini Papadopoulou	Northumbria University Northumbria University GCU, UK	
PPIE	Dr Felicity Shenton Ms Vivienne Hibberd	NIHR ARC NENC n/a	
Partner	Dr Hellen Barsosio Fiona Doring Prof Gerard Danjoux	Kenya Medical Research Institute (KEMRI) Impact Arts South Tees Hospitals NHS Foundation Trust	

Summary of research (abstract)

It is increasingly recognised that the ‘big problems’ in health and social care require well-designed complex solutions and robust evaluation, which itself is often complex. Realist evaluations were designed to take account of that complexity, offering an explanation of ‘what works, for who, in which circumstances and why?’. Concurrently, policymakers and research funders require economic appraisals to accompany evaluations of complex interventions, to inform difficult decisions in the context of resource scarcity. However, economic evaluation methods often ignore context and do not capture variations in resource use or outcomes across groups, or recognise the implications this may have for the relationship between resource use and outcomes. Additionally, realist evaluations do not tend to explicitly capture

the economic consequences of interventions. This research will develop realist economic evaluation methods (REEM) and guidance to better understand and evaluate the costs and consequences of complex interventions. It will bring together realist and economic evaluation to enable evaluators to establish what works, for whom, in which circumstances, whist integrating better understanding of costs and consequences (including opportunity cost). This research addresses the following questions in three phases:

- Phase 1 (Months 0-12): What are the theoretical, methodological and practical similarities and differences between realist and economic evaluations? This phase will use literature scoping, online discussion boards (including short activities), and facilitated virtual meetings to develop provisional REEM guidance for piloting.
- Phase 2 (Months 13-30 months): What lessons can we learn from using REEM in practice to improve it? This phase will pilot REEM and investigate its feasibility in research across three evaluations.
- Phase 3 (Months 31–36): How can we use empirical and expert knowledge to produce consensus REEM guidance? This phase will use an online Delphi method followed by a Consensus Development Conference to agree the core elements of REEM and produce finalised guidance.

The outputs from this research will be REEM guidance, a checklist, and summary notes for use by PPIE members. These outputs will allow REEM to be utilised by academics and scrutinised by research funders. Additional outputs include: peer reviewed academic journal articles, conference presentations, webinars for research funders, policy makers and commissioners. Short term impact will be facilitated through: stakeholder involvement from the outset of the research, development of international academic networks, and established international academic reach. Long term impact will include the use of REEM to: make better-informed commissioning and policy decisions; reduce research waste, and better target interventions to those that need them.

Background and rationale

Health and social care require well-designed complex interventions and robust evaluation, which itself is often complex [1-4]. Interventions are considered to be complex for several reasons including but not limited to: the number of components of the intervention, the expertise or skills required to deliver or receive the intervention, the scale of the intervention (numbers of groups or settings), and the flexibility of intervention [1]. Complex interventions thus have many implications for evaluation, not least, embracing and accounting for complexity. Realist evaluation was designed to evaluate complex social interventions [5-7]. Realist approaches that evaluate how and why interventions are effective for different groups and in different settings have increasingly been used by applied researchers and research funders [e.g., [8, 9]]. The National Institute for Health Research (NIHR) Health and Social Care Delivery Research (HSDR) Programme alone currently has 15 active or contracted projects using realist approaches. Concurrently, policymakers and research funders require economic appraisals to accompany evaluations of complex interventions to inform difficult decisions in the context of resource scarcity. However, realist evaluations do not tend to explicitly capture the economic consequences of interventions [7]. Additionally, economic evaluation methods often ignore context and do not capture variations in resource use or outcomes across groups, or recognise the implications this may have for the relationship between resource use and outcomes[10-12]. For example, not only is there likely to be variability in outcomes between groups and communities receiving the intervention, but the resources required to achieve those outcomes will vary. It is further argued that the results of economic evaluations in health and social care are often poorly generalisable (transferable) and should take better account of the role of context, population variations, and the conceptual challenges of evaluating complex interventions [13-17]. Yet the actual methods of economic evaluation have made very few advances in these directions. This is despite the potential for more useful, context-sensitive approximations of the ‘cost’ of complex interventions. This represents a

methodological gap in the evaluation of complex interventions and a key limitation to providing applicable evidence for policy and service delivery [1, 2, 16, 17].

The update of Medical Research Council (MRC) guidance outlining a new framework for developing and evaluating complex interventions states that: "*complex intervention research goes beyond asking whether an intervention works in the sense of achieving its intended outcome—to asking a broader range of questions (e.g. identifying what other impact it has, assessing its value relative to the resources required to deliver it, theorising how it works, taking account of how it interacts with the context in which it is implemented, how it contributes to system change, and how the evidence can be used to support real world decision making)*" [[1], pg.1]. The new framework outlines 6 core phases of evaluation to be considered to answer the following questions: 1) how does the intervention interact with its context? 2) what is the underpinning programme theory? 3) how can diverse stakeholder perspectives be included in the research? 4) what are the main uncertainties? 5) how can the intervention be refined? 6) do the effects of the intervention justify its cost? Drawing heavily on realist and economic evaluation methods, the updated MRC framework represents a significant step in bringing together these disciplines. Yet how to do this and overcome the barriers presented by differences in underpinning epistemology, ontology, and academic disciplinary roots (which prevent simply combining such methods) is unclear.

Currently, there is very little literature, theoretical or applied, that integrates realist and economic evaluation [13]. Recent examples are limited to guidance for synthesising realist and economic evidence in health and criminal justice (EEMIE) [18], the theoretical development of economic informed programme theories [19], and most recently, the development of economically optimised programme theories [20]. More commonly, where realist and economic evaluations are undertaken together, they are done so in parallel or sequentially, with each using narrow discipline-specific guidance (e.g. RAMESES II Quality Standards for Realist Evaluation [21], ISPOR Good Research Practices for Cost-Effectiveness Analysis [22]), resulting in incomplete knowledge for policy and decision makers and limiting opportunities to share valuable learning between the evaluations.

Despite their different (implicit or explicit) ontological and epistemological bases, there is considerable potential for realist and economic evaluations not only to learn from each other but to be merged. Our research will advance understanding of how these approaches can be integrated and provide added value to evaluators, decision-makers and funders by developing a form of evaluation that both enables economic evaluation to become more context-sensitive and explanatory, and realist evaluations to better capture the role of resources and the opportunity costs.

We will achieve this through the development of REEM, including methodological guidance to better evaluate and understand the costs and consequences of complex interventions. It will integrate the core elements of realist and economic approaches to enable evaluators to establish what works, for whom, in which circumstances, whist integrating better understanding of costs and consequences (including opportunity cost). It is important to acknowledge here that there is no one single way to undertake a realist evaluation and no single economic evaluation approach, and therefore, we do not anticipate agreeing a singular highly protocolised approach to REEM. Instead, REEM can be considered as augmenting theory of change and action (realist evaluation) on one hand and the theory of value creation (economic evaluation) on the other [23].

Ultimately, REEM will provide a useful way to evaluate complex interventions and ultimately provide policy makers and commissioners with the integrated information needed to make better decisions which are applicable in the 'real world'. This will enable better informed commissioning about complex health, social and civil society initiatives and will also provide

added value for funders who recognise the importance of producing research findings that are context-sensitive and cumulative.

Research Questions, Aims, and Objectives

Research Question 1: What are the theoretical, methodological and practical similarities and differences between realist and economic evaluations?
Aim 1: To understand and develop REEM, principles, and applications.

- **Objective 1**. To scope and map the evidence about current approaches and advances in realist and economic analysis in complex evaluations.
- **Objective 2.** To agree on a set of common definitions and principles for REEM and how they might be applied in practice.
- **Objective 3.** To use the outputs from Objectives 1 and 2 to develop provisional REEM guidance for piloting and an accompanying framework to evaluate the feasibility and value of using REEM in practice.

Research Question 2: What lessons can we learn from using REEM in practice to improve it?
Aim 2: To apply and evaluate the feasibility and value of using REEM in practice.

- **Objective 4.** To pilot the application of methods guidance developed in Aim 1, Objective 3 and evaluate the feasibility of applying REEM in practice (in terms of strengths, weaknesses, application, outcomes and value, to inform further refinement).

Research Question 3: How can we use empirical and expert knowledge to produce consensus REEM guidance?
Aim 3: To refine REEM principles and develop guidance for wider application and further development.

- **Objective 5.** To synthesise the findings from Aims 1 and 2 and agree the core elements of REEM guidance; integrate into existing guidance, quality and reporting standards for evaluation of complex interventions; and highlight further opportunities for development.

Overall Research plan (including data collection, data analysis and sampling)

SET-UP PHASE [3 Months -3 to 0]: We will establish an International Interdisciplinary Advisory Group (IIAG) to provide disciplinary and applied expertise to oversee the research. The group will also challenge the research team, ensuring rigour and that the methodological developments are consistent with realist and economic principles held by other scholars beyond the project team, as well as with the policy and funding context. Membership of the IIAG will include a mix of internationally renowned academic experts in either health economic or realist evaluation, policy-makers, and research funders. As we progress the research, we will seek to expand membership of the IIAG to ensure representation in terms of method and geographical location, including adding partners from the pilot evaluation sites in Phase 2. In addition, we have PPIE embedded throughout all the phases, further details of which are outlined in the PPIE strategy.

PHASE 1: TO UNDERSTAND AND DEVELOP REEM, PRINCIPLES, AND APPLICATIONS [Months 0 to 12]. Addresses Research Question 1, Aim 1 (Objectives 1 to 3).

Objective 1: To scope and map the evidence about current approaches and advances in realist research and economic analysis when evaluating complex interventions.

Data Collection. We will undertake a scoping review to: identify the available evidence, gaps in knowledge, clarify key definitions, and examine *what* the theoretical and methodological barriers and facilitators to integrating these methods are. This will include finding examples of *how* programme theory has been applied in economic evaluations and *how* costs/resource

use and cost-effectiveness have been captured in realist evaluations [24]. The searches will gather guidance documents and studies to address four sub-questions:

a) What are the recent developments in methods/guidance recommended for economic evaluation of complex interventions in health and social care?
b) How are realist concepts i.e., programme or intervention theory or context-dependency (a description of the causal association between programme components, resources and outcomes) captured theoretically or applied in the conduct of economic evaluations, including any examples that have demonstrated this?
c) What are the recent developments/guidance recommended for realist evaluations of complex health and social care interventions?
d) How are economic concepts i.e., resource use/impacts, outcome valuation and opportunity costs of interventions, captured theoretically or applied in the conduct of realist evaluations including any examples which have demonstrated this?

We are aware that a modest amount of relevant methods guidance exists for realist evaluations and for economic evaluation of complex interventions (Questions (a) and (c)). Conventional literature searches for documents mentioning 'methods' would retrieve a large amount of mostly irrelevant results. We will therefore use search methods recommended for conducting reviews of methods papers [25, 26] to gather methods documents efficiently for Questions (a) and (c). First, we will gather key guidance, methods papers and chapters from our IIAG and project team members. We will identify further relevant sources using forward and backwards citation search techniques using Science and Social Science Citation Indexes (Web of Science) and Google Scholar. Finally, a focussed literature search of databases (Assia, Medline, EconLit, Web of Science databases) and a Google search will supplement our collection of methods guidance. Searches for Question (b) will aim to identify published and unpublished studies that have attempted to use realist concepts or programme theory within economic evaluation. The databases and sources will be the same as those outlined above. The initial search strategy will include the search terms 'programme theory', 'causal mechanisms' and 'intervention theory', combined with a purposive search for economic evaluations. Searches for Question (d) will identify published and unpublished realist evaluations that capture costs or resource use. Searches will run in the sources listed above, using the search term 'realist evaluation' and words, synonyms and index terms for 'costs' or 'resource use'. This approach ensures we will draw on research, experience, and knowledge across multiple disciplines and countries (published and grey literature sources, ongoing projects, and training materials). Searches will be peer-reviewed by an independent information specialist. For all searches, an iterative approach will be used; features and key words from initially included studies will be used to re-seed searches for relevant studies.

Draft searches developed as part of this proposal are available in the additional files. They indicate a manageable workload for the scoping review, generating the following approximate number of abstracts: A (300), B (1000), C (200), and D (300). Additional references will be found through contact with the IIAG and citation searching. Given the overlap in search terms used for each question we anticipate finding considerable duplicate records. EndNote software will be used to manage and remove duplicates. Studies will be coded for which question(s) they relate to, noting that some studies/reports will be relevant to more than one review. The reviews will be led by RAn and JW who bring expertise in reviewing and literature searching in economic and realist evaluations, plus methodology reviews. The review protocol will be registered with the Open Science Framework and the reporting will follow PRISMA ScR guidelines [27].

Summary and synthesis. There will be no formal or standardised quality assessment; instead, each source will be judged in terms of their clarity of reporting (especially methods/recommendations) and its contribution to the emerging synthesis. This will involve tabulating the main characteristics of included papers/sources, grouping them, identifying outliers/discrepant points/evidence, and discussing them initially within the research team.

We will use the findings from the scoping review to produce four briefing papers, addressing questions (a) to (d) posed above. These briefing papers will be presented for debate in the online discussion board and meetings described below (Objective 2).

Objective 2: To agree on a set of common definitions and principles for REEM and how they might be applied in practice.

Data collection. The briefing papers produced in Objective 1 will be presented to the wider research team and the IIAG and form the basis for discussion and debate on the following topics:

a) How epistemological and ontological commonalities and differences between realist and economic evaluation undermine or support and shape the theoretical development of REEM;
b) How realist and economic approaches/methods are applied in evaluations (particularly of complex interventions and systems) and how they can be combined pragmatically in the development of REEM. This will include critiquing previous/ongoing attempts.
c) The language (semantics and terminology) to be used in the description and definition of REEM.

Discussion and debate will be facilitated through:

a) An online moderated discussion board hosted via Microsoft Teams. The functionality of Microsoft Teams matches the demands of the research (i.e., sharing of files, images and links) and all members of the research team and IIAG are familiar and have prior experience of using it. PPIE members will be given training where necessary.
b) Short activities (e.g., white board think aloud exercises, JamBoard, research critiques, etc.) will be posted on the online discussion board. An in-depth review and critique of any particularly significant papers identified through the scoping review will be conducted.
c) Three facilitated recorded virtual meetings (max 2 hours each) to allow further/wider discussion and elaboration on the discussion board topics and facilitated short activities.

Data analysis. Thematic synthesis [28] will be used to further understand and analyse all qualitative data sources (online moderated discussion board, facilitated short activities, facilitated recorded virtual meetings). This approach is usually used to synthesise published primary research studies, however it will help us to understand findings across several data sources in this research phase, as outlined here. Furthermore, thematic synthesis was developed out of a need to conduct reviews of qualitative research that addressed questions relating to need, appropriateness and acceptability, as well as those relating to effectiveness [29]. Thematic synthesis has three stages: the coding of text 'line-by-line'; the development of 'descriptive themes'; and the generation of 'analytical themes'. The analytical themes represent a stage of interpretation whereby the reviewers (AB, SD and SRA) 'go beyond' the primary studies and generate new interpretive constructs, explanations or hypotheses. This approach will allow us to synthesise learning across all data sources to draw together common operational definitions of and principles for applying REEM.

Sampling. A purposive sample consisting of all members of the research team, the IIAG and PPIE members will be invited to participate at all stages of data collection and to reflect on the analysis.

Objective 3: To develop provisional REEM guidance for piloting and an accompanying framework to evaluate the feasibility and value of using REEM in practice.

Data collection. The results of the thematic synthesis will be presented in a 2-day workshop at the end of Year 1/start of Year 2. Participants at the workshop will be asked to reflect on these findings to:

a) finalise the common operational definitions and principles of what REEM are.

b) guide the design of provisional REEM guidance for piloting and agree strategies to evaluate the feasibility of using REEM in practice (in Phase 2).

Data Analysis. The intention at this phase is to collate all the views of all participants, rather than to try to reach consensus. We anticipate that there will be disagreements between participants but contend that consensus at this point is not necessarily useful; if we can identify where there is agreement and disagreement, identify causes and the nature of disagreements, we can propose responses to these (e.g., more than one definition, principle or strategy; piloting multiple strategies of REEM within pilot evaluations, etc.).

Sampling. A purposive sample consisting of all members of the research team, the IIAG and PPIE members will be invited to participate in the workshop. To reduce unnecessary travel and costs, the workshop will be held alongside international conferences that participants anticipate attending. Virtual attendance will also be offered as an alternative.

Phase 1 output. The results from each of the objectives will be written into discussion papers and peer-reviewed publications (co-authored by all participants) where appropriate. Furthermore, from these results, we will produce provisional guidance on how realist and economic evaluation methods can be integrated; common operational definitions and principles; and how these can be applied in practice. These will be tested in Phase 2.

PHASE 2: TO APPLY AND EVALUATE REEM, AND UNDERLYING PRINCIPLES IN PRACTICE [18 Months 13 to 30]. Addresses Research Question 2, Aim 2 (Objective 4).

Objective 4. To pilot the application of methods guidance developed Phase 1 and evaluate the feasibility of applying REEM in practice (in terms of strengths, weaknesses, application, outcomes and value, to inform further refinement).

In this Phase we will conduct three parallel pilot evaluations (detailed below) to test the feasibility and value of applying the provisional guidance of REEM (developed in Phase 1), to inform its refinement. These pilots have been selected on the basis that they reflect a range of health and social care interventions and varied geographical locations. Hence the resulting evaluations will test REEM across this range of interventions and sites. The evaluations will be running in the proposed research time frame and there are ongoing relationships with the evaluation partners. The pilot evaluations are:

- **An NHS prehabilitation programme (Waiting Well) for patients undergoing surgery (Led by Northumbria University in partnership with South Tees NHS Trust, Teesside. SD, AB, AF).** Waiting Well is a service delivered by South Tees Hospitals and Public Health South Tees which is designed to support patients in improving their fitness, health and wellbeing before a planned surgical operation or treatment.
- **An e-health maternity screening programme (C-it Du-it) in a low-income country context (in partnership with the Kenya Medical Research Institute (KEMRI), Nairobi. MK, SR, GWe).** C-it Du-it aims to improve access to antenatal sonography for low-income women in Nairobi and is an important part of the national strategy more widely in Kenya to reduce maternal and neonatal mortality.
- **A community enterprise café aimed at reducing social isolation and loneliness among Housing Association service users. (Led by the Yunus Centre at GCU in partnership with ImpactArts, Edinburgh. RB, CD, GWo).**
  This pilot evaluation will focus on Craft Café workshops. The Cafés are a safe, welcoming spaces for people to socialise and express their creativity, bridging the gap between care and housing support, to reduce social isolation and loneliness, and increase wellbeing.

Letters from the pilot evaluation site partners detailing their support are attached. Pilot evaluation partners will provide support to the research including: staff time to attend meetings related to the evaluations and research interviews, facilitating access to internal data,

brokering initial contact and access to services and individuals between the research team and the interventions being evaluated, and hosting or providing access to facilities for the Research Assistant (RA) where required. A RA will be recruited to each pilot site for the duration of each pilot evaluation. In addition, the Senior Research Assistant (SRA) based at Northumbria University will take a leading role in working closely with the other RAs as a point of contact throughout the evaluations, ensuring complementarity and optimising opportunities for learning across pilot sites and synthesising learning at the end. We will attempt to recruit a range of skills across the RAs and will provide additional methodological training (costed).

Data collection and analysis. REEM will be applied in each of the pilot evaluation sites using the provisional guidance developed in Phase 1, thus we cannot specify REEM in detail now. However, data collection in REEM will be driven by the realist programme theories generated in each pilot evaluation. Initial programme theories (IPTs) will drive subsequent data collection and analysis, as is usual practice in a realist evaluation [5]. Therefore, whilst it is not possible at this stage to state what data we will need to collect and thus how we will analyse it, we will draw on established methods of data collection and analysis commonly utilised in realist and economic evaluations (specified below). Primary and secondary data collection methods will be used generate qualitative and quantitative data to identify and test the causal mechanisms, contexts, and outcomes (CMOs) and the societal costs and consequences (positive and negative, intended and unintended) associated with the intervention in question. These methods are outlined in the following iterative and cyclical steps which incorporate the six core elements of the new MRC framework for developing and evaluating complex interventions [1]:

- **Step 1. Identify and define the scope/boundaries of the evaluation and stakeholders.**
  Data collection methods and data sources: literature scoping, documentary review and analysis, realist theory gleaning interviews, and participant observation.
- **Step 2. Develop initial programme theories identifying economic costs and consequences.**
  Data collection methods and data sources: stakeholder workshops.
- **Step 3. Test initial programme theories, measuring and valuing economic costs and consequences.**
  Data collection methods and data sources: realist refining interviews, routine health and social care data, outcome measurement and valuation, economic modelling.
- **Step 4. Refine REEM programme theories**
  Data collection methods and data sources: realist theory consolidation interviews, sensitivity analysis.

The specific methods of data collection and analysis adopted within the steps above will be decided in light of Phase 1 findings but our preliminary approaches are outlined below.

a) Realist interviews: (n=20 in each pilot evaluation). Realist theory gleaning, refining and consolidation interviews [30] with stakeholders (programme architects and/or implementers and those receiving the programme) will be used to provide a nuanced understanding of the generative mechanisms and associated contexts leading to outcomes (positive, negative, intended, and unintended). They will also allow elicitation of information about resource use and available cost data. We will adopt a realist purposive sampling strategy, which is determined through the programme theories to be investigated [13] for identifying programme theories. We will adhere to RAMESES II quality and reporting standards [21].

b) Stakeholder workshops: Two workshops will be held with the identified stakeholders. Workshop 1 will allow informal consultation with stakeholders (programme architects and implementers [30, 31]) to develop initial programme theories (IPTs). Initial explorations of the economic costs and consequences will be considered alongside the development of IPTs and will be further nuanced in Workshop 2. Though it is not possible at this stage to state what the costs and consequences will be (as they will be bound up in the development of the IPTs), in Workshop 2 we will seek to identify: direct and indirect costs

and consequences, relevant data sources, gaps in data, and suitable approaches to valuation of consequences and outcomes, etc.

c) **Realist Analysis:** Realist interviews [30] and stakeholder workshops will be transcribed verbatim and imported into NVivo. All qualitative data will be analysed in Nvivo using a realist CMO lens [32]. Analysis will move iteratively between analysis of particular examples, refinement of programme theory, and further iterative searching for data to test [33]. Throughout the data collection period, the pilot evaluation teams will partake in a deeply reflexive, iterative data analysis approach, to spark insight and develop meaning [34]. It consists of multiple rounds, revisiting the data as new additional questions emerge and connections are established, thus deepening the understanding and meaning of the findings

d) **Administrative data:** Pilot site administrative data sets will be used to extract both cost and outcome data. Participant records from the pilot evaluations will be accessed through the pilot evaluation sites and pseudonymised data will be transferred to the research team in Microsoft Excel. Data will be analysed using SPSS or Stata, as appropriate.

e) **Routine health and social care data:** Routine health and social care data sources (such as Hospital Episode Statistics, Secondary Uses Service, NHS reference costs, and unit costs of health and social care (PSSRU)) will be used to analyse cost and outcome data, where available and necessary. Data will be extracted from routine data sources using standardised codes. Version control will be applied to the datasets.

f) **Participant surveys or proforma:** will be used to supplement cost and outcome data collection. This may include: validated measures of health, quality of life, wellbeing, or capabilities (where relevant), standardised participant costing proforma, and in(direct) outcome valuation methods such as stated preference measures (e.g. discrete choice experiments, willingness to pay).

g) **Economic modelling:** Economic models may be used in this phase to depict the complexity of the intervention, using insights drawn from the realist analysis to define and compare the model pathways and parameters and capture the associated costs and consequences. The choice of model will vary between the evaluation sites and will depend on the final evaluation question and data generated earlier stages of the evaluation. Examples of cohort or individual-based models more suited to complex interventions include system dynamic models, agent-based models, and social network models [35].

The pilot evaluation partners will work closely with the research team and have costed for their time to do so. The project partners will act as the gatekeepers to recruit participants for the evaluation. They will distribute the participant information and gain consent. The contact details of the RA at each of the pilot evaluation sites will be shared and participants will opt in.

The pilot evaluation partners will receive a full final evaluation report and be offered co-authorship on subsequent outputs. Participants will be asked if and how they would like to be informed of the evaluation findings (full report, lay summary), provided via email or hardcopy.

Phase 2 output: Lessons on the application of REEM from the pilot evaluation sites will be brought together in a discussion paper and used to update the provisional REEM guidance in line with these lessons. This will be circulated to the IIAG and used as a basis to develop methodological guidance in Phase 3. In addition, the pilot evaluation reports will be shared with project partners and published in peer reviewed journals.

PHASE 3: TO REFINE REEM AND PRINCIPLES, AND DEVELOP GUIDANCE FOR WIDER APPLICATION AND FURTHER DEVELOPMENT [6 Months 30 to 36]. Addresses research question 3, Aim 3, Objective 5.

Objective 5: To synthesise the findings from Aims 1 and 2 and agree the core elements of REEM guidance; integrate into existing guidance, quality and reporting standards

for evaluation of complex interventions; and highlight further opportunities for development.

Data Collection. We will use an online Delphi method followed by a Consensus Development Conference (CDC) [36] to synthesise the outputs from Phases 1 and 2 and agree the core requirements for integrating and applying REEM in practice. We reviewed several deliberative consensus development methods including the Delphi Method, the RAND/UCL appropriateness method, Nominal Group Techniques and the CDC. Due to the infancy of REEM and the two different approaches of realist and economic evaluators, we believe that it will be more conducive to have complex discussions and agree consensus in person and thus propose the addition of the CDC, where iterative feedback is generated through several rounds of group discussion, which will build on the Delphi.

We will use an online Delphi method which has previously been successfully used to develop quality and reporting standards in the RAMESES II [21]. The Delphi panel will be run online in 2 rounds (or more if required) using Survey Jisc or similar. In round 1, panel participants will be provided with briefing materials including the updated version of the provisional REEM guidance from Phase 2 and invited to rate the importance of each updated item in the REEM guidance. Participants will also be given opportunities to provide additional suggestions for revisions or new items which will form the basis of the beginning of discussions in the CDC (below). Responses will be analysed (as below) and fed into the design of questionnaire items for round 2. In round 2, participants will be asked to rank each potential item twice on a Likert scale (strongly disagree to strongly agree), to agree 1) which are relevant (i.e., should an item on this theme/topic be included at all in the guidance?), and 2) item definitions (i.e., to what extent do you agree with the specific wording for the guidance?). Those who agree that an item is relevant, but disagree with the definition/wording, will be invited to suggest changes via a free-text comments box. In this second round, participants will again be invited to suggest additional topic areas and items which will be further discussed in the CDC.

The online Delphi will be followed by a CDC. The CDC is a rapid data synthesis method used for the collation of balanced advice about a technology or approach, and for the definition of the need for further information and research [37]. The aim of the CDC will be to achieve consensus amongst members of the CDC, who come from different epistemological backgrounds. We believe that achieving consensus will benefit from in-person discussion and have costed for this to be conducted as a 2-day face-to-face meeting. To reduce unnecessary travel and costs, the CDC will be held alongside international conferences that members anticipate attending. However, it can also be delivered entirely virtually using online consensus methods if required. Following analysis, the results of the Delphi will be fed into the CDC for further discussion with the project team and IIAG. The CDC will involve reviewing the results of the Delphi and further rounds of moderated discussion for items where consensus was not reached through the Delphi. AB and SD, along with the SRA, will chair and facilitate the CDC using audio recording alongside extensive notes and live editing of the REEM guidance on a large screen for members to view, consider and input further. The CDC will use reflective practice [38] drawing on the Theory of Consensus [39] to consider REEM guidance and engage in a process of continuous learning from one another to reach consensus. Throughout the CDC, SD and AB will articulate proposals related to specific parts of REEM guidance and facilitate the 'process model of consensus building' outlined in Figure 2. Use of this model will allow participants to reflect on their experience and expertise, drawing in values and theories which inform their approach, to lead towards informed consensus (or disagreement) for REEM guidance. Using reflective practice and the process model of consensus building AB and SD will draw together arguments, synthesise and reflect these back to participants, highlighting potential consensus and disagreement. Using a cyclical reflection process will allow participants' arguments to be conveyed back to them to move towards consensus. We plan to report residual non-consensus 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 those who use REEM.

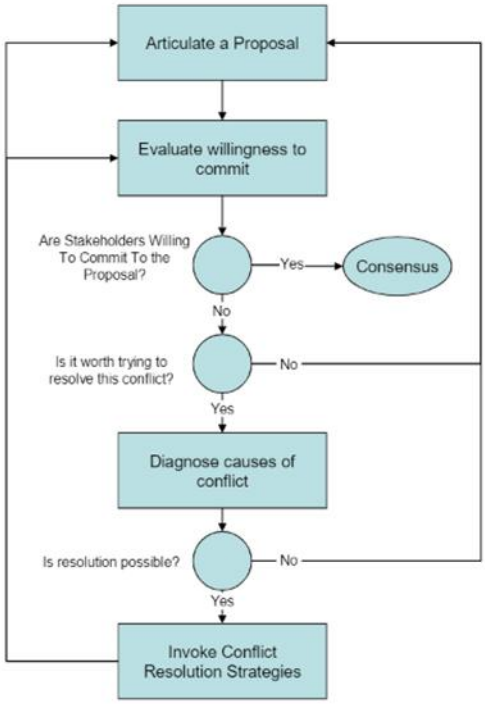

Figure 2. A process model of consensus building to be used in CDC (a rectangle indicates an activity, a circle a decision).

Data Analysis. For the online Delphi, each participant's responses will be collated and the numerical rankings entered onto an Excel spreadsheet. The response rate, average, mode, median and IQR for each participant's response to each item will be calculated. Items that score low on relevance will be omitted from subsequent rounds. We will invite further online discussion on items that score high on relevance but low on validity (indicating that a rephrased version of the item may be needed) and on those where there was wide disagreement about relevance or validity. The panel members' free text comments will also be collated and analysed thematically. Consensus will be considered reached in each Delphi round if at least 70% of the participants strongly agree/disagree [40]. The results of the vote will be written into any subsequent publications in order to show transparency in agreement

(or disagreement). Where agreement is not reached or there are suggestions for altered or new items, these will be debated further in the CDC. Where disagreement is high then this will be detailed thoroughly in subsequent reports and publications [37].

The final template will be agreed within the CDC but also distributed to participants for approval after the CDC, allowing for time for further reflection.

Sampling. The Delphi panel will include members of the research team, the IIAG and participation will be widened to include a further 20 people identified as key academics from papers identified in the scoping synthesis, (a minimum of 40 potential participants in total). The CDC participants will include a purposive sample of the research team, members of the IIAG, project partners and PPIE members who will all be invited to join.

Phase 3 output. The agreed and finalised guidance will outline the definitions, principles, and methodological quality of REEM with respect to a) planning, b) conducting, and c) reporting. We recognise that any REEM guidance should facilitate fidelity in the application of REEM, whilst also allowing diversity in methods and the continuation of its development, rather than imposing rigid methods. To ensure that REEM are subsequently used in practice, we will compile a REEM checklist similar to the updated CHEERS II Statement [41] and produce summary notes for PPIE members. We will share the findings with research funders, policymakers, and academics (see dissemination strategy). Furthermore, we will ensure that the guidance is integrated in line with further developments of the RAMESES project [21], MRC Complex Intervention Guidance [1], and CHEERS II Statement [41].

Dissemination Strategy and Outputs

We have identified our primary audience (academics, research funders, policymakers and commissioners) prior to attaining funding and invited them to input into the research via the IIAG. This means that they will be involved from the planning of the study through to dissemination, with regular communication and input (online discussion boards, regular emails, virtual meetings) throughout the research, across all three phases.

A range of targeted outputs will be produced to correspond with the identified research users:

- The primary outputs will include the REEM guidance, checklist and the summary notes for use by PPIE members. These outputs will allow REEM to be utilised by academics and scrutinised by research funders. A webinar will be hosted for research funders, policymakers and commissioners (beyond those involved in the IIAG) to promote the REEM Guidance, check list and PPIE guidance notes. Funders with whom the research team have worked with (including: MRC, ESRC, NIHR, CSO, The Health Foundation) will be invited, amongst others (such as the Association of Medical Research Charities) and those in local health and social care commissioning (networks contacted through the IIAG).
- We will ensure that the guidance is integrated in line with further developments of the RAMESES project [21], MRC Complex Intervention Guidance [1], and CHEERS II Statement [41].
- PPIE guidance notes will be co-produced with PPIE members and shared in a webinar via PPIE networks (e.g. the ARC NENC Public Advisory Network and the Fuse (Centre for Translational Research in Public Health) Public Involvement and Engagement Committee.
- In addition to the final report, the outputs from all three Phases will be published in leading peer-reviewed journals in health economics and evaluation, notably Value in Health and Social Science and Medicine.
- Abstracts will be submitted to national and international conferences in realist and health economic methods (such as International Realist Research, International Health Economic Association, and Health Economists Study Group). In addition, REEM workshops at these conferences will be scheduled within 2 years of the conclusion of the study.

- We will develop and maintain a research website and host an online seminar to disseminate the results and promote the outputs listed in this section. Social media channels including Twitter, and the active international RAMESES JISCmail and Health Economics distribution lists, will be used to promote the research and outputs.
- The research team will seek further funding to test the principles of REEM in future empirical work outside of health and social care and further build and develop our relationships and activities with key stakeholders (e.g. policymakers) to ensure REEM is used, useful and useable.

The research team has established and strong links with several international realist and economic research groups important disseminating this research including the RAMESES project and RAMESES JISCMail listserv which has over 1296 realist researchers world-wide, The Realist Research Evaluation and Learning Initiative at Charles Darwin University, Northern Realist Research Team Hub (NoRTH), Fuse (The Centre for Translational Research in Public Health), a UK Clinical Research Collaboration (UKCRC) Centre of Public Health Research Excellence, ARC NENC, Health Economists' Study Group and International Health Economics Association. In addition, IIAG members are in a position to promote and disseminate REEM through established networks, with discussion papers being developed into workshops, seminars and publications where appropriate.

Project management and timetable

SD and AB will jointly manage the project to ensure sufficient representation of realist and economic expertise. This will be achieved through 2 weekly meetings with the SRA, and 2 monthly co-applicant meetings. Regular monthly mentoring meetings will also take place with CD and Gwo throughout the research, with capacity to add ad-hoc meetings where necessary. In Phase 2, SD and AB will also meet with the pilot evaluation RAs and leads (MK and RB) once every two weeks, to ensure progress and capture feedback. A detailed timetable is outlined below:

Timetable (36 Months. Start 01 September 2022).

Month	Activities	Events
-3-0	Set-up Phase. Convene IIAG and PPIE members. Seek ethical approvals. Recruit SRA. Submit study protocol.	Research Team meeting
1-5	Phase 1. Objective 1. Conduct literature scoping and narrative synthesis analysis. Findings to be written into briefing papers.	Launch Meet (M1)
6-11	Phase 1. Objective 2. Share briefing papers, conduct short online activities (JamBoards, reviews of papers etc) and host 3 virtual meetings with research team, IIAG, and PPIE members. Synthesise learning across all data sources to draw together common operational definitions of and principles for applying REEM.	3 virtual meetings (M6,8,10)
10-12	Phase 2. Objective 4. Set-up of pilot evaluation sites with project partners. Recruit RAs. Seek ethical approvals.	
12	Phase 1. Objective 3. Host 2-day workshop with research team, IIAG, and PPIE members to finalise the common operational definitions and principles of REEM and agree strategies to evaluate the feasibility of using REEM in practice (in Phase 2). Produce provisional guidance for applying REEM.	2-day IIAG workshop
13-15	Phase 2. Objective 4. Evaluation Step 1. Identify and define scope/boundaries of the evaluation through literature scoping, documentary review and analysis, realist theory gleaning interviews, and participant observation.	Training for RAs. Virtual meetings with project partners

16-18	Phase 2. Objective 4. Evaluation Step 2. Develop programme theory and identify possible economics costs and consequences through stakeholder workshops	
19-24	Phase 2. Objective 4. Evaluation Step 3. Test programme theories and measure and value the economic costs and consequences through realist theory refining interviews, routine health and social care data, direct and indirect valuation methods, validated outcome measures, economic modelling.	
25-27	Phase 2. Objective 4. Evaluation Step 4. Refine programme theory through realist theory consolidation interviews.	
27-30	Draw together findings and share with project partners. Produce discussion paper regarding the application of REEM and update provisional guidance in line with the lessons learnt from the pilot evaluations.	
31	Phase 3. Objective 5. Circulate the updated provisional REEM guidance	
32	Phase 3. Objective 5. Refine and develop guidance 2-day CDC and Delphi	2-day CDC
33	Phase 3. Objective 5. Circulate final guidance to IIAG and research team for final reflection	
34-36	Finalise framework and guidance; produce final report for HSDR.	

Ethics

Ethical approval will be sought from Northumbria University Ethical Approval System for the full project, excluding Pilot Evaluations 2 and 3 (Phase 2). Ethical approval for the Pilots will be sought from their respective institutional ethical review panels, namely: Strathmore University Institutional Review Board and Glasgow Caledonian University School of Health and Life Sciences Ethics Panel, and the Health Research Authority Research Ethics Service for data collection involving NHS staff or patient data where applicable. All research and the pilot evaluations will be conducted in line with ethical principles of anonymity and confidentiality, and all participants will have to provide informed consent to participate.

References

1. Skivington, K., et al., A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ, 2021. 374(2061).
2. Shiell, A., P. Hawe, and L. Gold, Complex interventions or complex systems? Implications for health economic evaluation. BMJ,, 2008: p. 1281-83.
3. Petticrew, M., Public Health Evaluation: epistemological challenges to evidence production and use. Evidence and Policy, 2013: p. 87-95.
4. Rutter, H., The need for a complex systems model of evidence for public health. The Lancet, 2017: p. 2602-2604.
5. Pawson, R. and N. Tilley, Realistic Evaluation. 1997, London: Sage.
6. Fletcher, A., et al., Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. Evaluation, 2016: p. 286-303.
7. Dalkin, S., et al., What's in a mechanism? Development of a key concept in realist evaluation. Implementaion Science, 2015. 10(49).
8. Fisher, R., et al., Large-scale implementation of stroke early supported discharge: the WISE realist mixed-methods study. Health Services and Delivery Research 2021. 9(22).
9. Cornes, M., et al., Improving care transfers for homeless patients after hospital discharge: a realist evaluation. Health Services and Delivery Research, 2021. 9(17).
10. Anderson, R., Systematic reviews of economic evaluations: utility or futility,. Health Economics, 2010: p. 350-364.
11. Culyer, A., Cost, context, and decisions in health economics and health technology assessment. 2018: p. 434-441.
12. Franklin, M., J. Lomas, and G. Richardson, Conducting Value for Money Analyses for Non-randomised Interventional Studies Including Service Evaluations: An Educational Review with Recommendations. Pharmacoeconomics, 2020.
13. Anderson, R., et al., Using realist approaches to explain the costs and cost-effectiveness of programmes.,, in Doing Realist Research, J.G. N. Emmel, A. Manzano, M. Monaghan and S. Dalkin, Editor. 2018, Sage: London.
14. Anderson, R. and I. Shemilt, The Role of Economic Perspectives and Evidence in Systematic Review. Medicine, 2010.
15. Normand, C., Measuring Outcomes in Palliative Care: Limitations of QALYs and the Road to PalYs. 38, 2009. 1.
16. Anderson, R., New MRC guidance on evaluating complex interventions: clarifying what interventions work by researching how and why they are effective. BMJ, 2008. 1937.
17. Greenhalgh, T. and C. Papoutsi, Studying complexity in health services reserach: desperately seeking an overdue paradigm shift BMC Medical, 2018. 16(95).
18. Johnson, S., N. Tilley, and K. Bowers, Introducing EMMIE: an evidence rating scale to encourage mixed-method crime prevention synthesis reviews. . Journal of Experimental Criminology, 2015. 11(3): p. 459-73.
19. Anderson, R. and R. Hardwick, Realism and resources: towards more explanatory economic evaluation. Evaluation, 2016: p. 323-341.
20. Rees, C., et al., Unpacking economic programme theory for supervision training: Preliminary steps towards realist economic evaluation. Medical Education, 2021.
21. Wong, G., et al., Quality and reporting standards, resources, training materials and information for realist evaluation: the RAMESES II project, in Health Services and Delivery Research. 2017.
22. Ramsey, S., R. Willke, and H. Glick, Cost-effectiveness analysis alongside clinical trials II: an ISPOR Good Research Practices Task Force Report. . Value Health, 2015. 18(2): p. 161-172.
23. King, J., Using Economic Methods Evaluatively. American Journal of Evaluation, 2016. 38(1): p. 101-113.

24. Cookson, R., et al., Using Cost-Effectivness Analysis to Address Health Equity Concerns Value in Health, 2017. 20(2): p. 206-212.
25. Gentles, S., et al., Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research. Systematic Reviews, 2016. 5.
26. Martin, G., et al., Toward a framework for the design, implementation, and reporting of methodology scoping reviews. Journal of Clinical Epidemiology., 2020(127): p. 191-7.
27. Tricco, A., et al., PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Research and Reporting Methods, 2018.
28. Thomas, J. and A. Harden, Methods for the thematic synthesis of qualitative reserach in systematic reviews BMC Medical Research Methodology, 2008. 8(45).
29. Barnett-Page, E. and J. Thomas, Methods for the synthesis of qualitative reserach: a critical reveiw BMC Medical Research Methodology, 2009. 9(59).
30. Manzano, A., The craft of interviewing in realist evaluation. Evaluation, 2016: p. 342-360.
31. Mukumbang, F., et al., Using the realist interview approach to maintain theoretical awareness in realist studies. Qualitative Research, 2019. 20(4): p. 485-515.
32. Jackson, S. and G. Kolla, A New Realistic Evaluation Analysis Method: Linked Coding of Context, Mechanism, and Outcome Relationships. American Journal of Evaluation, 2012. 33(3).
33. Wong, G., et al., Interventions to improve antimicrobial prescribing of doctors in training: the IMPACT (IMProving Antimicrobial presCribing of doctors in Training) realist review. . BMJ Open, 2015. 5(10).
34. Jagsoh, J., Retroductive theorizing in Pawson and Tilley's applied scientific realism. Journal of Critical Realism, 2020.
35. Hoang, V., et al., A systematic review of modelling approaches in economic evaluations of health interventions for drug and alcohol problems. BMC Health Services Research 2016. 16(127).
36. Ives, J., et al., Standards of practice in empirical bioethics research: towards a consensus. BMC Medical Ethics, 2018. 19(68).
37. McGlynn, E., J. Kosecoff, and R. Brook, Format and Conduct of Consensus Development Conferences: A Multination Comparison, in Improving Consensus Development for Health Technology Assessment: An International Perspective., Institute of Medicine (US) Council on Health Care Technology, C. Goodman, and S. Baratz, Editors. 1990, National Academic Press (US): Washington DC.
38. Bolton, G., Reflective Practice: Writing and Professional Development. 3 ed. 2010, London: SAGE Publications.
39. Briggs, R., G. de Vreede, and G. Kolfschoten. Toward a Theoretical Model of Consensus Building. in Proceedings of the Eleventh Americas Conference on Information Systems. 2005. Omaha, USA.
40. Slade, S., et al., Standardised method for reporting exercise programmes: protocol for a modified Delphi study. BMJ Open, 2014. 4.
41. Husereau, D., et al., CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. Value Health, 2022. 25(1): p. 3-9.