RAMESES II: Realising the potential of realist research for improving the delivery of health services

Research summary

This study has three linked workstreams:

Workstream 1: Develop quality standards and publication guidance for realist evaluation

With expert librarian help, we will identify reviews, scholarly commentaries, models of good practice and examples of (alleged) misapplication of realist evaluation. We will summarise a) current best practice, b) what experts and other researchers believe count as high quality research and what aspects of a study they believe needs to be reported; and c) what issues researchers struggle with. The last of these will be identified both from the literature and from a thematic analysis of the RAMESES JISCmail list archive). The purpose of this step is not to produce an encyclopaedic summary of all realist evaluations (or commentaries on this method) ever undertaken but to prepare briefing materials for the Delphi panel.

We will recruit approximately 35 panellists from academia, policymaking, research funders, research publishers and research users (including patient organisations). As in our previous NIHR-finded RAMESES I study, we will do this online using SurveyMonkey [9]. In round 1, we will send out briefing materials and invite suggestions to inform item construction for a Delphi survey.

Responses will be analysed and fed into the design of questionnaire items for round 2. Panel members will be invited to rank items for [a] relevance (should we include this topic/theme?) and [b] content (should we word it like this?). High scores for relevance *and* content will mean the item will be included 'as is'. High scores for relevance but low scores for content will mean we need to word the item differently (we will ask for suggestions). Low scores for relevance mean the item will be dropped. We will continue until no further move towards consensus is possible.

As well as writing up the standards and guidance for publication in peer-reviewed journals, we will invite our patient and public panel to help us develop of a lay summary of these outputs.

Workstream 2: Provide bespoke support for realist evaluations

As in RAMESES I, the Delphi panel will be fed not only 'clean' ideas and perspectives on realist evaluation from the scholarly literature but also 'messy' ones from ongoing evaluations [9]. To sample up to ten realist evaluations that unfold in parallel with our Delphi exercise, we will a) ask NIHR to link us with planned evaluations funded by them that align with our own timeline; b) ask on the RAMESES JISCmail list; c) capture unsolicited requests for help (of which we receive many). We will aim for maximum variety in experience of research teams, topics, settings and approach to patient and public involvement. We will work flexibly with teams, mostly by phone, Skype and email, to support them with methodological advice and troubleshooting. We will systematically capture the questions and issues from these ten primary studies and feed them into the deliberations of the Delphi panel (where timings permit) and also into the design of resources, training materials and courses in workstream 3.

Workstream 3: Develop training resources and train the trainers

Building on our background in research design and methodological support (e.g. the principal applicant, TG, was previously Co-Director of the East London branch of the Research Design Service), we will seek examples of the kinds of requests made by aspiring researchers for methodological support and the level of confidence that designers and trainers feel in dealing with them. For this, we will draw on our extensive contacts with trainers in realist research, our empirical data from the RAMESES list archive and the outputs of workstream 2 on the real-world struggles of realist researchers, plus our literature review and Delphi panel. We will synthesise these to inform development of training materials and resources (see example from RAMESES I: http://www.ramesesproject.org/media/Realist_reviews_training_materials.pdf).

We will run 3 x 2-day 'how to do a realist evaluation' workshops and 3 x 2-day 'training the trainers' workshops (one of each in London, Leeds and Liverpool). The workshops will welcome staff from the Research Design Service who seek to support realist studies, as well as staff from other research support organisations and researchers, practitioners, policymakers and patient advocates. The detailed

curriculum for the training the trainers workshops will emerge from our empirical work, but will include all the steps needed to set up and run a responsive service to support and evaluate realist and related evaluations and realist reviews, including costing different components of the support.

We will also run up to 6 workshops for patient organisations on how to support, participate in and critique realist evaluation (and, if desired, realist reviews), and we will develop additional resources (e.g. worked-up case examples of realist evaluations and resources for patients and lay members) to be placed with training materials on open access websites.

Background and rationale (why this research is needed now)

Realist research – an approach whose time has come

Realist review (for secondary research) and realist evaluation (for primary research) ask "what works, for whom, in what circumstances, in what respects and how?" [21]. These approaches are rapidly growing in popularity, in parallel with the rise of applied programmes of work at the interface between research, service development and education (e.g. Collaborations for Leadership in Applied Health Research and Care [CLAHRCs], Academic Health Science Centres/Networks [AHSC/Ns], and Health Innovation and Education Clusters [HIECs]), which are difficult to evaluate in randomised controlled trials [22-24].

There are a number of advantages for the health sector in supporting a realist approach. Firstly, complex programmes introduced in complex settings do not always work in or through the expected ways. Understanding how and why programmes work (or why they don't work as expected) can assist in refining programme design. Secondly, almost all complex interventions generate different outcomes for different populations in different contexts. A realist approach offers an approach to data analysis that not merely demonstrates that context matters, but makes sense of *how* and *why* context affects programme outcomes. This greater understanding can then be used to refine policy, to inform choice of programmes for particular populations or contexts, and to refine programme design to ensure best fit with particular target populations or local contexts.

Realist evaluation – a brief introduction by Ray Pawson

Realist evaluation was developed by Pawson and Tilley in the 1990s, originally in the field of criminology for the evaluation of "what works for whom in what circumstances and how?" in criminal justice interventions [25]. Realist review was developed as a form of secondary research to summarise and synthesise studies of programmes, with an early (and widely cited) monograph in 2005 [26]. This early work made the following points:

- Social programmes (closely akin to what health services researchers call complex interventions) are an attempt to address an existing social problem – i.e. to create some level of social change.
- Programmes 'work' by enabling participants to make different choices (although choice-making is always constrained by participants' previous experiences, beliefs and attitudes, opportunities and access to resources).
- Making and sustaining different choices requires a change in a participant's reasoning (for example, in their values, beliefs, attitudes, or the logic they apply to a particular situation) and/or the resources (eg information, skills, material resources, support) they have available to them. This combination of 'reasoning and resources' is what enables the programme to 'work' and is known as a programme 'mechanism'.
- Programmes 'work' in different ways for different people (that is, programmes can trigger different change mechanisms for different participants).
- The contexts in which programmes operate make a difference to the outcomes they achieve. Programme contexts include features such as social, economic and political structures, organizational context, programme participants, programme staffing, geographical and historical context, and so on.
- Some factors in the context may enable particular mechanisms to be triggered. Other aspects of the context may prevent particular mechanisms from being triggered. That is, there is always an interaction between context and mechanism, and that interaction is what creates the programme's impacts or outcomes: Context + Mechanism = Outcome.
- Because programmes work differently in different contexts and through different change mechanisms, programmes cannot simply be replicated from one context to another and automatically achieve the same outcomes. Theory-based understandings about 'what works for whom, in what contexts, and how' are, however, transferable.

- Therefore, one of the tasks of evaluation is to learn more about 'what works for whom', 'in which contexts particular programmes do and don't work', and 'what mechanisms are triggered by what programmes in what contexts'.

A realist approach assumes that programmes are "theories incarnate". That is, whenever a programme is implemented, it is testing a theory about what 'might cause change', even though that theory may not be explicit. One of the tasks of a realist evaluation is therefore to make the theories within a programme explicit, by developing clear hypotheses about how, and for whom, programmes might 'work'. The implementation of the programme, and the evaluation of it, then tests those hypotheses. This means collecting data, not just about programme impacts, or the processes of programme implementation, but about the specific aspects of programme context that might impact on programme outcomes, and about the specific mechanisms that might be creating change.

Pawson and Tilley also argue that a realist approach has particular implications for the design of an evaluation and the roles of participants. For example, rather than comparing changes for participants who have undertaken a programme with a group of people who have not (as is done in randomised controlled or quasi-experimental designs), a realist evaluation compares mechanisms and outcomes within programmes. It may ask, for example, whether a programme works differently in different localities (and if so, how and why); or for different population groups (for example, men and women, or groups with differing socio-economic status). Further, they argue that different stakeholders will have different information and understandings about how programmes are supposed to work and whether they in fact do so. Data collection processes (interviews, focus groups, questionnaires and so on) should be constructed to collect the particular information that those stakeholder groups will have, and thereby to refute or refine theories about how and for whom the programme 'works'.

The philosophical underpinnings of a realist approach

"Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the philosophy of science and social science." [27]

Philosophically speaking, realism sits between positivism ('there is a real external world which we can come to know directly through experiment and 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 holds that there is a real social world *but* that our knowledge of it is amassed and interpreted (sometimes partially and/or imperfectly) via our senses and brains, filtered through our language, culture and past experience.

In other words, realism sees the human agent as suspended in a wider social reality, encountering experiences, opportunities and resources and interpreting and responding to the social world within particular personal, social, historical and cultural frames. For this reason, different people in different social, cultural and organisational settings respond differently to the same experiences, opportunities and resources. Hence, a programme (or, in the language of health services research, a complex intervention) aimed at improving health outcomes is likely to have different levels of success with different participants in different contexts – and even in the same context at different times.

What realists mean by 'mechanisms'

To understand how programmes unfold differently (and in particular how they produce different outcomes) in different contexts, realism uses the concept of *mechanism*. While realists argue about the exact definition of a mechanism, they agree that, broadly speaking, the term relates to how human agents interpret the world around them and respond to the opportunities and resources available to them. To put it another way, mechanisms are what *generate* outcomes (more or less successfully and as planned – and/or in ways that are unplanned and unanticipated) when a programme is put in place.

For example, a school-based feeding programme for disadvantaged children may work by short-term hunger relief in a low-income rural setting where famine has produced overt nutritional deficiencies (the 'intended' mechanism), but the same programme in a troubled inner-city community where many children are disaffected may work by a different mechanism – making pupils feel valued and nurtured (a mechanism that was unanticipated by the programme's planners but which nevertheless had positive effects on children's learning) [17].

Another example of mechanisms, drawn from medical education, relates to asking learners on a widening access course for medicine to write free text responses to the question: 'How have you changed as a result of coming on this course?' [28]. Students might variously respond: 'I started to think more deeply about scientific problems'; 'I met real medical students and saw that you could still have fun whilst learning medicine'; 'I started to believe in my own ability', or 'I made some friends and we are planning to keep in touch on Facebook as we prepare for the medical school entrance tests'. These responses give an inkling of the complex outcomes that might be generated by a widening-access course and also suggest potential mechanisms by which the opportunities provided by the course might improve students' likelihood of applying to medical school and their competitiveness for places, thereby widening access. Expressed at a slightly higher level of abstraction, these mechanisms might be described as 'promoting reflection and deep learning', 'increasing motivation through vicarious experience', 'building confidence' and 'providing mutual support'.

The above examples illustrate the key features of realist 'mechanisms', which have been defined as *"... underlying entities, processes or [social] structures which operate in particular contexts to generate outcomes of interest"* [29]. Here 'entities' refers to things like an individual's norms or belief systems (e.g. students from some socioeconomic groups might not believe they are meant to go to university); 'processes' are sequences where later events depend on earlier ones (e.g. students will need to believe they have a chance of entry in to medical school before they will be willing to make the effort to apply); and social structures may refer to things such as gender, class, or cultural patterns of relationships (e.g. students will have specific individuals or groups of individuals whom they will consider their 'peers' and thus to have opinions and advice that matter). Like mechanisms in natural sciences (e.g. gravity), realist mechanisms possess a number of features: they are not 'visible' or directly measurable, but must be inferred from the observable data; they are context sensitive; and (to the extent that they operate in a particular scenario) they generate outcomes [29].

Realist research in healthcare

Many health problems facing patients and the NHS today are complex. Much health need results from the effects of smoking, suboptimal diets (including obesity), alcohol excess, inactivity or adverse family circumstances (e.g. partner violence). Interventions or programmes designed to tackle these challenges are themselves complex, since they have multiple components that are delivered individually and/or targeted at communities or populations. The success of such interventions depends both on individuals' responses and on the wider context in which people strive (or not) to live healthy lives. As we all know from experience, what works for a young professional woman in Knightsbridge may not work for a middle-aged unemployed man in Moss Side.

Similarly, the 'wicked problems' of contemporary health services research – how to improve quality and assure patient safety consistently across the service; how to meet rising need from a shrinking budget; and how to realise the potential of information and communication technologies (which often promise more than they deliver) – require complex delivery programmes with multiple components that engage with people and the particularities of context. What works for the paediatric service in hospital A may not work for the ophthalmology service in hospital B.

Designing and evaluating complex interventions is challenging [30, 31]. Randomised trials that compare 'intervention on' with 'intervention off', and their secondary research equivalent, meta-analyses of such trials, may produce statistically accurate but unhelpful statements (e.g. that the intervention works 'on average') which leave us none the wiser about where to target resources or how to maximise impact [32, 33]. The Medical Research Council's Guidelines for the Development and Evaluation of Complex Interventions was the result of a systematic and iterative attempt by expert panellists to advance the science of complex intervention research [31]. This widely cited guidance document acknowledged that a key step in the evaluation of complex interventions is to document the range of outcomes observed in different settings. The authors also commented that *"a good theoretical understanding is needed of how the intervention causes change"* and that "*ensuring strict standardisation may be inappropriate; the intervention may work better if a specified degree of adaptation to local settings is allowed for"* [31], but did not go as far as offering a theorisation of mechanisms of change or theory-driven approaches to adapting interventions to context. Realist evaluation, a relatively new approach (in health services research), based on realist philosophy (see above), aims to advance understanding of why complex interventions work, how, for whom, in what context and to what extent – and also to explain the many situations in which a programme fails to achieve the anticipated benefit.

Recent realist evaluations have addressed numerous topics of central relevance to the NHS, including what works for whom when 'modernising' health services [4], designing integrated care pathways [34], introducing breastfeeding support groups [18], using communities of practice to drive change [19] and involving patients and the public in research [20]. The methodology has spread extensively worldwide, especially to Australia and Canada where it is widely used in health services research.

The RAMESES I study

From the mid 2000s there was increasing interest in the use of realist reviews to make sense of complex interventions. However, there were also growing concerns that this method was being misapplied and that the projects that had been funded were not being carried out to the highest standards. In 2011, NIHR HS&DR funded us to undertake the RAMESES I study (www.ramesesproject.org) in recognition that standards and guidelines were urgently needed for realist reviews [9].

RAMESES I successfully produced, within budget and on schedule, consensus derived quality standards and publication guidelines [10, 11] and a menu of training resources. These resources are freely available online and highly accessed (e.g. the publication standards have had 18,000 hits in 13 months). We are aware of numerous ongoing realist reviews using the RAMESES standards, including several funded by NIHR [12-15]. Note that we are not claiming to have set the last word here – the final 'E' in RAMESES stands for 'evolving' [standards], on the assumption that as the knowledge base grows, the RAMESES standards will be periodically updated.

As part of RAMESES I, we set up (and continue to run) an active email discussion list (www.jiscmail.ac.uk/RAMESES), whose 300+ members include academics, policymakers, service clinicians and patient advocates. In developing this application, list members' top priority was that we should seek funding to do for realist evaluation what RAMESES I did for realist review – develop robust quality standards, publication guidance and training materials. From queries posted to the RAMESES list, it has become clear that researchers, editors and research users (policymakers, clinicians, patient advocates) are unclear what constitutes rigour when conducting and reporting realist evaluations.

Just as the RAMESES I quality standards and publication guidelines have quickly become the current benchmark for realist secondary research, we would now like to produce a similar set of resources that will help researchers produce consistently high quality realist evaluations. In RAMESES II, we also want to go beyond RAMESES I in one specific aspect – increasing the emphasis on patient and public involvement, since (as explained on page 5), realist approaches are particularly helpful in PPI.

Concerns about quality in realist evaluation

The RAMESES JISCmail listserv postings suggest that enthusiasm for realist evaluation and belief in its potential in applied health research have outstripped the development and application of robust quality standards in the field. There are also academic papers that have demonstrated this systematically.

Pawson and Manzano-Santaella in their paper 'A realist diagnostic workshop' use case examples of flawed realist evaluations to highlight three common errors in such studies [8]. Their first case illustrates that whilst it is possible to show associations and correlations in data from many types of evaluation, the focus of a *realist* evaluation is to explore and explain *why* such associations occur. As the authors put it, *"Realist evaluation is avowedly theory driven; it searches for and refines explanations of programme effectiveness"*. Their second case illustrates what may constitute valid data for use in realist evaluation. In the case they cite, only self reported data were collected and used to justify findings. Producing a realist explanation requires the use of data not just to describe and demonstrate what has happened, but also to *show how* it has. Hence a mix of data types, not merely self-reports, needs to be collected. Their final case example addresses context-mechanism-outcome configurations. The authors note that some realist evaluations have become bogged down in finely detailed analyses of such configurations but failed to produce a coherent explanation of how these configurations were linked and related (or not) to each other. Pawson and Manzano-Santaella call for greater emphasis on elucidating programme theory (the theory about what a programme or intervention is expected to do and in some cases, how it is expected to work).

In the same year that the above paper was published, Marchal et al. from Belgium undertook a review of the realist evaluation literature to quantify and analyse the field [6]. From an extensive search, they identified 18 realist evaluations and noted a range of challenges that arose for researchers. Absence of prior theoretical and methodological guidance appeared to have led to recurring problems in the realist evaluations they appraised. Firstly, *"The philosophical principles that underlie realist evaluation are*

variably interpreted and applied to different degrees. Most authors only fleetingly refer to the philosophical foundation of realist evaluation, which arguably is among its most distinctive features and provides much of its explanatory power". In addition, they noted that different researchers had conceptualised concepts used in realist evaluation, such as 'middle-range theory', 'mechanism' and 'context' differently. This, they concluded, was often related to fundamental misunderstandings. Where misunderstandings occurred, rigour of the realist evaluation undertaken often suffered.

These two papers show that realist evaluation is often an intellectually challenging task. Both sets of authors point out that more guidance is needed to allay misunderstandings about the purpose, underlying philosophical assumptions and analytic concepts and processes of realist evaluation.

In sum, there is strong evidence of unmet need for methodological and other support for primary realist research. We think a good way of building capacity to meet such need would be to generate resources and training materials and offer 'training the trainers' courses for people and organisations who deliver research design and methodological support. We developed our initial bid in consultation with the London Research Design Service and refined it in consultation with the RDS nationally.

Realist research and patient and public involvement

Realist research has the advantage of – potentially – having high patient and policy relevance because its driving question is "what works, for whom, in what circumstances and how?". Realist approaches have been used in particular to study this set of questions in patient and public involvement (see for example these studies [1-5]). There is huge potential for extending the application of realist methods to maximise patient and public involvement in the evaluation and improvement of health services.

Realist evaluations in health services research are increasingly being conducted with the active involvement of non-academic stakeholders (patients, community leaders, hospital managers and others) as research partners. This is partly because the complexity inherent in such realist research (in terms of what works, for whom in what circumstances) greatly benefits from multi-stakeholder research teams that involve 'insiders' who can help to identify key contextual factors and mechanisms at play in the settings and interventions they are intimately connected to. But despite the appropriateness of involving patients and the public in realist research, the current state of literature and training materials for lay reading of realist evaluation is greatly lacking. In RAMESES II, we will create new materials for understanding and implementing realist evaluation that will be accessible to people of diverse educational, professional and cultural backgrounds. In addition, we will draw on the expertise in our patient panel to extend and improve explanatory materials on realist synthesis intended for non-academic audiences.

Aims and objectives

AIM

To realise the potential of realist research for improving the delivery of health services.

STRATEGIC OBJECTIVES

1. To develop quality standards, publication guidance and training materials for realist evaluation.
2. To build capacity in health services research for supporting and assessing realist approaches to research.
3. Acknowledging the unique potential of realist research to address the patient's agenda ("what will work for us in our circumstances?"), to produce resources and training materials for lay participants, and those seeking to involve them, in research.

OPERATIONAL OBJECTIVES

1. Establish a management and governance infrastructure, including a project steering group with lay representation and a patient/service user panel.
2. Recruit an interdisciplinary Delphi panel consisting of researchers, people who support and help design research studies, publishers, peer reviewers, policymakers, patient advocates and practitioners with (various types of) experience relevant to realist evaluation.
3. Summarise the current literature and expert opinion on best practice in realist evaluation, to serve as a baseline / briefing document for the panel.

4. Run three (and more if needed) rounds of the online Delphi panel to generate and refine items for a set of quality standards and publication guidance.
5. In parallel with the Delphi panel:
    a. Provide ongoing advice and consultancy to up to ten realist evaluations, including any funded by NIHR, thereby capturing the 'real world' problems and challenges of this methodology.
    b. Host the RAMESES Jiscmail list on realist research, capturing relevant discussions about theoretical, methodological and practical issues.
    c. Feed problems and insights from 5a and 5b into the deliberations of the Delphi panel and the design of training materials and courses.
6. Write up the quality standards and guidance for publication in an open-access journal.
7. Collate examples of learning/training needs for researchers, postgraduate students, reviewers and lay steering group members in relation to realist evaluation.
8. Develop, deliver and evaluate training materials for realist evaluation. Deliver 3 x 2-day 'realist evaluation' workshops AND 3 x 2-day 'training the trainers' workshops for a range of audiences (including interested NIHR Research Design Service staff).
9. Develop, deliver and evaluate information and resources for patients and other lay participants in realist evaluation. In particular, draft template information sheets and consent forms that could be adapted for ethics and governance activity, and deliver up to 6 workshops for PPI organisations.
10. Disseminate training materials and other resources – e.g. via public access websites.

ANTICIPATED OUTPUTS

1. Quality standards, publication guidelines and training materials for realist evaluation.
2. Bespoke methodological support to up to ten realist evaluations (with NIHR funded studies prioritised).
3. Significant increase in the capacity of the health services research community to support and evaluate realist research.
4. Accessible, plain-English resources for patients and the public who are asked to participate in realist research or oversee its management and governance

Research Plan / Methods

THEORETICAL BACKGROUND

Quality standards and publication guidelines are common (and, increasingly, expected) in health services research - see for example CONSORT for randomised controlled trials [35], AGREE for clinical guidelines [36], PRISMA for Cochrane-style systematic reviews [37], SQUIRE for quality improvement studies [38] and RAMESES for realist reviews [11]. They have two main purposes: they help researchers design, undertake and report robust studies, and they help reviewers and potential users of research outputs assess validity and reliability. This project seeks to produce a comparable set of quality standards and publication guidelines for realist evaluation.

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 [39]. It was used, for example, to set the original care standards that formed the basis of the Quality and Outcomes Framework for United Kingdom general practitioners [40]. Increasingly, and largely to overcome the practical challenges of bringing experts together repeatedly to discuss an issue, Delphi panels are conducted online. In the study protocol for RAMESES I, we justified the use of an online Delphi panel as follows (page 3) [9]:

"*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 [41]. Asynchronous online communication has well-established benefits in promoting reflection and knowledge construction [42]. There are over 100 empirical examples of successful online Delphi studies conducted between geographically dispersed participants [9 examples cited]. 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 inter- national 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.' [43]".*

The above text was written before the RAMESES I study was complete but our own experience of using an online Delphi for that and other studies confirms the very positive comments on this method that we found in the literature. In short, we anticipate that the online Delphi method will be highly acceptable to participants, have a low withdrawal rate and generate high-quality data for our consensus process.

STUDY DESIGN AND METHODS

The study's 10 operational objectives (listed on page 6) will be delivered in three workstreams, underpinned by a management and governance infrastructure. The detail is set out below.

Objective 1 *Establish a management and governance infrastructure, including a project steering group with lay representation and a patient/service user panel*

Described in 'Management and Governance' below.

WORKSTREAM 1 (Objectives 2, 3 and 4)

Objective 2 *Recruit an interdisciplinary Delphi panel consisting researchers, people who support and help design research studies, publishers, peer reviewers, policymakers, patient advocates and practitioners with (various types of) experience relevant to realist evaluation.*

For the online Delphi panel, we will apply the same successful approach as we did for the RAMESES I study [9]. We will recruit 35 panellists the groups listed in the objective above (including patient organisations). Recruitment will be done by the core working group, drawing on our knowledge of the field, our different professional networks, the RAMESES JISCmail listserv and our links to user organisations. Input from a wide range of experts in relevant fields will be sought. Those who meet one or more criteria for expertise will be briefed on the project, what is expected from them and informed that participation is voluntary and unpaid and that they may withdraw at any time. We will ensure representation from all relevant stakeholder groups, if necessary by asking existing panel members to nominate and invite others.

Objective 3 *Summarise the current literature and expert opinion on best practice in realist evaluation, to serve as a baseline / briefing document for the panel*

With expert librarian help, we will identify reviews, scholarly commentaries, models of good practice and examples of (alleged) misapplication of realist evaluation [6-8]. To identify the relevant documents we will refine and develop the search used by Marchal et al for a previous review on a similar topic [6], and also apply contemporary search methods designed to identify 'richness' when exploring complex interventions [44, 45]. We will thematically summarise [a] what is considered by experts to be current best practice (and the range and diversity of such practice); [b] what experts and other researchers believe count as high quality and needs to be reported; and [c] what issues researchers struggle with (based on thematic analysis of postings on the RAMESES JISCmail list archive as well as the published literature). The purpose of this step is not to produce definitive answers to these questions but to prepare a baseline set of briefing materials for the Delphi panel, who will deliberate on them and add to them in the next step.

Objective 4 *Run three (and more if needed) rounds of the online Delphi panel to generate and refine items for a set of quality and publication standards*

As in RAMESES I, the Delphi panel will be run using the online survey tool SurveyMonkey [9]. In round 1, we will send out briefing materials and invite suggestions from the panel members on what topics and themes they think should be included in the standards and guidance. Responses will be analysed and fed into the design of questionnaire items for round 2. Panel members will rank items for [a] relevance (should we include this topic/theme?) and [b] content (should we word it like this?) on a 7-point Likert scale from 1 = strongly disagree to 7 = strongly agree. High scores for relevance and content will mean the item will be included. High scores for relevance but low scores for content will mean we need to word the item differently (we will ask for suggestions). Low scores for relevance mean the item will be dropped. The SurveyMonkey software will allow panel members to offer suggestions in free-text boxes on item wording or any other issue related to the item.

At the end of round 2, we will analyse responses (including free text), dropping and/or modifying items to reflect panel members' recommendations. This will generate an 'end of round 2' briefing document for the panel. This will consist of each item's wording; the response rate; mode, median and interquartile range values of the ratings; the actions we took as a result of the panel's ratings and an anonymised list of all the free text comments made. Each participant will receive the scores they gave each item alongside the average for the panel, allowing them to see how (if at all) they deviated from the majority.

We will then produce a new, shorter questionnaire for round 3 that contains only items where consensus has not yet been reached. We will continue the same process until no further move towards consensus is possible, though in practice very few Delphi panels go beyond 3 rounds. A Delphi panel seeks to work towards consensus but it should not be seen as having failed if consensus is not reached on some items, especially since a Delphi exercise has a secondary aim of mapping the level of residual dissent. For example, in a previous study, we used a Delphi panel to develop quality standards for narrative research [46]. In that study, the panel included people from a wide range of academic disciplines and some people who self identified as 'service users'. It would be very surprising if they had agreed on all aspects of the topic. After three rounds of Delphi, we reached high agreement on almost all items but for one or two items, there remained a wide spread of views. To deal with that, we simply reported the finding and concluded that to set hard and fast rules on these aspects of narrative research would be unwise.

WORKSTREAM 2 (Objectives 5 and 6)

Objective 5 *In parallel with the Delphi panel:*

a. *Provide ongoing advice and consultancy to up to ten realist evaluations, including any funded by NIHR, thereby capturing the 'real world' problems and challenges of this methodology.*
b. *Host the RAMESES Jiscmail list on realist research, capturing relevant discussions about theoretical, methodological and practical issues.*
c. *Feed problems and insights from 5a and 5b into the deliberations of the Delphi panel and the design of training resources and courses*

We will provide advice and or methodological support to up to 10 realist evaluations. To sample ten that unfold in parallel with our Delphi exercise, we will [a] ask NIHR to link us with planned evaluations funded by them that align with our own timeline; [b] ask on the RAMESES list; [c] capture unsolicited requests for help (of which we receive many). We will aim for maximum variety in experience of research teams, topics, settings and approach to patient and public involvement. We will work flexibly with teams, mostly by phone, Skype and email, to support them with methodological advice and troubleshooting. We will systematically capture the questions and issues from these ten primary studies and feed them into the deliberations of the Delphi panel (where timings permit) and, if relevant, the training materials and courses described below.

We provided a comparable service to realist review teams in the RAMESES I study, and plan to follow a similar approach. In RAMESES I, there was considerable variation in the level of expertise and confidence in the research teams. Some were highly skilled and used our input mainly as 'sounding board' for their own developing ideas and methodology. Others lacked basic understanding of realist concepts and methods; they were offered face-to-face training workshops and bespoke support with data analysis and interpretation. All but one of the teams completed their study and submitted a publication to a peer-reviewed journal, sometimes with one of our team as a co-author [5, 47-51]; the final team experienced a delay and may still complete. We captured numerous methodological issues that fed into the design of training materials and also informed some methodological papers by our team and the teams we worked with (some of whom have now joined this new collaborative bid) [16, 52]. We will aim for a similar set of outputs in this work package in RAMESES II.

Objective 6 *Write up the quality standards and publication guidelines for an open-access journal*

Again, we will follow the method applied successfully in RAMESES I to produce an account of the background, methods, main findings and conclusions of the Delphi project, including publishing a detailed protocol in an open access journal [9-11] and engaging the editors of specialist journals in potential parallel publication to reach an extended range of readers [53, 54]. We will also, as in RAMESES I, enter into dialogue with the EQUATOR network (http://www.equator-network.org), a

clearinghouse for publication standards which is used as a first port of call by researchers seeking such standards, and which already lists the RAMESES standards for secondary research.

Whilst we do not anticipate major problems in the writing up of this work for publication, we believe that achieving consensus on both quality standards and publication guidelines may be more difficult for realist evaluation than it was for realist review, since the former covers a huge variety of settings, topics, approaches and configurations [7]. Hence it is possible that, unlike in RAMESES I, consensus among Delphi panel members may not be achieved for all items. This is not inherently a problem: in a previous Delphi study to develop standards for undertaking and reporting narrative research, we simply reported, and commented on, the areas of residual disagreement between panel members, which were explained by their different disciplinary and/or sectoral backgrounds [46].

WORKSTREAM 3 (Objectives 7-10)

Objective 7 *Collate examples of learning / training needs for researchers, postgraduate students, reviewers and lay steering group members in relation to realist evaluation*

We will seek examples of the kinds of requests that are made by researchers for support on realist evaluation. We already have a rich archive of postings on the RAMESES JISCmail listserv from both novice and highly experienced researchers, going back three years. We will also proactively ask the list members for additional examples; use our empirical data from workstream 2 on the real-world struggles of realist researchers (see Objective 5 above); and draw on our literature review (Objective 3) and Delphi panel discussions (Objective 4), to identify relevant examples. Finally, we will seek input from Research Design Service staff, particularly with those who respond to an invitation sent out by the RDS Steering Group on our behalf. We will ask such RDS staff (some of whom are already members of the RAMESES list) to describe the kind of problems people bring to them, and where they feel that further guidance, support and resources are needed.

We will use a thematic approach to classify examples into a coherent taxonomy of problems and issues, each with a corresponding training need(s). This will be developed iteratively in regular meetings chaired by TG. At least two researchers will independently classify examples within this taxonomy and through subsequent discussion with the wider team, both the taxonomy and the classification of examples within it will be refined. The goal of this step will be to feed into a coherent and comprehensive curriculum for training realist researchers and for 'training the trainers'.

Objective 8 *Develop, deliver and evaluate training materials for realist evaluation. Deliver 3 x 2-day 'realist evaluation' workshops AND 3 x 2-day 'training the trainers' workshops for a range of audiences (including interested NIHR Research Design Service staff)*

To develop training materials, we will analyse and take forward various problems, issues and learning needs raised in the examples identified in Objective 7. Some will be philosophical or theoretical, some methodological, some practical, some ethical, and so on. Different kinds of learning need require different materials and resources and delivered by different media (face-to-face, internet) and in different learning arrangements (self-study, online drill-and-practice, interactive group tasks and so on). Developing the resources will involve setting specific learning objectives, preparing study notes (e.g. explanations, diagrams) and developing and piloting exercises to engage learners. For each main challenge, we will produce a menu of materials oriented to different audiences and learning styles. Several of the applicants on this bid are experienced trainers and consultants on realist evaluation; we will draw on, and refine, the existing training materials that we have developed and acquired over the years.

It is important to stress that realist research cannot be achieved simply by following a protocol in a technically correct manner. Rather, becoming competent at realist research involves acquiring the ability to think, reflect, and interpret data in a way that is resonant with realist philosophy and principles. For this reason, much of the workshops will take the form of "show and tell", facilitated discussion and "apprenticeship" to experienced and skilled realist researchers.

We will run 3 x 2-day 'how to do a realist evaluation' workshops for a main audience of researchers and evaluators, and including research users – both lay and professional and 3 x 2-day 'training the trainers in realist research' workshops for a main audience of those who train and support such work. In both sets of workshops, diversity of background will be used productively in group-based case discussions

and other hands-on, interactive formats. We will run one of each workshop in London, Leeds and Liverpool (the cities where the UK members of our collaboration are based).

The training the trainers workshops in particular will be open to RDS staff who seek to become confident in supporting realist studies; they will also seek interdisciplinary participation from researchers, practitioners, policymakers and patient advocates. The detailed curriculum for the workshops will emerge from our empirical work, but the training the trainers programme will include all the steps needed to set up and run a responsive service to support and evaluate realist reviews and evaluations, including costing different components of support.

Objective 9 *Develop, deliver and evaluate information and resources for patients and other lay participants in realist evaluation. In particular, draft template information sheets and consent forms that could be adapted for ethics and governance activity, and deliver up to 6 workshops for PPI organisations.*

We will engage with our patient/service user panel to help us develop resources that are relevant, understandable and useful to this group. Examples are: the quality and publication standards; some of the training resources, especially lay summaries of what a realist evaluation is; template information sheets and consent forms for participants in realist evaluations.

As well as developing 'generic' patient/lay resources, we will offer up to six half-day workshops on realist evaluation for patient organisations. We will work with each organisation to develop a curriculum and format. Organisations for these workshops will not be formally sampled as we have found in the past that we receive 'ad hoc' requests for such input, which we often have to turn down because of lack of protected time. Hence this will be a responsive component of the study, dependent on which organisations approach us. Those who do so will probably hear about us from the following sources: [a] the RAMESES listserv, whose membership includes a number of patient/lay advocates; [b] our patient panel and their personal networks; [c] social media invitations (e.g. the PI for this study has an active presence on Twitter and more than 10,000 followers, many of whom represent patient organisations); and [d] newsletters and email feeds from organisations such as INVOLVE.

Objective 10 *Disseminate training materials and other resources – e.g. via public access websites.*

Our previous approach of dissemination of the RAMESES I resources was very successful (e.g. the publication guidelines were classified as 'highly accessed' by BMC Medicine within 24 hours of publication and accessed 18,000 times in the subsequent 13 months); in RAMESES II we will replicate the approach we used for that study. We used a combination of publication of the standards in peer-reviewed journals (including parallel publication in a nursing journal); a RAMESES project website providing open access to all resources; the RAMESES JISCMail list (on which we posted the links to the above); and submission of the academic papers to the EQUATOR NETWORK (an international clearinghouse for peer-reviewed publication standards, http://www.equator-network.org).

Because RAMESES II is placing more emphasis on patient and lay involvement than RAMESES I, we will have an additional component to dissemination – development, piloting and publishing of lay summaries of the key publications. Depending on the journal, it may be possible to publish these lay summaries alongside the academic papers (e.g. New England Journal of Medicine offers such an option). We will make lay summaries available on the RAMESES website, and will negotiate with COREC (research ethics) and INVOLVE to publish templates of information sheets and consent forms for patient participants in realist evaluation. We will ask the Research Design Service to link to resources relevant to their staff and clients (and have agreement from the RDS to do this in principle).

SUMMARY OF HOW WE WILL INVOLVE PATIENTS AND THE PUBLIC

As explained on page 5 ('Realist research and patient and public involvement'), realist evaluation not only accommodates but effectively *requires* the active involvement of patients and other non-academic participants, since a thorough realist evaluation necessarily seeks the perspective of everyone involved. Realist evaluation is a relatively new approach in health services research, hence may be unfamiliar to many patients and members of the public, even those who are familiar with other research designs such as clinical trials. Those who do not understand the principles or methods of realist evaluation cannot give full informed consent to participation, hence training materials and resources are urgently needed to explain in lay terms what a realist evaluation is and what it can (and cannot) contribute to the assessment and improvement of services from the patient's perspective.

We plan six components of user involvement, to ensure that the patients and public are involved throughout the project and in each of the workstreams. (This is an aspect of the RAMESES work that we have developed and extended since RAMESES I, which did not have a strong focus on PPI).

1. A patient / service user panel to whom we will present progress and findings every 6 months. This will have cross-membership with the project steering group (i.e. some but not all members of the patient panel will also join the steering group) – see Management and Governance, page 12.
2. Patient advocates as members of the Delphi panel – see Objective 2, page 7. We anticipate recruiting one or more people who are working in an evaluation role for patient organisations.
3. Patient/lay input to the design of selected training resources (especially training materials for supporting patient involvement in realist evaluations) – see Objectives 8 and 9, page 9-10.
4. Workshops for patient organisations on realist evaluation (see Objective 9, page 10).
5. Patient/lay input to a lay summary of our findings – see Objective 10, page 10.
6. Patient/lay input to disseminating our findings e.g. via lay press and patient portals – see Objective 10, page 10.

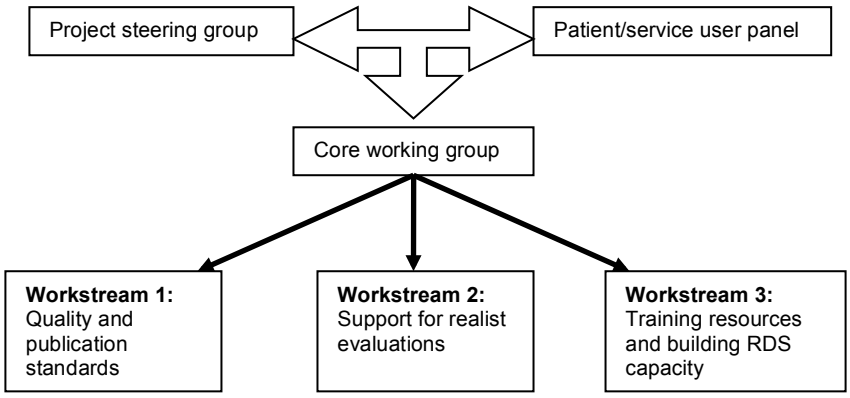
For recruitment of lay participants, see 'Management and Governance' below.

We believe that lay involvement will benefit the research in four main ways:

- Maintain a firm focus on the practical usefulness of the guidance and resources we produce for patients and the public. This is partly to avoid falling into the trap of producing 'self-serving' resources that are relevant only to the academic community and linked evaluation experts.
- Building capacity in patient and user organisations to participate in, contribute to, and critique realist evaluations relevant to their area of interest.
- Jargon busting. As with many research methods, realist evaluation has its own 'jargon'. Researchers new to it may struggle with realist concepts and assumptions. The best people to identify obscure or meaningless terminology are those who are not closely familiar with the use of such terms, and we will value the input of lay members here.
- Extending dissemination to a wider range of audiences. Dissemination via traditional academic channels (e.g. the academic journals) is less likely to reach patients and the public – but also less likely to reach many 'professional' users. Increasingly, clinicians, managers and academics first learn of innovations in their field when they read an article in the *lay* press. Hence, working with a patient panel to produce the highest quality outputs targeted at a 'lay' audience will also help disseminate to a range of clinical and policy audiences.

Management and governance

Research management arrangements



We plan to use the same management model as RAMESES I, which enabled us to complete that project on time and within budget. The core working group will meet fortnightly, and the steering group (with lay representation) and a separate patient / service user panel will each meet 6 monthly. This infrastructure will advise and support (but not replace) regular meetings among the researchers, as needed, to execute

the study, conduct the data analysis, discuss emerging findings and prepare outputs. The management structure is shown above.

The core working group, chaired by TG, will include GWong, JJ, JG and AM-S. GWesthorp will join this group when relevant. To minimise travel expenses and to save time, we will use Skype Premium, which allows high-quality remote interaction at modest cost. This group will plan and monitor day to day progress, ensure ongoing communication among researchers, review quality and timeliness of outputs, and manage day to day risks and issues. An experienced project manager will (in close discussion with TG) clarify the project brief, roles and responsibilities, terms of reference, communication plans and so on and monitor project progress. Governance and financial management will be formally overseen by the University of Oxford.. All data will be handled in accordance with the Data Protection Policies of our respective institutions. In house expert librarians/informaticists will assist with literature searching.

The steering group will have wide cross-sector representation (including experts in realist evaluation, research support, NHS professionals and representatives from the patient panel – see below). It will monitor progress against milestones and spend against budget, provide advice, promote the project, communicate with stakeholders and help maximise dissemination and impact of findings. In addition, where needed it will act as a sounding board and 'critical friend' to the project team.

The patient panel will provide advice and feedback to the working group to on how to present the study and findings in a way that is maximally accessible to lay people. Representatives from it will attend the wider steering group (with training and support if required). Where necessary we will provide induction and training to the group members and ensure that they are made aware that their participation is entirely voluntary and may withdraw at any time. Lay participants will be drawn from patient organisations with whom we are already working (e.g. Asthma UK, assistive technology research, diabetes service development groups), our social medial contacts and the INVOVLE network of lay participants from whom we have recruited very successfully in the past.

Timetable

The key tasks and their timings are set out in detail in the Gantt chart. Briefly, they comprise:

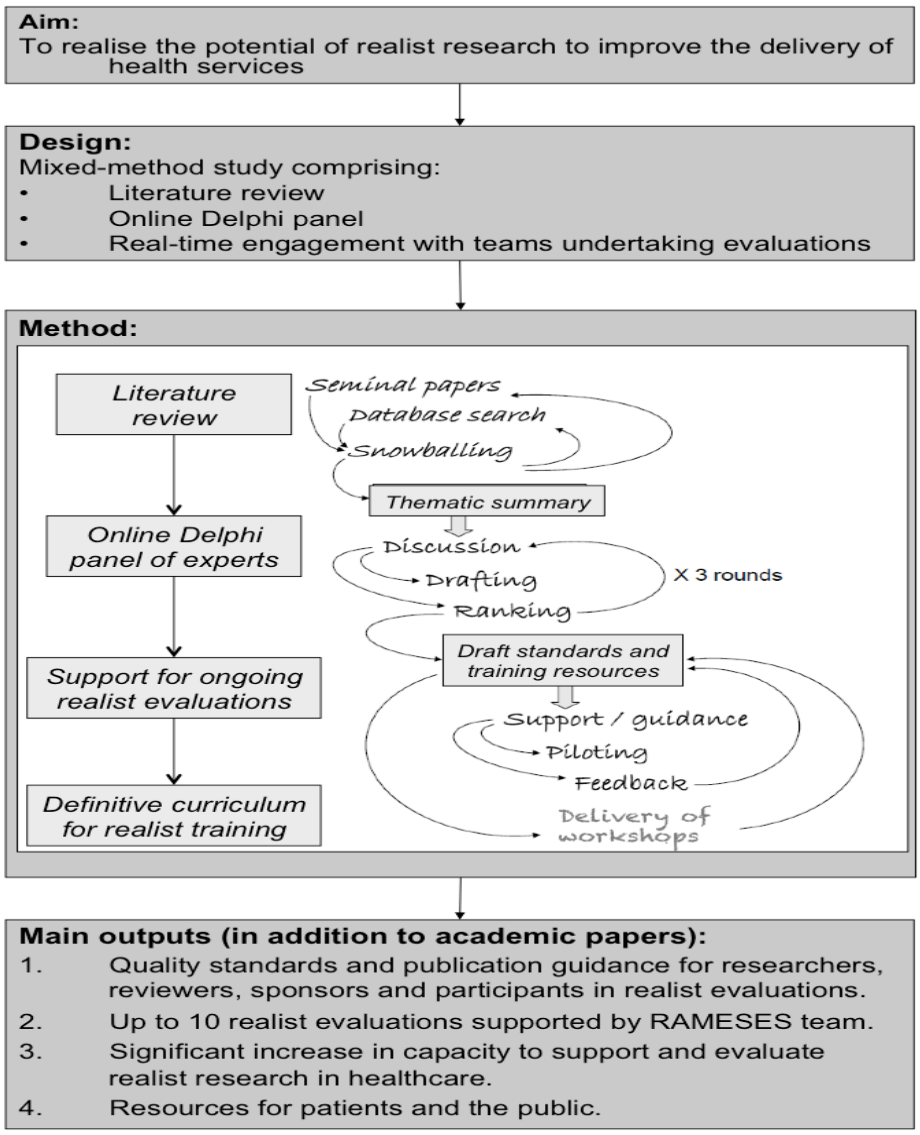
MONTHS 0-3
Set up project, appoint steering group and patient / service user panel
Confirm all ethics approvals in place
Clarify brief, allocate roles and responsibilities
Provide staff training if needed
Commence literature review and prepare background briefing
Recruit Delphi panel
Recruit realist evaluation studies that seek support over an appropriate timescale

MONTHS 3-15
Run 3 rounds of Delphi panel with cycles of item generation, panel surveys and refinement of items
Provide methodological support to up to 10 realist evaluations
Run the RAMESES email list
Feed real-time problems and issues into deliberations of panel
Run realist evaluation workshops x3

MONTHS 15-21
Draft academic papers and executive and lay summaries
Prepare draft training materials and resources and circulate for feedback
Run training the trainers workshops x3
Run PPI workshops x6

MONTHS 21-24
Finalise and publish all web resources
Finalise and publish academic papers
Write final report

Project flowchart



Gantt chart

TASK	MONTH							
	3	6	9	12	15	18	21	24
ADMINISTRATION / GOVERNANCE								
Seek ethics exemption and any necessary approvals								
Recruit and set up steering group and patient panel								
Steering group and patient panel / reports to funder	X X		X X		X X		X X	Final report
WORKSTREAM 1: DEVELOPMENT REALIST EVALUATION QUALITY AND PUBLICATION STANDARDS								
Clarification of brief and focus								
Searching / appraising published realist evaluations								
Preparing briefing documents for Delphi panel								
Recruiting to Delphi Panel								
Delphi round 1								
Delphi round 2								
Delphi round 3								
Preparing standards and guidance for realist evaluation								
WORKSTREAM 2: Bespoke support for realist evaluations								

Identify and contact ongoing realist evaluations								
Provide advice and methodological support								
WORKSTREAM 3: DEVELOPING TRAINING RESOURCES AND BUILDING CAPACITY								
Collate examples of learning/training needs								
Iteratively produce and refine training materials								
Run training workshops (3 on RE, 3 training the trainers, 6 for PPI organisations)								
Develop, deliver and evaluate information and resources for patients and other lay participants								
DISSEMINATION								
Draft dissemination strategy (and refine as needed)								
Consult with patient panel on dissemination strategy								
Writing up definitive quality and publication standards								
Adapting training materials for suitable audiences – patients, researchers and RDS								
Writing up definitive training materials								

Outputs, dissemination and impact

Intended audiences for this research

Via the RAMESES JISCmail list, and also the wider community of practice of realist researchers and research users, there is already an active and interested international network of academics, policymakers, practitioners and patient advocates who will be following the study as it unfolds (some of whom will be participants in the Delphi panel and/or attending the training courses) and who will be ready to help disseminate its findings. Our specific dissemination plans by audience are:

AUDIENCE 1: THE RESEARCH AND EVALUATION COMMUNITY

This is the audience to whom we are already most closely associated and widely networked, as we are realist researchers and evaluators ourselves. The channels listed below build on existing national and international networks of realist researchers and formal academic outlets familiar to this audience.

Quality standards and publication guidance for realist evaluation will be disseminated via:

- One or more publications in an open-access journal (e.g. BMC Research Methodology, who published RAMESES I)
- Submission to the EQUATOR Network, a repository of standards for different types of health research (http://www.equator-network.org)
- RAMESES project website (www.ramesesproject.org)
- RAMESES JICSMail listserv (www.jiscmail.ac.uk/RAMESES)
- 'Realist hive' blog run by colleagues at University of Exeter (http://blogs.exeter.ac.uk/realisthive/)

Training days in realist evaluation

- Three x 2-day workshops (London, Leeds, Liverpool) on how to do a realist evaluation

AUDIENCE 2: ORGANISATIONS THAT SUPPORT RESEARCH DESIGN

Although the NIHR Research Design Service is not a collaborator or partner (because it is against their policies to 'partner' research bids from HEIs in the conventional sense), *"RDS Strategy Group have agreed to support this application by disseminating requests, from the research team, to all RDS staff interested in participating in the Delphi panel. NIHR RDS staff will also be encouraged to participate in the workshops outlined in work package 3. When the study is complete, RDS will distribute the outputs across our networks and include a link to resources produced by RAMESES II on the NIHR RDS webpages*". (email from Wendy Baird to Prof Greenhalgh 28.3.14)

This generous offer from RDS will enable us to widely publicise any workshops we run across the RDS network. However, our workshops will not be limited to RDS staff but will include any organisation that seeks to support the design and delivery of realist research. For all such organisations, we plan to disseminate and build capacity as follows:

- Three x 2-day 'training the trainers' workshops for research support staff

- Online resources in a format that such staff will be able to use in their practical work of supporting teams seeking to design a study (see for example the 'Better Research' page for RDS London http://www.rdslondon.co.uk/Online-Resources/better-research.aspx)
- Newsletters, email circulars and social media feeds from both RDS and the RAMESES team

AUDIENCE 3: SPONSORS AND FUNDERS OF REALIST EVALUATIONS
To assist them in deciding when realist evaluations should be undertaken, we will produce executive / lay summaries of when (and when not) to use realist evaluation and quality checklists to apply to grant proposals, disseminated via (e.g.) the RAMESES website. These audiences will include organisations with whom we have worked in the past to advise on the place of realist evaluation and review and the quality of research bids and proposals submitted to them (including the Research Councils, National Institute for Health and Care Excellence and Health Foundation).

AUDIENCE 4: PATIENTS AND THE PUBLIC
Patients and the public have a tradition of active involvement in realist research, as it is not programmes per se that achieve impact but how the programme resources and infrastructure are interpreted, drawn upon and applied by human actors, including patients and the public. We anticipate that the dissemination channels listed below will reach this audience. However, we will consult with our patient / service user panel to ask for their advice on whether we should be using any other dissemination channels. 'Plain English' guides to realist evaluation along with template information sheets and consent forms for use in ethics / governance activity will be disseminated via

- INVOLVE website and activities
- Individual patient organisations (via bespoke workshops)
- Patient/public attendance at 'realist evaluation' workshops (see above)

Expected outputs

PUBLICATIONS
- Quality standards for realist evaluation
- Publication guidance for realist evaluation
- Theoretical/methodological papers on realist approaches
- 'Plain English' lay summary of standards and guidelines
- Executive summary for research sponsors and funders

TRAINING MATERIALS AND RESOURCES
- Curriculum, reading materials, pre-course exercises, powerpoints and interactive group activities for workshops (see below);
- RDS resources on developing, supporting and getting funding for a realist evaluation, including indicative costings for such support.

TRAINING PROGRAMMES
- 3 x 2-day workshops 'How to do a realist evaluation'
- 3 x 2-day workshops 'How to support and evaluate a realist evaluation'
- Bespoke training (up to 6 half-day workshops) for patient organisations on realist research and how it may help achieve their mission

TOOLS AND CHECKLISTS
- Checklist for writing grant applications for realist evaluation (for researchers)
- Checklist for assessing grant applications for realist evaluations (for funders/reviewers, including plain English version for lay members on award panels)
- Checklist for assessing a paper submitted for publication (for editors/reviewers)
- Checklist for patients who may be asked to participate in a realist evaluation (for patients, researchers and ethics committees)

Anticipated impacts
Research that is not undertaken to the highest accepted standards may have negative impacts. Researchers misapplying realist evaluation will end up with findings that could mislead and confuse. It is these very findings that policy and decision makers (e.g. such as GPs in the new clinical commissioning

groups) will draw in when making decisions about what to do when faced with health care service delivery issues. For funders of research, the opportunity cost of funding a poorly executed realist evaluation must be considered. Patients and the public are often involved in realist evaluations. It would be unethical to involve them in research that is not carried out rigorously. We believe that this project has the potential to impact on the areas listed below:

1. Academic impact:
    a. Greater understanding of realist approaches in health services research;
    b. Greater quality and consistency of realist evaluations;
    c. Empirical evidence on what works for whom in what circumstances, potentially much more useful to policymakers than studies that measure average 'effect size'.
2. Impact on research policy / funding:
    a. Better targeting of research funds to the appropriate and rigorous use of realist methods;
    b. Greater chance that funded studies will have economic and societal impact (see below).
3. Impact on health policy and services:
    a. Greater chance that those developing policy and/or redesigning services will find a relevant evidence base to inform decisions;
    b. Greater potential for evidence-informed customisation of interventions for particular contexts and settings.
4. Impact on health service users:
    a. Greater chance that an intervention they are offered will work (or can be adapted to maximise its chances of working) for them;
    b. Greater potential to have meaningful input to the design and customisation of a complex intervention.

Success criteria and risks

How we will measure the success of the study

We will consider this project a success when the following outputs are produced:

a. Robust quality standards for realist evaluation
b. Evidence-based publication guidelines for realist evaluation
c. Significant capacity within the health services research community for supporting the production of high-quality realist research studies.

Barriers to success and project risks

We do not anticipate major barriers since the study is a follow-on from RAMESES I which we undertook with a similar budget, timescale, management arrangements and stakeholder group. The main challenge we anticipate is in the recruitment of our Delphi panel, which (to be valid) needs to comprise a diverse group of experts who are able and willing to commit to the timescale of the study. However, from RAMESES I we have past experience in successful recruitment to a Delphi panel. We are aware that that within the realist community of practice, there is significant interest in both the development of quality and publication standards for realist evaluation and training practices. We will draw on this interest to recruit. The project team has extensive professional links to a broad range of individuals who would be suitable for Delphi panel recruitment. We were able to successfully recruit in RAMESES I through direct personal contact from the project team to potential panel members. The asynchronous and online nature of the Delphi panel also helps us to recruit as it offers flexibility to busy individuals – enabling them to participate and complete their ratings regardless of where they may be located and at a time of their choosing. In RAMESES I, we found that some panel members we had recruited were very helpful in assisting to identify 'gaps' in our panel and in suggesting potential candidates. We will seek the help of or panel members in a similar way if the need arises. In this project we have the additional option of using the RAMESES JISCMail listserv to recruit potential panel members. RDS nationally has offered to help recruit from among its senior staff.

We did not encounter any problems with participation within the Delphi panels in RAMESES I. We were able to get overall response rates for each round of over 90%. We did this through reminder emails and also by warning participants in advance that their involvement would be needed soon, providing a realistic time frame for completion and making the process as clear and simple as possible. We will use

the same approaches in this project to ensure we get high and consistent participation in the Delphi panels.

Lay involvement is very important to this project. We anticipate that lay members of our steering group and patient panel will need training and support to understand the principles of realist evaluation and provide input to the study's progress. We have allocated time and a budget for this activity and recruited JJ who is a specialist in participatory research to ensure that our involvement of lay members is democratic and maximally effective. We will also draw on the practical experience and expertise of all the applicants of working with lay members. For example, in the HS&DR funded realist review, 'Motivations, mandates and accountability in demand management' (HS&DR - 11/1022/04 – Leeds team), lay members are being successful involved in the steering group. In the ongoing HTA funded, 'Does therapeutic writing help people with long-term conditions? Systematic review, realist synthesis and economic modelling' (HTA - 11/70/01 – London team), lay members have contributed on a number of fronts, from helping in the research itself (e.g. devise programme theory and set inclusion / exclusion criteria), to co-authoring the protocol and sections of the final project report.

REFERENCES

1. Gagliardi AR, Legare F, Brouwers MC, Webster F, Wiljer D, Badley E, Straus S: **Protocol: developing a conceptual framework of patient mediated knowledge translation, systematic review using a realist approach**. *Implementation science : IS* 2011, **6**:25.
2. Best A, Greenhalgh T, Lewis S, Saul JE, Carroll S, Bitz J: **Large-system transformation in health care: a realist review**. *Milbank Q* 2012, **90**(3):421-456.
3. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J: **How outcomes are achieved through patient portals: a realist review**. *Journal of the American Medical Informatics Association : JAMIA* 2014.
4. Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C, Pawson R: **How do you modernize a health service? A realist evaluation of whole-scale transformation in london**. *Milbank Q* 2009, **87**(2):391-416.
5. Doherty C, Stavropoulou C: **Patients' willingness and ability to participate actively in the reduction of clinical errors: A systematic literature review**. *Social science & medicine* 2012, **75**(2):257-263.
6. Marchal B, van Belle S, van Olmen J, Hoerée T, Kegels G: **Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research**. *Evaluation* 2012, **18**(2):192-212.
7. Pawson R: **The science of evaluation: a realist manifesto**: Sage; 2013.
8. Pawson R, Manzano-Santaella A: **A realist diagnostic workshop**. *Evaluation* 2012, **18**(2):176-191.
9. 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.
10. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R: **RAMESES publication standards: meta-narrative reviews**. *BMC Med* 2013, **11**:20.
11. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R: **RAMESES publication standards: realist syntheses**. *BMC Med* 2013, **11**:21.
12. Meads C, Nyssen O, Wong G, Steed L, Bourke L, Ross C, Hayman S, Field V, Lord J, Greenhalgh T: **Protocol for an HTA report: Does therapeutic writing help people with long-term conditions? Systematic review, realist synthesis and economic modelling**. *BMJ Open* 2014, **4**(2):e004377.
13. Barolia RI, Clark AM, Higginbottom GM: **Protocol for a qualitative study on promoting dietary change and positive food choices for poor people with low income who experience cardiovascular disease in Pakistan**. *BMJ Open* 2013, **3**(12):e004176.
14. Rohrbasser A, Mickan S, Harris J: **Exploring why quality circles work in primary health care: a realist review protocol**. *Systematic reviews* 2013, **2**:110.
15. Yardley S, Cottrell E, Protheroe J: **Understanding success and failure in multimorbidity: protocol for using realist synthesis to identify how social learning and workplace practices can be optimised**. *Systematic reviews* 2013, **2**:87.
16. Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Kegels G, Pawson R: **Realist RCTs of complex interventions - An oxymoron**. *Social science & medicine (1982)* 2013.
17. Greenhalgh T, Kristjansson E, Robinson V: **Realist review to understand the efficacy of school feeding programmes**. *BMJ* 2007, **335**(7625):858-861.

18. Hoddinott P, Britten J, Pill R: **Why do interventions work in some places and not others: a breastfeeding support group trial**. *Social science & medicine (1982)* 2010, **70**(5):769-778.
19. Ranmuthugala G, Cunningham FC, Plumb JJ, Long J, Georgiou A, Westbrook JI, Braithwaite J: **A realist evaluation of the role of communities of practice in changing healthcare practice**. *Implementation science : IS* 2011, **6**:49.
20. Cowe A, Cowe M, Goodman C, Kendal S, Mathie E, McNeilly E, Munday D, Peckham S, Polland F, Staniszewska S *et al*: **RAPPORT: ReseArch with Patient and Public invOlvement: A Realist evaluaTion. National Institute for Health Research Health Services and Delivery Research Programme (10/2001/36).** In: *NIHR INVOLVE Conference: 2012; Leeds*: NIHR; 2012.
21. Pawson R, Greenhalgh T, Harvey G, Walshe K: **Realist synthesis: an introduction**. *ESRC Research Methods Programme Manchester: University of Manchester* 2004.
22. Rycroft-Malone J, Wilkinson J, Burton CR, Harvey G, McCormack B, Graham I, Staniszewska S: **Collaborative action around implementation in Collaborations for Leadership in Applied Health Research and Care: towards a programme theory**. *Journal of health services research & policy* 2013, **18**(3 Suppl):13-26.
23. Harvey G, Fitzgerald L, Fielden S, McBride A, Waterman H, Bamford D, Kislov R, Boaden R: **The NIHR Collaborations for Leadership in Applied Health Research and Care (CLAHRC) for Greater Manchester: combining empirical, theoretical and experiential evidence to design and evaluate a large-scale implementation strategy**. *Implementation science : IS* 2011, **6**:96.
24. Rowley E, Morriss R, Currie G, Schneider J: **Research into practice: Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for Nottinghamshire, Derbyshire, Lincolnshire (NDL)**. *Implementation science : IS* 2012, **7**:40.
25. Pawson R, Tilley N: **Realistic evaluation**: Sage; 1997.
26. 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.
27. Pawson R: **Evidence ‐ based policy: A realist perspective.** . London: Sage; 2006.
28. 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.
29. Astbury B, Leeuw FL: **Unpacking black boxes: mechanisms and theory building in evaluation**. *American Journal of Evaluation* 2010, **31**(3):363-381.
30. Shiell A, Hawe P, Gold L: **Complex interventions or complex systems? Implications for health economic evaluation**. *BMJ: British Medical Journal* 2008, **336**(7656):1281.
31. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M: **Developing and evaluating complex interventions: the new Medical Research Council guidance**. *BMJ (Clinical research ed)* 2008, **337**.
32. Rawlins M: **De testimonio: on the evidence for decisions about the use of therapeutic interventions**. *Clinical Medicine* 2008, **8**(6):579-588.
33. Berwick DM: **The science of improvement**. *Jama* 2008, **299**(10):1182-1184.
34. Dalkin SM, Jones D, Lhussier M, Cunningham B: **Understanding integrated care pathways in palliative care using realist evaluation: a mixed methods study protocol**. *BMJ Open* 2012, **2**(4).
35. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: **CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials**. *International journal of surgery (London, England)* 2012, **10**(1):28-55.
36. Collaboration A: **Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project**. *Quality & safety in health care* 2003, **12**(1):18-23.
37. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration**. *BMJ (Clinical research ed)* 2009, **339**:b2700.
38. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney SE: **Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project**. *BMJ* 2009, **338**:a3152.
39. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique**. *Journal of advanced nursing* 2000, **32**(4):1008-1015.
40. 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**. *Quality in health care : QHC* 1999, **8**(1):6-15.

41. 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**. *Medical care* 2003, **41**(12):1374-1381.
42. 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**(4):465-478.
43. Holliday C, Robotin M: **The Delphi process: a solution for reviewing novel grant applications**. *International journal of general medicine* 2010, **3**:225-230.
44. Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E: **Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER)**. *BMC medical research methodology* 2013, **13**:118.
45. Greenhalgh T, Peacock R: **Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources**. *BMJ (Clinical research ed)* 2005, **331**(7524):1064-1065.
46. Greenhalgh T, Wengraf T: **Collecting stories: is it research? Is it good research? Preliminary guidance based on a Delphi study**. *Med Educ* 2008, **42**(3):242-247.
47. Yen HW, G. et al: **How design of places promotes or inhibits mobility of older adults: Realist synthesis of 20 years of research**. *Journal of Ageing and Health* 2014, **in press**.
48. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T: **Risk models and scores for type 2 diabetes: systematic review**. *BMJ* 2011, **343**:d7163.
49. Kristjansson E, Welch V, Greenhalgh T: **Realist review of pre-school feeding programs for disadvantaged children**. *In preparation* 2014.
50. MacDonald M, Jackson A: **Relevance of complexity concepts and systems thinking to public and population health intervention research: A metanarrative synthesis**. *Submitted to Social Science and Medicine* 2014.
51. Bradshaw Bea: **Mining and Aboriginal Community Health: A systemaic review**. *In preparation* 2014.
52. Jagosh J, Pluye P, Wong G, Cargo M, Salsberg J, Bush PL, Herbert CP, Green LW, Greenhalgh T, Macaulay AC: **Critical reflections on realist review: insights from customizing the methodology to the needs of participatory research assessment**. *Research Synthesis Methods* 2013:n/a-n/a.
53. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R: **RAMESES publication standards: meta-narrative reviews**. *Journal of advanced nursing* 2013.
54. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R: **RAMESES publication standards: realist syntheses**. *Journal of advanced nursing* 2013.