

# BMJ Open

## Protocol – The RAMESES II study: Developing guidance and reporting standards for realist evaluation

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2015-008567
Article Type:	Protocol
Date Submitted by the Author:	22-Apr-2015
Complete List of Authors:	Greenhalgh, Trisha; University of Oxford, Nuffield Department of Primary Care Health Sciences Wong, Geoff; University of Oxford, Nuffield Department of Primary Care Health Sciences Jagosh, Justin; University of Liverpool, for Advancement in Realist Evaluation and Syntheses (CARES) Greenhalgh, Joanne; University of Leeds, Sociology and Social Policy Santaella, Ana Manzano; University of Leeds, Sociology and Social Policy Westhorp, Gill; Community Matters, Pawson, Ray; University of Leeds, Sociology and Social Policy
<b>Primary Subject Heading</b>:	Health services research
Secondary Subject Heading:	Research methods
Keywords:	Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MEDICAL JOURNALISM, EDUCATION & TRAINING (see Medical Education & Training)

SCHOLARONE™  
Manuscripts

Original manuscript submitted to BMJ Open April 2015

**Title**

Protocol – The RAMESES II study: Developing guidance and reporting standards for realist evaluation

**Authors**

Trisha Greenhalgh - trish.greenhalgh@phc.ox.ac.uk [1]

Geoff Wong\* - grckwong@gmail.com [1]

Justin Jagosh - Justin.Jagosh@liverpool.ac.uk [2]

Joanne Greenhalgh - J.Greenhalgh@leeds.ac.uk [3]

Ana Manzano - A.Manzano@leeds.ac.uk [3]

Gill Westhorp - gill@communitymatters.com.au [4]

Ray Pawson - R.D.Pawson@leeds.ac.uk [3]

**Authors' affiliations**

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2] Centre for Advancement in Realist Evaluation and Syntheses (CARES), University of Liverpool, Liverpool, United Kingdom

[3] Sociology and Social Policy, University of Leeds, Leeds, United Kingdom

[4] Community Matters, Mt Torrens, South Australia, Australia

Word count: 5200

\*corresponding author

Nuffield Department of Primary Care Health Sciences

New Radcliffe House, Radcliffe Observatory Quarter, Woodstock Road, OX26GG

(T) 01865 289300

Email: grckwong@gmail.com

1  
2  
3 48 **ABSTRACT**

4 49  
5 50 **Introduction**

6  
7  
8 51  
9  
10 52 Realist evaluation is an increasingly popular methodology in health services research. For  
11  
12 53 realist evaluations (RE) this project aims to: develop quality and reporting standards and  
13  
14 54 training materials; build capacity for undertaking and critically evaluating them; produce  
15  
16 55 resources and training materials for lay participants, and those seeking to involve them.  
17  
18  
19  
20 56

21  
22 57 **Methods**

23  
24 58  
25  
26  
27 59 To achieve our aims, we will: [1] Establish management and governance infrastructure; [2]  
28  
29 60 Recruit an interdisciplinary Delphi panel of 35 participants with diverse relevant experience  
30  
31 61 of RE; [3] Summarise current literature and expert opinion on best practice in RE; [4] Run an  
32  
33 62 online Delphi panel to generate and refine items for quality and reporting standards; [5]  
34  
35 63 Capture 'real world' experiences and challenges of RE – e.g. by providing ongoing support to  
36  
37 64 realist evaluations, hosting the RAMESES JISCmail list on realist research, and feeding  
38  
39 65 problems and insights from these into the deliberations of the Delphi panel; [6] Produce  
40  
41 66 quality and reporting standards; [7] Collate examples of the learning and training needs of  
42  
43 67 researchers, students, reviewers and lay members in relation to RE; [8] Develop, deliver and  
44  
45 68 evaluate training materials for RE and deliver training workshops; and [9] Develop and  
46  
47 69 evaluate information and resources for patients and other lay participants in RE (e.g. draft  
48  
49 70 template information sheets and model consent forms) and; [10] Disseminate training  
50  
51 71 materials and other resources.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 73 Planned outputs:

4  
5 74 1. Quality and reporting standards and training materials for RE.

6  
7  
8 75 2. Methodological support for RE.

9  
10 76 3. Increase in capacity to support and evaluate RE.

11  
12 77 4. Accessible, plain-English resources for patients and the public participating in RE.

13  
14  
15 78

## 17 79 **Discussion**

18  
19  
20 80

21  
22 81 The realist evaluation is a relatively new approach to evaluation and its overall place in the

23  
24 82 is not yet fully established. As with all primary research approaches, guidance on quality

25  
26 83 assurance and uniform reporting is an important step towards improving quality and

27  
28  
29 84 consistency.

30  
31  
32 85

33  
34 86 Keywords: Realist evaluation, Realistic evaluation, Quality standards, Reporting standards,

35  
36 87 Delphi method

37  
38  
39 88

40  
41  
42 89

43  
44  
45 90

46  
47  
48 91

49  
50 92

## 51 93 **BACKGROUND**

52  
53  
54 94

## 55 95 **Introduction**

56  
57  
58 96

1  
2  
3 97 Many of the problems confronting researchers today are complex. For example, much  
4  
5 98 health need results from the effects of smoking, suboptimal diets (including obesity),  
6  
7 99 alcohol excess, inactivity or adverse family circumstances (e.g. partner violence) – all of  
8  
9  
10 100 which in turn have multiple causes operating at both individual and societal level.  
11  
12 101 Interventions or programmes designed to tackle such problems are themselves complex,  
13  
14 102 having multiple, interconnected components delivered individually or targeted at  
15  
16  
17 103 communities or populations. Their success depends both on individuals' responses and on  
18  
19 104 the wider context in which people strive (or not) to live meaningful and healthy lives. What  
20  
21 105 works in one family, or one organisation, or one city may not work in another.  
22  
23  
24 106  
25  
26 107 Similarly, the 'wicked problems' of contemporary health services research – how to improve  
27  
28 108 quality and assure patient safety consistently across the service; how to meet rising need  
29  
30 109 from a shrinking budget; and how to realise the potential of information and  
31  
32 110 communication technologies (which often promise more than they deliver) – require  
33  
34 111 complex delivery programmes with multiple, interlocked components that engage with the  
35  
36 112 particularities of context. What works in hospital A may not work in hospital B.  
37  
38  
39 113  
40  
41 114 Designing and evaluating complex interventions is challenging. Randomised trials that  
42  
43 115 compare 'intervention on' with 'intervention off', and their secondary research equivalent,  
44  
45 116 meta-analyses of such trials, may produce statistically accurate but unhelpful statements  
46  
47 117 (e.g. that the intervention works 'on average') which leave us none the wiser about where  
48  
49 118 to target resources or how to maximise impact.  
50  
51  
52  
53  
54  
55 119  
56  
57  
58  
59  
60

1  
2  
3 120 A relatively new approach (especially in health services research) to addressing these  
4  
5 121 problems is realist evaluation. A form of theory-driven evaluation, based on realist  
6  
7 122 philosophy (1), it aims to advance understanding of why these complex interventions work,  
8  
9  
10 123 how, for whom, in what context and to what extent – and also to explain the many  
11  
12 124 situations in which a programme fails to achieve the anticipated benefit.  
13

14  
15 125  
16  
17 126 Realist evaluation assumes both that social systems and structures are 'real' (because they  
18  
19 127 have real effects) and also that human actors respond differently to interventions in  
20  
21 128 different circumstances. To understand how an intervention might generate different  
22  
23 129 outcomes in different circumstances, realism introduces the concept of *mechanisms* –  
24  
25 130 underlying changes in the reasoning and behaviour of participants that are triggered in  
26  
27 131 particular contexts. For example, a school-based feeding programme may work by short-  
28  
29 132 term hunger relief in young children in a low-income rural setting where famine has  
30  
31 133 produced overt nutritional deficiencies, but for teenagers in a troubled inner-city  
32  
33 134 community where many young people are disaffected, it may work chiefly by making pupils  
34  
35 135 feel valued and nurtured (2).  
36  
37  
38  
39  
40  
41  
42

43 137 Realist evaluations have addressed numerous topics of central relevance in health services  
44  
45 138 research, including what works for whom when 'modernising' health services (3),  
46  
47 139 introducing breastfeeding support groups (4), using communities of practice to drive change  
48  
49 140 (5), involving patients and the public in research (6), how robotic surgery impacts on team  
50  
51 141 working and decision making within the operating theatre (7) and fines for delays in  
52  
53 142 discharge from hospitals (8).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1444  
5 145 **What is realist evaluation?**  
67  
8 1469  
10 147 Realist evaluation was developed by Pawson and Tilley in the 1990s, originally in the field of  
11  
12 148 criminology, to address the question “what works for whom in what circumstances and  
13  
14 149 how?” in criminal justice interventions (9). This early work made the following points:

- 15
- 
- 16
- 
- 17 150 • Social programmes (closely akin to what health services researchers call complex
- 
- 18
- 
- 19 151 interventions) are an attempt to address an existing social problem – i.e. to create
- 
- 20
- 
- 21 152 some level of social change.
- 
- 22
- 
- 23 153 • Programmes ‘work’ by enabling participants to make different choices (although
- 
- 24
- 
- 25 154 choice-making is always constrained by such things as participants’ previous
- 
- 26
- 
- 27 155 experiences, beliefs and attitudes, opportunities and access to resources).
- 
- 28
- 
- 29 156 • Making and sustaining different choices requires a change in a participant’s
- 
- 30
- 
- 31 157 reasoning (for example, in their values, beliefs, attitudes, or the logic they apply to a
- 
- 32
- 
- 33 158 particular situation) and/or the resources (e.g. information, skills, material resources,
- 
- 34
- 
- 35 159 support) they have available to them. This combination of ‘reasoning and resources’
- 
- 36
- 
- 37 160 is what enables the programme to ‘work’ and is known as a ‘mechanism’.
- 
- 38
- 
- 39 161 • Programmes ‘work’ in different ways for different people (that is, the contexts within
- 
- 40
- 
- 41 162 programmes can trigger different change mechanisms for different participants).
- 
- 42
- 
- 43 163 • The contexts in which programmes operate make a difference to the outcomes they
- 
- 44
- 
- 45 164 achieve. Programme contexts include features such as social, economic and political
- 
- 46
- 
- 47 165 structures, organizational context, programme participants, programme staffing,
- 
- 48
- 
- 49 166 geographical and historical context, and so on.
- 
- 50
- 
- 51
- 
- 52
- 
- 53
- 
- 54
- 
- 55
- 
- 56
- 
- 57
- 
- 58
- 
- 59
- 
- 60

- 1  
2  
3 167 • Some factors in the context may enable particular mechanisms to be triggered.  
4  
5 168 Other aspects of the context may prevent particular mechanisms from being  
6  
7 169 triggered. That is, there is always an interaction between context and mechanism,  
8  
9  
10 170 and that interaction is what creates the programme's impacts or outcomes: Context  
11  
12 171 + Mechanism = Outcome.  
13  
14  
15 172 • Because programmes work differently in different contexts and through different  
16  
17 173 change mechanisms, programmes cannot simply be replicated from one context to  
18  
19 174 another and automatically achieve the same outcomes. Theory-based  
20  
21  
22 175 understandings about 'what works for whom, in what contexts, and how' are,  
23  
24 176 however, transferable.  
25  
26  
27 177 • Therefore, one of the tasks of evaluation is to learn more about 'what works for  
28  
29 178 whom', 'in which contexts particular programmes do and don't work', and 'what  
30  
31 179 mechanisms are triggered by what programmes in what contexts'.  
32  
33  
34 180  
35  
36 181  
37  
38 182 A realist approach assumes that programmes are 'theories incarnate'. That is, whenever a  
39  
40 183 programme is implemented, it is testing a theory about what 'might cause change', even  
41  
42 184 though that theory may not be explicit. One of the tasks of a realist evaluation is therefore  
43  
44 185 to make the theories within a programme explicit, by developing clear hypotheses about  
45  
46 186 how, and for whom, programmes might 'work'. The implementation of the programme, and  
47  
48 187 the evaluation of it, then tests those hypotheses. This means collecting data, not just about  
49  
50 188 programme impacts, or the processes of programme implementation, but about the specific  
51  
52 189 aspects of programme context that might impact on programme outcomes, and about the  
53  
54 190 specific mechanisms that might be creating change.  
55  
56  
57  
58  
59  
60



1  
2  
3 191  
4  
5 192 Pawson and Tilley also argue that a realist approach has particular implications for the  
6  
7 193 design of an evaluation and the roles of participants. For example, rather than comparing  
8  
9 194 changes for participants who have undertaken a programme with a group of people who  
10  
11 195 have not (as is done in randomised controlled or quasi-experimental designs), a realist  
12  
13 196 evaluation compares context-mechanism-outcome configurations within programmes. It  
14  
15 197 may ask, for example, whether a programme works more or less well, and/or through  
16  
17 198 different mechanisms, in different localities (and if so, how and why); or for different  
18  
19 199 population groups (for example, men and women, or groups with differing socio-economic  
20  
21 200 status). Further, they argue that different stakeholders will have different information and  
22  
23 201 understandings about how programmes are supposed to work and whether they in fact do  
24  
25 202 so. Data collection processes (interviews, focus groups, questionnaires and so on) should be  
26  
27 203 constructed partly to identify the particular information that those stakeholder groups will  
28  
29 204 have, and thereby to refute or refine theories about how and for whom the programme  
30  
31 205 'works'. The philosophical underpinnings of realist evaluation maybe found in Box 1.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

#### 207 **Box 1: The philosophical underpinnings of realist evaluation**

208

44 "Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the  
45 philosophy of science and social science." (10)  
46

47  
48 Philosophically speaking, realism can be thought of as sitting between positivism ('there is a  
49 real external world which we can come to know directly through experiment and  
50 observation') and constructivism ('given that all we can know has been interpreted through  
51 human senses and the human brain, we cannot know for sure what the nature of reality is').  
52 Realism holds that there is a real social world but that our knowledge of it is amassed and  
53 interpreted (sometimes partially and/or imperfectly) via our senses and brains, filtered  
54 through our language, culture and past experience.  
55

56  
57 In other words, realism sees the human agent as suspended in a wider social reality,  
58  
59  
60

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.

209

### 210 **The need for standards and training materials in realist evaluation**

211

212 The RAMESES JISCmail listserv ([www.jiscmail.ac.uk/RAMESES](http://www.jiscmail.ac.uk/RAMESES) - an email list for discussing  
213 realist approaches) postings suggest that enthusiasm for realist evaluation and belief in its  
214 potential for application in many fields have outstripped the development and application  
215 of robust quality standards in the field. Two recent publications have systematically shown  
216 that many so-called 'realist evaluations' were not applying the concepts appropriately and  
217 were (as a result) producing misleading findings and recommendations (11;12) .

218

219 Pawson and Manzano-Santaella in their paper 'A realist diagnostic workshop' used case  
220 examples of flawed realist evaluations to highlight three common errors in such studies  
221 (12). Their paper illustrates three common issues. Firstly, whilst it is possible to show  
222 associations and correlations in data from many types of evaluation, the focus of a realist  
223 evaluation is to explore and explain why such associations occur. Secondly, they explain  
224 what may constitute valid data for use in realist evaluation. Producing a realist explanation  
225 requires a mix of data types, not only qualitative data, to provide explanations and support  
226 for the relationships within and between context mechanisms outcome configurations.  
227 Thirdly, realist explanations require context-mechanism-outcome configurations to be  
228 produced. They note that some realist evaluations have become bogged down in finely

1  
2  
3 229 detailed lists of contexts, mechanisms and outcomes but failed to produce a coherent  
4  
5 230 explanation of how these Cs Ms and Os were linked and related (or not) to each other.  
6  
7  
8 231 Pawson and Manzano-Santaella call for greater emphasis on elucidating programme theory  
9  
10 232 (the theory about what a programme or intervention is expected to do and in some cases,  
11  
12 233 how it is expected to work) expressed as CMO configurations.  
13  
14 234  
15  
16  
17 235 Marchal et al. undertook a review of the realist evaluation literature to quantify and analyse  
18  
19 236 the field (11). They identified 18 realist evaluations and noted a range of challenges that  
20  
21 237 arose for researchers. Absence of prior theoretical and methodological guidance appeared  
22  
23 238 to have led to recurring problems in the realist evaluations they appraised. Firstly, "The  
24  
25 239 philosophical principles that underlie realist evaluation are variably interpreted and applied  
26  
27 240 to different degrees. Most authors only fleetingly refer to the philosophical foundation of  
28  
29 241 realist evaluation, which arguably is among its most distinctive features and provides much  
30  
31 242 of its explanatory power". In addition, they noted that different researchers had  
32  
33 243 conceptualised concepts used in realist evaluation, such as 'middle-range theory',  
34  
35 244 'mechanism' and 'context' differently. This, they concluded, was often related to  
36  
37 245 fundamental misunderstandings. Where misunderstandings occurred, rigour of the realist  
38  
39 246 evaluation undertaken often suffered.  
40  
41 247  
42  
43 248 These two papers show that realist evaluation is often an intellectually challenging task.  
44  
45 249 Both sets of authors point out that more guidance is needed to allay misunderstandings  
46  
47 250 about the purpose, underlying philosophical assumptions and analytic concepts and  
48  
49 251 processes of realist evaluation.  
50  
51  
52  
53  
54  
55  
56  
57 252  
58  
59  
60

253

**254 The online Delphi method**

255

256 To develop our quality and reporting standards we will use the online Delphi method. We  
257 had previously successfully used this method to develop quality and reporting standards and  
258 training materials for meta-narrative reviews and realist syntheses in the RAMESES I (Realist  
259 And Meta-narrative Evidence Syntheses: Evolving Standards) project (13).

260

261 In brief, the essence of the Delphi technique is to engender reflection and discussion  
262 amongst a panel of experts with a view to getting as close as possible to consensus. Both the  
263 agreements reached and the nature and extent of residual disagreement are documented  
264 (14). It was used, for example, to set the original care standards which formed the basis of  
265 the Quality and Outcomes Framework for United Kingdom general practitioners (15). Our  
266 experience and the evidence indicate that the online medium is more likely to improve than  
267 jeopardise the quality of the development process. Delphi panels conducted at a distance  
268 have been shown to be as reliable as face-to-face panels (16) and offer advantages, such as  
269 less cost, speed and greater flexibility for those involved (17). Our experiences of using the  
270 online Delphi method chimes with that of others and indicate that it is the underlying design  
271 and rigour of the Delphi process which is key to quality and not the medium through which  
272 it happens (18;19).

273

274

**275 METHODS / DESIGN**

276

1  
2  
3 277  
4  
5

6 278 **Study design**

7  
8 279  
9

10 280 Mixed-methods study comprising literature review, online Delphi panel, real time  
11  
12 281 engagement with teams undertaking realist evaluations and training workshops (Figure 1).  
13  
14

15 282

16  
17 283  
18

19 284 Figure 1: Study protocol  
20  
21

22 285  
23

24 286  
25

26 287 **Study aims**

27  
28 288  
29

30  
31 289 This project sets out to:  
32

- 33  
34 290 • Develop quality standards, reporting guidance and training materials for realist  
35  
36 291 evaluation  
37  
38 292 • Build capacity for undertaking and critically evaluating realist evaluation in the  
39  
40 293 healthcare context  
41  
42  
43 294 • Produce resources and training materials for lay participants, and those seeking to  
44  
45 295 involve them, in realist evaluations.  
46  
47

48 296  
49

50 297 The project has 10 operational objectives which are described in detail below. The project's  
51  
52 298 10 operational objectives will be delivered in three workstreams, underpinned by a  
53  
54 299 management and governance infrastructure. The detail is set out below.  
55  
56

57 300  
58  
59  
60

1  
2  
3 3014  
5 302 *Objective 1 Establish a management and governance infrastructure, including a project*6  
7 303 *advisory group with lay representation and a patient/service user panel*8  
9 30410  
11 305 A core working group will meet fortnightly, and the advisory group (with lay representation)12  
13 306 and a separate patient / service user panel will each meet 6 monthly. This infrastructure will14  
15 307 advise and support (but not replace) regular meetings among the researchers, as needed, to16  
17 308 execute the study, conduct the data analysis, discuss emerging findings and prepare18  
19 309 outputs.20  
21 31022  
23 311 The project advisory group will have wide cross-sector representation (including experts in24  
25 312 realist evaluation, research support, NHS professionals and representatives from the patient26  
27 313 panel). It will monitor progress against milestones and spend against budget, provide28  
29 314 advice, promote the project, communicate with stakeholders and help maximise30  
31 315 dissemination and impact of findings. In addition, where needed it will act as a sounding32  
33 316 board and 'critical friend' to the project team.34  
35 31736  
37 318 The patient panel will provide advice and feedback to the working group to on how to38  
39 319 present the study and findings in a way that is maximally accessible to lay people.40  
41 320 Representatives from it will attend the project advisory group (with training and support if42  
43 321 required). Where necessary we will provide induction and training to the group members44  
45 322 and ensure that they are made aware that their participation is entirely voluntary and may46  
47 323 withdraw at any time.48  
49 32450  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 325

4  
5 326 Workstream 1 (Objectives 2, 3 and 4)

6  
7 327

8  
9 328 *Objective 2 Recruit an interdisciplinary Delphi panel consisting researchers, people who*

10  
11 329 *support and help design research studies, publishers, peer reviewers,*

12  
13 330 *policymakers, patient advocates and practitioners with (various types of)*

14  
15 331 *experience relevant to realist evaluation.*

16  
17 332

18  
19 333 For the online Delphi panel, we will apply the same successful approach as we did for the

20  
21 334 RAMESES study (13). We will recruit 35 panellists the groups listed in the objective above

22  
23 335 (including patient organisations). Recruitment will be done by the core working group,

24  
25 336 drawing on our knowledge of the field, our different professional networks, the RAMESES

26  
27 337 JISmail listserv and our links to user organisations. Input from a wide range of experts in

28  
29 338 relevant fields will be sought. Those who meet one or more criteria for expertise will be

30  
31 339 briefed on the project, what is expected from them and informed that participation is

32  
33 340 voluntary and unpaid and that they may withdraw at any time. We will ensure

34  
35 341 representation from all relevant stakeholder groups, if necessary by asking existing panel

36  
37 342 members to nominate and invite others.

38  
39 343

40  
41 344

42  
43 345 *Objective 3 Summarise the current literature and expert opinion on best practice in realist*

44  
45 346 *evaluation, to serve as a baseline / briefing document for the panel*

46  
47 347

48  
49

50  
51

52  
53

54  
55  
56  
57  
58  
59  
60

1  
2  
3 348 With expert librarian help, we will identify reviews, scholarly commentaries, models of good  
4  
5 349 practice and examples of (alleged) misapplication of realist evaluation (11;12). To identify  
6  
7  
8 350 the relevant documents we will refine and develop the search used by Marchal et al for a  
9  
10 351 previous review on a similar topic (11), and also apply contemporary search methods  
11  
12 352 designed to identify 'richness' when exploring complex interventions (20;21). We will  
13  
14 353 thematically summarise [a] what is considered by experts to be current best practice (and  
15  
16 354 the range and diversity of such practice); [b] what experts and other researchers believe  
17  
18 355 count as high quality and needs to be reported; and [c] what issues researchers struggle  
19  
20 356 with (based on thematic analysis of postings on the RAMESES JISCmail list archive as well as  
21  
22 357 the published literature). The purpose of this step is not to produce definitive answers to  
23  
24 358 these questions but to prepare a baseline set of briefing materials for the Delphi panel, who  
25  
26 359 will deliberate on them and add to them in the next step.  
27  
28  
29  
30  
31  
32  
33  
34  
35

360

361

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

364

365 The Delphi panel will be run online using SurveyMonkey (Survey Monkey, Palo Alto, CA,  
366 USA). Participants in Round 1 will be provided with briefing materials and invited to suggest  
367 what might be included in the reporting standards. Responses will be analysed and fed into  
368 the design of questionnaire items for round 2.

369

370 In round 2 of the Delphi Panel participants will be asked to rank each potential item twice  
371 on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree), once for relevance (i.e.

372

373

374

375

376

377



1  
2  
3 372 should an item on this theme/topic be included at all in the guidance?) and once for validity  
4  
5 373 (i.e. to what extent do you agree with this item as currently worded?). Those who agreed  
6  
7  
8 374 that an item was relevant, but disagreed on its wording, will be invited to suggest changes  
9  
10 375 to the wording via a free-text comments box. In this second round, participants will again be  
11  
12 376 invited to suggest additional topic areas and items.  
13

14 377  
15  
16  
17 378 Each participant's responses will be collated and the numerical rankings entered onto an  
18  
19 379 Excel spreadsheet. The response rate, average, mode, median and interquartile range for  
20  
21 380 each participant's response to each item will be calculated. Items that score low on  
22  
23 381 relevance will be omitted from subsequent rounds. We will invite further online discussion  
24  
25 382 on items that score high on relevance but low on validity (indicating that a rephrased  
26  
27 383 version of the item may be needed) and on those where there was wide disagreement  
28  
29 384 about relevance or validity. The panel members' free text comments will also be collated  
30  
31 385 and analysed thematically.  
32  
33  
34  
35

36 386  
37  
38 387 Following analysis and discussion within the project team we will then draw up a second list  
39  
40 388 of statements and will be circulated for ranking (round 3). Round 3 will only contain items  
41  
42 389 where consensus has not yet been reached. We plan that the process of collation of  
43  
44 390 responses, further e-mail discussion, and re-ranking will be repeated until a maximum  
45  
46 391 consensus is reached (round 4 et seq.). In practice, very few Delphi panels, online or face to  
47  
48 392 face, go beyond three rounds as participants tend to 'agree to differ' rather than move  
49  
50 393 towards further consensus. We will use email reminders to optimise our response rate from  
51  
52 394 Delphi panel members. We will consider consensus to have been achieved when the median  
53  
54 395 score is 6 or above.  
55  
56  
57  
58  
59  
60

396

397 We plan to report residual non-consensus as such and the nature of the dissent described.

398 Making such dissent explicit tends to expose inherent ambiguities (which may be

399 philosophical or practical) and acknowledges that not everything can be resolved; such

400 findings may be more use to those who use realist evaluation than a firm statement that

401 implies that all tensions have been fixed.

402

403

404

405 Workstream 2 (Objectives 5 and 6)

406

407 *Objective 5 In parallel with the Delphi panel:*

408 *a) Provide ongoing advice and consultancy to up to ten realist evaluations,*  
409 *thereby capturing the 'real world' problems and challenges of this*  
410 *methodology.*

411 *b) Host the RAMESES JISCMail list on realist research, capturing relevant*  
412 *discussions about theoretical, methodological and practical issues.*

413 *c) Feed problems and insights from 5a and 5b into the deliberations of the*  
414 *Delphi panel and the design of training resources and courses*

415

416 We will provide advice and or methodological support to up to 10 realist evaluations. To

417 sample ten that unfold in parallel with our Delphi exercise, we will [i] ask NIHR to link us

418 with planned evaluations funded by them that align with our own timeline; [ii] ask on the

419 RAMESES list; [iii] capture unsolicited requests for help (of which we receive many). We will

1  
2  
3 420 aim for maximum variety in experience of research teams, topics, settings and approach to  
4  
5 421 patient and public involvement. We will work flexibly with teams, mostly by phone, Skype  
6  
7 422 and email, to support them with methodological advice and troubleshooting. We will  
8  
9  
10 423 systematically capture the questions and issues from these ten primary studies and feed  
11  
12 424 them into the deliberations of the Delphi panel (where timings permit) and, if relevant, the  
13  
14 425 training materials and courses described below.  
15  
16

17 426  
18  
19 427 We provided a comparable service to realist review teams in the RAMESES I study, and plan  
20  
21 428 to follow a similar approach (13). In RAMESES I, there was considerable variation in the level  
22  
23 429 of expertise and confidence in the research teams. Some were highly skilled and used our  
24  
25 430 input mainly as 'sounding board' for their own developing ideas and methodology. Others  
26  
27 431 lacked basic understanding of realist concepts and methods; they were offered face-to-face  
28  
29 432 training workshops and bespoke support with data analysis and interpretation. We captured  
30  
31 433 numerous methodological issues that fed into the design of training materials and also  
32  
33 434 informed some methodological papers by our team and the teams we worked with (some of  
34  
35 435 whom have now joined this new collaborative bid) (22;23). We will aim for a similar set of  
36  
37 436 outputs in this work package in RAMESES II.  
38  
39  
40  
41  
42

43 437  
44  
45 438

46  
47 439 *Objective 6 Write up the quality standards and reporting guidelines for an open-access*  
48  
49  
50 440 *journal*  
51

52 441

53  
54  
55 442 We will follow the method applied successfully in RAMESES I to produce an account of the  
56  
57 443 background, methods, main findings and conclusions of the Delphi project, including  
58  
59  
60

1  
2  
3 444 publishing a detailed protocol in an open access journal (24;25) and engaging the editors of  
4  
5 445 specialist journals in potential parallel publication to reach an extended range of readers  
6  
7 446 (26;27). We will also, as in RAMESES I, enter into dialogue with the EQUATOR network  
8  
9  
10 447 (<http://www.equator-network.org>), a clearinghouse for reporting standards which is used as  
11  
12 448 a first port of call by researchers seeking such standards, and which already lists the  
13  
14 449 RAMESES standards for secondary research.  
15

16  
17 450

18  
19 451 Achieving consensus on both quality standards and reporting guidelines may be more  
20  
21 452 difficult for realist evaluation than it was for realist review, since the former covers a huge  
22  
23 453 variety of settings, topics, approaches and configurations (1). Hence it is possible that, unlike  
24  
25 454 in RAMESES I, consensus among Delphi panel members may not be achieved for all items.  
26  
27 455 This is not inherently a problem: in a previous Delphi study to develop standards for  
28  
29 456 undertaking and reporting narrative research, we simply reported, and commented on, the  
30  
31 457 areas of residual disagreement between panel members, which were explained by their  
32  
33 458 different disciplinary and/or sectoral backgrounds (28).  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 459

44 460

45 461 Workstream 3 (Objectives 7-10)

46 462

47 463 *Objective 7 Collate examples of learning / training needs for researchers, postgraduate*  
48  
49 *students, reviewers and lay members in relation to realist evaluation*  
50  
51

52 465

53  
54  
55 466 We will seek examples of the kinds of requests that are made by researchers for support on

56  
57 467 realist evaluation. We already have a rich archive of postings on the RAMESES JISCMail  
58  
59  
60

1  
2  
3 468 listserv from both novice and highly experienced researchers, going back three years. We  
4  
5 469 will also proactively ask the list members for additional examples; use our empirical data  
6  
7 470 from workstream 2 on the real-world struggles of realist researchers (see Objective 5  
8  
9 471 above); and draw on our literature review (Objective 3) and Delphi panel discussions  
10  
11 472 (Objective 4), to identify relevant examples. Finally, we will seek input from UK Research  
12  
13 473 Design Service (RDS) staff, particularly with those who respond to an invitation sent out by  
14  
15 474 the RDS Steering Group on our behalf. We will ask such RDS staff (some of whom are  
16  
17 475 already members of the RAMESES list) to describe the kind of problems people bring to  
18  
19 476 them, and where they feel that further guidance, support and resources are needed.  
20  
21  
22 477  
23  
24 478 We will use a thematic approach to classify examples into a coherent taxonomy of problems  
25  
26 479 and issues, each with a corresponding training need(s). This will be developed iteratively in  
27  
28 480 regular meetings of the research team. At least two researchers will independently classify  
29  
30 481 examples within this taxonomy and through subsequent discussion with the wider team,  
31  
32 482 both the taxonomy and the classification of examples within it will be refined. The goal of  
33  
34 483 this step will be to feed into a coherent and comprehensive curriculum for training realist  
35  
36 484 researchers and for 'training the trainers'.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

487 *Objective 8 Develop, deliver and evaluate training materials for realist evaluation. Deliver 3*  
488 *x 2-day 'realist evaluation' workshops AND 3 x 2-day 'training the trainers'*  
489 *workshops for a range of audiences (including interested NIHR Research Design*  
490 *Service staff)*  
491

1  
2  
3 492 To develop training materials, we will analyse and take forward various problems, issues and  
4  
5 493 learning needs raised in the examples identified in Objective 7. Some will be philosophical or  
6  
7 494 theoretical, some methodological, some practical, some ethical, and so on. Different kinds  
8  
9  
10 495 of learning need require different materials and resources and delivered by different media  
11  
12 496 (face-to-face, internet) and in different learning arrangements (self-study, online drill-and-  
13  
14 497 practice, interactive group tasks and so on). Developing the resources will involve setting  
15  
16  
17 498 specific learning objectives, preparing study notes (e.g. explanations, diagrams) and  
18  
19 499 developing and piloting exercises to engage learners. For each main challenge, we will  
20  
21  
22 500 produce a menu of materials oriented to different audiences and learning styles. Several of  
23  
24 501 the applicants on this bid are experienced trainers and consultants on realist evaluation; we  
25  
26  
27 502 will draw on, and refine, the existing training materials that we have developed and  
28  
29 503 acquired over the years.

30  
31 504

32  
33 505 It is important to stress that realist evaluation cannot be achieved simply by following a  
34  
35 506 protocol in a technically correct manner. Rather, becoming competent at realist evaluation  
36  
37 507 involves acquiring the ability to think, reflect, and interpret data in a way that is resonant  
38  
39 508 with realist philosophy and principles. For this reason, much of the workshops will take the  
40  
41 509 form of “show and tell”, facilitated discussion and “apprenticeship” to experienced and  
42  
43 510 skilled realist researchers.

44  
45  
46  
47 511

48  
49 512 We will run 3 x 2-day ‘how to do a realist evaluation’ workshops for a main audience of  
50  
51 513 researchers and evaluators, and including research users – both lay and professional and 3 x  
52  
53 514 2-day ‘training the trainers in realist research’ workshops for a main audience of those who  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 515 train and support such work. In both sets of workshops, diversity of background will be used  
4  
5 516 productively in group-based case discussions and other hands-on, interactive formats.  
6  
7 517  
8  
9  
10 518 The training the trainers workshops in particular will be open to RDS staff who seek to  
11  
12 519 become confident in supporting realist studies; they will also seek interdisciplinary  
13  
14 520 participation from researchers, practitioners, policymakers and patient advocates. The  
15  
16 521 detailed curriculum for the workshops will emerge from our empirical work, but the training  
17  
18 522 the trainers programme will include all the steps needed to set up and run a responsive  
19  
20 523 service to support and evaluate realist reviews and evaluations, including costing different  
21  
22 524 components of support.  
23  
24  
25  
26  
27  
28  
29  
30

31 527 *Objective 9 Develop, deliver and evaluate information and resources for patients and other*  
32  
33 528 *lay participants in realist evaluation. In particular, draft template information*  
34  
35 529 *sheets and consent forms that could be adapted for ethics and governance*  
36  
37 530 *activity, and deliver up to six workshops for PPI organisations.*  
38  
39  
40  
41  
42

43 532 We will engage with our patient/service user panel to help us develop resources that are  
44  
45 533 relevant, understandable and useful to this group. Examples are: the quality and reporting  
46  
47 534 standards; some of the training resources, especially lay summaries of what a realist  
48  
49 535 evaluation is; template information sheets and consent forms for participants in realist  
50  
51 536 evaluations.  
52  
53  
54

55 537  
56  
57  
58  
59  
60

1  
2  
3 538 As well as developing 'generic' patient/lay resources, we will offer up to six half-day  
4  
5 539 workshops on realist evaluation for patient organisations. We will work with each  
6  
7 540 organisation to develop a curriculum and format. Organisations for these workshops will not  
8  
9 541 be formally sampled as we have found in the past that we receive 'ad hoc' requests for such  
10  
11 542 input, which we often have to turn down because of lack of protected time. Hence this will  
12  
13 543 be a responsive component of the study, dependent on which organisations approach us.  
14  
15 544 Those who do so will probably hear about us from the following sources: [a] the RAMESES  
16  
17 545 listserv, whose membership includes a number of patient/lay advocates; [b] our patient  
18  
19 546 panel and their personal networks; [c] social media invitations (e.g. TG has an active  
20  
21 547 presence on Twitter and more than 10,000 followers, many of whom represent patient  
22  
23 548 organisations); and [d] newsletters and email feeds from organisations such as INVOLVE  
24  
25 549 ([www.invo.org.uk](http://www.invo.org.uk)).

30 550

33 551

36 552 *Objective 10 Disseminate training materials and other resources – e.g. via public access*  
37  
38 553 *websites.*

40 554

43 555 We will replicate the dissemination approach we used for the RAMESES I study, namely: [a]  
44  
45 556 publish the standards in a peer-reviewed journal (in parallel if possible); [b] develop the  
46  
47 557 existing RAMESES project website to host and facilitate open access to all resources; [c]  
48  
49 558 continue to run the RAMESES JISmail list (on which we posted the links to the above); and  
50  
51 559 [d] submit the reporting standards to the EQUATOR NETWORK (an international  
52  
53 560 clearinghouse for peer-reviewed reporting standards, <http://www.equator-network.org>).

56 561



1  
2  
3 562 In addition, we will emphasise the development, piloting and publishing of lay summaries of  
4  
5 563 the key publications. Depending on the journal, it may be possible to publish these lay  
6  
7 564 summaries alongside the academic papers (e.g. New England Journal of Medicine offers  
8  
9  
10 565 such an option). We will make lay summaries available on the RAMESES project website,  
11  
12 566 and will negotiate with COREC (research ethics) and INVOLVE to publish templates of  
13  
14 567 information sheets and consent forms for patient participants in realist evaluation. We will  
15  
16  
17 568 ask the Research Design Service to link to resources relevant to their staff and clients (and  
18  
19 569 have agreement from the RDS to do this in principle).  
20  
21  
22 570  
23  
24 571

## 26 572 **DISCUSSION**

28 573  
29  
30  
31 574 Realist evaluation is a relatively new approach to evaluation, especially in health services  
32  
33 575 research. It potentially offers great promise in unpacking the 'black box' of the many  
34  
35 576 complex interventions that are increasingly being used to improve health and patient  
36  
37 577 outcomes. As relatively experienced users of this approach, we have noted a number of  
38  
39 578 common and recurrent challenges that face grant awarding bodies, peer-reviewers,  
40  
41 579 reviewers and users. These centre on two closely related questions, namely how to judge if  
42  
43 580 a realist evaluation, or a proposal for such an evaluation, is of 'high quality' (including, for  
44  
45 581 completed evaluations, how 'credible' and 'robust' findings are) and how to undertake such  
46  
47 582 evaluations. Our experience to date suggests that we can go a long way towards answering  
48  
49 583 these questions by giving due consideration to the theoretical and conceptual  
50  
51 584 underpinnings of realist evaluation, outlined briefly below.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 586 Realist evaluation is based on a realist philosophy of science, which permeates and informs  
4  
5 587 its underlying epistemological assumptions, methodology and quality considerations. One of  
6  
7  
8 588 the most common misapplications we have noted is that evaluators have not always  
9  
10 589 appreciated the underlying philosophical basis of this approach (and the implications of  
11  
12 590 these for how the evaluation should be conducted). Instead, they have based their  
13  
14 591 evaluations explicitly or implicitly on fundamentally different philosophical assumptions -  
15  
16  
17 592 most commonly the positivist notion that interventions in and of themselves cause  
18  
19 593 outcomes.  
20  
21  
22 594

23  
24 595 Even when a realist philosophy of science has been adhered to in a realist evaluation,  
25  
26 596 reviewers - ourselves included - often struggle with recurring conceptual and  
27  
28 597 methodological issues. 'Mechanisms' present a particular challenge in realist evaluation -  
29  
30 598 how to define them, where to locate them, how to identify them and how to test and refine  
31  
32 599 them (29). Realist evaluation trades on the use of theoretical explanations to make sense of  
33  
34 600 the observed data. Realist evaluators commonly grapple with how to define a theory (what,  
35  
36 601 for example, is the difference between a 'programme theory' and a 'middle-range theory')  
37  
38 602 and what level of abstraction is appropriate in what circumstances. On a more pragmatic  
39  
40 603 level, those who seek to undertake realist evaluations wrestle with a broad range of 'how to'  
41  
42 604 issues: how to produce a programme theory; what type of data needs to be collected; how  
43  
44 605 to use collected data to refine a programme theory; how and to what extent to refine the  
45  
46 606 scope as the evaluation as it unfolds; what changes can legitimately be made to data  
47  
48 607 collected methods; how to organise, analyse and synthesise the collected data; how to  
49  
50 608 make recommendations that are academically defensible and useful to policymakers and  
51  
52 609 the research community; and so on.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 6104  
5 6116  
7  
8 612 **CONCLUSION**9  
10 61311  
12 614 Whilst realist evaluation holds much promise for developing theory and informing policy in13  
14 615 many fields of research, misunderstandings and misapplications of this approach is15  
16 616 common. The time is ripe to start on the iterative journey of producing guidance on quality17  
18 617 and reporting standards as well as developing quality-assured learning resources to ensure19  
20 618 that funding decisions, execution, reporting and use of this evaluation approach is21  
22 619 optimised. Acknowledging that research is never static, the RAMESES II project does not23  
24 620 seek to produce the last word on this topic but to capture current expertise and establish an25  
26 621 agreed 'state of the science' on which future researchers will no doubt build.27  
28 62229  
30 623 We anticipate that the Delphi panel will commence in September 2015 (at the latest) and31  
32 624 that a paper describing the guidance will be submitted by April 2016. The online discussion33  
34 625 forum is open to anyone with an interest in realist evaluation and may be found at35  
36 626 <http://www.jiscmail.ac.uk/RAMESES>.37  
38 62739  
40 62841  
42 62943  
44 63045  
46 63147  
48 63249  
50 63351  
52 63453  
5455  
5657  
5859  
60

1  
2  
3 635 **Competing interests**

4 636 All the authors provide, or may provide, training in realist research and evaluation methods.  
5 637 All the authors work in organisations which tender to undertake realist evaluations or apply  
6 638 for research funding for realist research projects. All the authors have agreed to contribute  
7 639 intellectual property to the project. Products from the project will be provided for free to  
8 640 the international community.  
9 641 The views and opinions expressed therein are those of the authors and do not necessarily  
10 642 reflect those of the United Kingdom's National Institute of Health Research Health Services  
11 643 and Deliver Research (NIHR HS&DR), NIHR, National Health Service (NHS) or the Department  
12 644 of Health.

13 645  
14 646 **Authors' contributions**

15 647 TG conceptualised the study with input from GWo, JJ, AMS, JG, GWe and RP. TG wrote the  
16 648 first draft and GWo, JJ, AMS, JG, GWe and RP critically contributed to and refined this  
17 649 manuscript. TG, GWo, JJ, AMS, JG, GWe and RP have read and approved the final  
18 650 manuscript.  
19 651

20 652 **Acknowledgements and funding**

21 653 This project was funded by the United Kingdom's National Institute of Health Research  
22 654 Health Services and Deliver Research (NIHR HS&DR) Programme (14/19/19).  
23 655 The views and opinions expressed therein are those of the authors and do not necessarily  
24 656 reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.  
25 657

26 658 **Ethics approval**

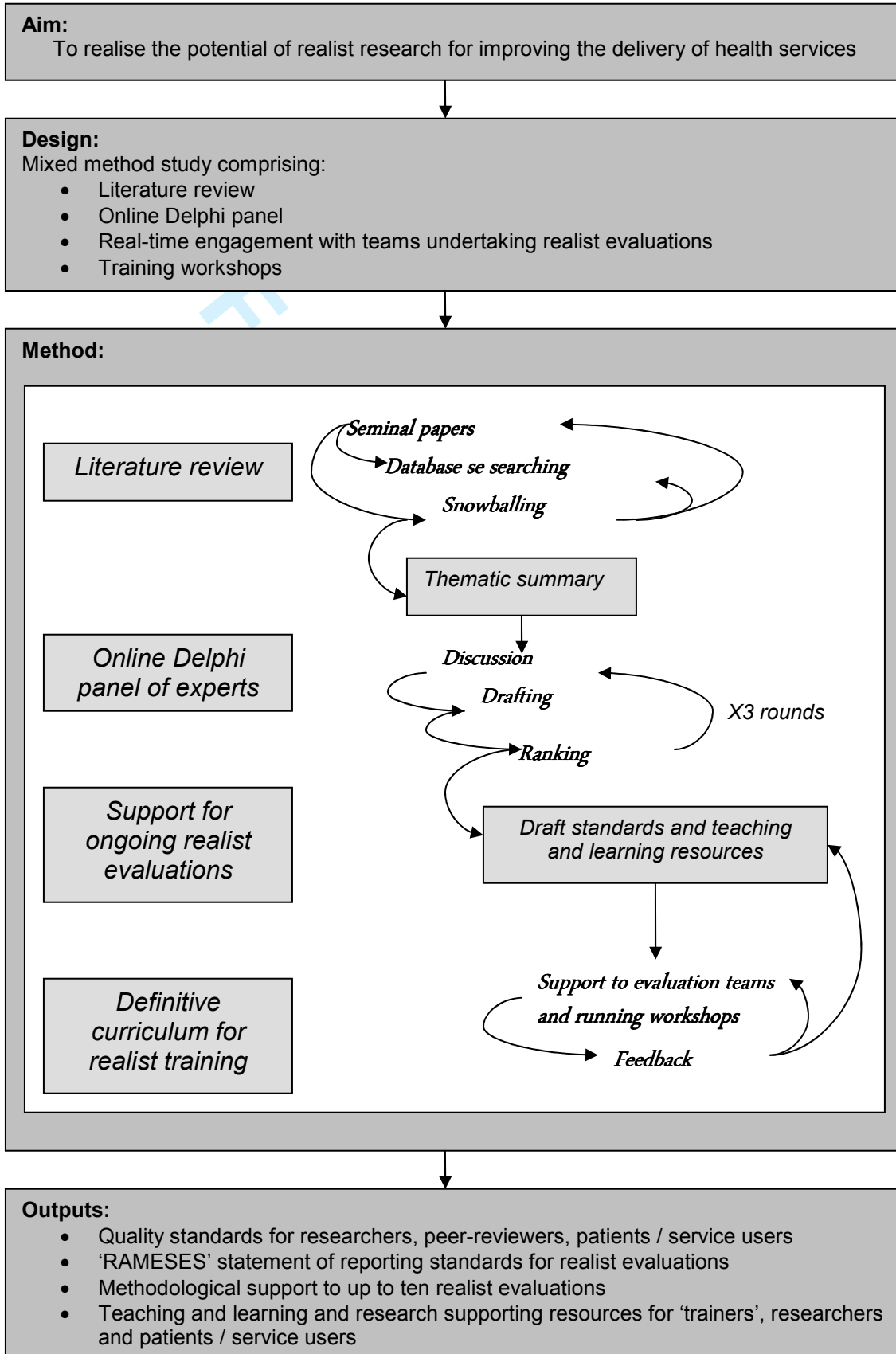
27 659 Ethics approval is currently being sought at the University of Oxford.  
28 660  
29 661  
30 662  
31 663  
32 664

33 665 **REFERENCES**

- 34 666  
35 667  
36 668  
37 669 (1) Pawson R. The Science of Evaluation: A Realist Manifesto. London: Sage; 2013.  
38 670  
39 671 (2) Greenhalgh T, Kristjansson E, Robinson V. Realist review to understand the efficacy  
40 672 of school feeding programmes. *BMJ*. 2007 Oct 27;335(7625):858-61.  
41 673  
42 674 (3) Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C, Pawson R. How do you  
43 675 modernize a health service? A realist evaluation of whole-scale transformation in  
44 676 London. *The Milbank Quarterly*. 2009;87(2):391-416.  
45 677  
46 678 (4) Hoddinott P, Britten J, Pill R. Why do interventions work in some places and not  
47 679 others: a breastfeeding support group trial. *Social Science & Medicine*.  
48 680 2010;70(5):769-78.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 678 (5) Ranmuthugala G, Cunningham F, Plumb J, Long J, Georgiou A, Weatbrook J, et al. A  
4 679 realist evaluation of the role of communities of practice in changing healthcare  
5 680 practice. *Implementation Science*. 2011;6(49).
- 6  
7  
8 681 (6) Cowe A, Cowe M, Goodman C, Kendal S, Mathie E, McNeilly E, et al. RAPPOR:  
9 682 ReseArch with Patient and Public invOLvement: A Realist evaluaTion. NIHR INVOLVE  
10 683 conference. Leeds 2012.
- 11  
12 684 (7) Randell R, Greenhalgh J, Hindmarch J, Dowding D, Jayne D, Pearman A, et al.  
13 685 Integration of robotic surgery into routine practice and impacts on communication,  
14 686 collaboration, and decision making: a realist process evaluation protocol.  
15 687 *Implementation Science*. 2014;9(52).
- 16  
17  
18 688 (8) Manzano-Santaella A. A realistic evaluation of fines for hospital discharges:  
19 689 Incorporating the history of programme evaluations in the analysis. *Evaluation*.  
20 690 2011;17(1):21-36.
- 21  
22 691 (9) Pawson R, Tilley N. *Realistic evaluation*. London: Sage; 1997.
- 23  
24 692 (10) Pawson R. *Evidence-based policy: A realist perspective*. London: Sage; 2006.
- 25  
26 693 (11) Marchal B, van Belle S, van Olmen J, Hoérée T, Kegels G. Is realist evaluation keeping  
27 694 its promise? A review of published empirical studies in the field of health systems  
28 695 research. *Evaluation*. 2012;18(2):192-212.
- 29  
30  
31 696 (12) Pawson R, Manzano-Santaella A. A realist diagnostic workshop. *Evaluation*.  
32 697 2012;18(2):176-91.
- 33  
34 698 (13) Wong G, Greenhalgh T, Westhorp G, Pawson R. Development of methodological  
35 699 guidance, publication standards and training materials for realist and meta-narrative  
36 700 reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses - Evolving  
37 701 Standards) project. *Health Serv Deliv Res*. 2014;2(30).
- 38  
39  
40 702 (14) Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey  
41 703 technique. *J Adv Nurs*. 2000 Oct;32(4):1008-15.
- 42  
43 704 (15) Campbell S, Roland M, Shekelle P, Cantrill J, Buetow S, Cragg D. Development of  
44 705 review criteria for assessing the quality of management of stable angina, adult  
45 706 asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual  
46 707 Health Care*. 1999;8:6-15.
- 47  
48  
49 708 (16) Washington D, Bernstein S, Kahan J, Leape L, Kamberg C, Shekelle P. Reliability of  
50 709 clinical guideline development using mail-only versus inperson expert panels. *Med  
51 710 Care*. 2003;41:1374-81.
- 52  
53  
54 711 (17) Holliday C, Robotin M. The Delphi process: a solution for reviewing novel grant  
55 712 applications. *Int J Gen Med*. 2010;3:225-30.
- 56  
57  
58  
59  
60

- 1  
2  
3 713 (18) Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey  
4 714 technique. *J Adv Nurs*. 2000 Oct;32(4):1008-15.  
5  
6 715 (19) Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the  
7 716 Delphi technique in nursing research. *J Adv Nurs*. 2006 Jan;53(2):205-12.  
8  
9  
10 717 (20) Booth A, Harris J, Crott E, Springett J, Campbell F, Wilkins E. Towards a methodology  
11 718 for cluster searching to provide conceptual and contextual "richness" for systematic  
12 719 reviews of complex interventions: case study (CLUSTER). *BMC Medical Research*  
13 720 *Methodology*. 2013;13(118).  
14  
15 721 (21) Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in  
16 722 systematic reviews of complex evidence: audit of primary sources. *BMJ*. 2005 Nov  
17 723 5;331(7524):1064-5.  
18  
19 724 (22) Jagosh J, Pluye P, Wong G, Cargo M, Salsberg J, Bush P, et al. Critical reflections on  
20 725 realist review: insights from customizing the methodology to the needs of  
21 726 participatory research assessment. *Research Synthesis Methods*. 2014;5(2):131-41.  
22  
23 727 (23) Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Pawson R. Realist RCTs  
24 728 of complex interventions - an oxymoron. *Social Science & Medicine*. 2013;94:124-8.  
25  
26 729 (24) Greenhalgh T, Wong G, Westhorp G, Pawson R. Protocol - Realist and Meta-narrative  
27 730 Evidence Synthesis: Evolving Standards(RAMESES). *BMC Medical Research*  
28 731 *Methodology*. 2011;11:115.  
29  
30 732 (25) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
31 733 realist syntheses. *BMC Medicine*. 2013;11:21.  
32  
33 734 (26) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
34 735 meta-narrative reviews. *Journal of Advanced Nursing*. 2013;69(5):987-1004.  
35  
36 736 (27) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
37 737 realist syntheses. *Journal of Advanced Nursing*. 2013;69(5):1005-22.  
38  
39 738 (28) Greenhalgh T, Wengraf T. Collecting stories: is it research? Is it good research?  
40 739 Preliminary guidance based on a Delphi study. *Med Educ*. 2008 Mar;42(3):242-7.  
41  
42 740 (29) Dalkin S, Greenhalgh J, Jones D, Cunningham B, Lhussier M. What's in a mechanism?  
43 741 Development of a key concept in realist evaluation. *Implementation Science*.  
44 742 Forthcoming 2015.  
45 743  
46 744  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

## Protocol – The RAMESES II study: Developing guidance and reporting standards for realist evaluation

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2015-008567.R1
Article Type:	Protocol
Date Submitted by the Author:	09-Jun-2015
Complete List of Authors:	Greenhalgh, Trish; University of Oxford, Nuffield Department of Primary Care Health Sciences Wong, Geoff; University of Oxford, Nuffield Department of Primary Care Health Sciences Jagosh, Justin; University of Liverpool, for Advancement in Realist Evaluation and Syntheses (CARES) Greenhalgh, Joanne; University of Leeds, Sociology and Social Policy Santaella, Ana Manzano; University of Leeds, Sociology and Social Policy Westhorp, Gill; Community Matters, Pawson, Ray; University of Leeds, Sociology and Social Policy
<b>Primary Subject Heading</b>:	Health services research
Secondary Subject Heading:	Research methods
Keywords:	Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MEDICAL JOURNALISM, EDUCATION & TRAINING (see Medical Education & Training)

SCHOLARONE™  
Manuscripts



Revised manuscript 4<sup>th</sup> June 2015

**Title**

Protocol – The RAMESES II study: Developing guidance and reporting standards for realist evaluation

**Authors**

Trisha Greenhalgh - trish.greenhalgh@phc.ox.ac.uk [1]

Geoff Wong\* - grckwong@gmail.com [1]

Justin Jagosh - Justin.Jagosh@liverpool.ac.uk [2]

Joanne Greenhalgh - J.Greenhalgh@leeds.ac.uk [3]

Ana Manzano - A.Manzano@leeds.ac.uk [3]

Gill Westhorp - gill@communitymatters.com.au [4]

Ray Pawson - R.D.Pawson@leeds.ac.uk [3]

**Authors' affiliations**

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2] Centre for Advancement in Realist Evaluation and Syntheses (CARES), University of Liverpool, Liverpool, United Kingdom

[3] Sociology and Social Policy, University of Leeds, Leeds, United Kingdom

[4] Community Matters, Mt Torrens, South Australia, Australia

Word count: 5200

\*corresponding author

Nuffield Department of Primary Care Health Sciences

New Radcliffe House, Radcliffe Observatory Quarter, Woodstock Road, OX26GG

(T) 01865 289300

Email: grckwong@gmail.com

1  
2  
3 48 **ABSTRACT**

4 49  
5 50 **Introduction**

6  
7  
8 51  
9  
10 52 Realist evaluation is an increasingly popular methodology in health services research. For  
11  
12 53 realist evaluations (RE) this project aims to: develop quality and reporting standards and  
13  
14 54 training materials; build capacity for undertaking and critically evaluating them; produce  
15  
16 55 resources and training materials for lay participants, and those seeking to involve them.  
17  
18  
19  
20 56

21  
22 57 **Methods**

23  
24 58  
25  
26  
27 59 To achieve our aims, we will: [1] Establish management and governance infrastructure; [2]  
28  
29 60 Recruit an interdisciplinary Delphi panel of 35 participants with diverse relevant experience  
30  
31 61 of RE; [3] Summarise current literature and expert opinion on best practice in RE; [4] Run an  
32  
33 62 online Delphi panel to generate and refine items for quality and reporting standards; [5]  
34  
35 63 Capture 'real world' experiences and challenges of RE – e.g. by providing ongoing support to  
36  
37 64 realist evaluations, hosting the RAMESES JISCmail list on realist research, and feeding  
38  
39 65 problems and insights from these into the deliberations of the Delphi panel; [6] Produce  
40  
41 66 quality and reporting standards; [7] Collate examples of the learning and training needs of  
42  
43 67 researchers, students, reviewers and lay members in relation to RE; [8] Develop, deliver and  
44  
45 68 evaluate training materials for RE and deliver training workshops; and [9] Develop and  
46  
47 69 evaluate information and resources for patients and other lay participants in RE (e.g. draft  
48  
49 70 template information sheets and model consent forms) and; [10] Disseminate training  
50  
51 71 materials and other resources.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 73 Planned outputs:

4  
5 74 1. Quality and reporting standards and training materials for RE.

6  
7  
8 75 2. Methodological support for RE.

9  
10 76 3. Increase in capacity to support and evaluate RE.

11  
12 77 4. Accessible, plain-English resources for patients and the public participating in RE.

13  
14  
15 78

## 17 79 **Discussion**

18  
19  
20 80

21  
22 81 The realist evaluation is a relatively new approach to evaluation and its overall place in the

23  
24 82 is not yet fully established. As with all primary research approaches, guidance on quality

25  
26 83 assurance and uniform reporting is an important step towards improving quality and

27  
28  
29 84 consistency.

30  
31  
32 85

33  
34 86 Keywords: Realist evaluation, Realistic evaluation, Quality standards, Reporting standards,

35  
36 87 Delphi method

37  
38  
39 88

40  
41  
42 89

43  
44  
45 90

46  
47  
48 91

49  
50 92

## 51 93 **BACKGROUND**

52  
53 94

## 55 95 **Introduction**

56  
57  
58 96

59  
60

1  
2  
3 97 Many of the problems confronting researchers today are complex. For example, much  
4  
5 98 health need results from the effects of smoking, suboptimal diets (including obesity),  
6  
7 99 alcohol excess, inactivity or adverse family circumstances (e.g. partner violence) – all of  
8  
9  
10 100 which in turn have multiple causes operating at both individual and societal level.  
11  
12 101 Interventions or programmes designed to tackle such problems are themselves complex,  
13  
14 102 having multiple, interconnected components delivered individually or targeted at  
15  
16  
17 103 communities or populations. Their success depends both on individuals' responses and on  
18  
19 104 the wider context in which people strive (or not) to live meaningful and healthy lives. What  
20  
21 105 works in one family, or one organisation, or one city may not work in another.  
22  
23 106  
24  
25  
26 107 Similarly, the 'wicked problems' of contemporary health services research – how to improve  
27  
28 108 quality and assure patient safety consistently across the service; how to meet rising need  
29  
30 109 from a shrinking budget; and how to realise the potential of information and  
31  
32 110 communication technologies (which often promise more than they deliver) – require  
33  
34 111 complex delivery programmes with multiple, interlocked components that engage with the  
35  
36 112 particularities of context. What works in hospital A may not work in hospital B.  
37  
38 113  
39  
40  
41  
42 114 Designing and evaluating complex interventions is challenging. Randomised trials that  
43  
44 115 compare 'intervention on' with 'intervention off', and their secondary research equivalent,  
45  
46 116 meta-analyses of such trials, may produce statistically accurate but unhelpful statements  
47  
48 117 (e.g. that the intervention works 'on average') which leave us none the wiser about where  
49  
50 118 to target resources or how to maximise impact.  
51  
52  
53  
54  
55 119

1  
2  
3 120 A relatively new approach (especially in health services research) to addressing these  
4  
5 121 problems is realist evaluation. A form of theory-driven evaluation, based on realist  
6  
7 122 philosophy (1), it aims to advance understanding of why these complex interventions work,  
8  
9  
10 123 how, for whom, in what context and to what extent – and also to explain the many  
11  
12 124 situations in which a programme fails to achieve the anticipated benefit.  
13

14  
15 125  
16  
17 126 Realist evaluation assumes both that social systems and structures are 'real' (because they  
18  
19 127 have real effects) and also that human actors respond differently to interventions in  
20  
21 128 different circumstances. To understand how an intervention might generate different  
22  
23 129 outcomes in different circumstances, realism introduces the concept of *mechanisms* –  
24  
25 130 underlying changes in the reasoning and behaviour of participants that are triggered in  
26  
27 131 particular contexts. For example, a school-based feeding programme may work by short-  
28  
29 132 term hunger relief in young children in a low-income rural setting where famine has  
30  
31 133 produced overt nutritional deficiencies, but for teenagers in a troubled inner-city  
32  
33 134 community where many young people are disaffected, it may work chiefly by making pupils  
34  
35 135 feel valued and nurtured (2).  
36  
37  
38  
39  
40  
41  
42

43 137 Realist evaluations have addressed numerous topics of central relevance in health services  
44  
45 138 research, including what works for whom when 'modernising' health services (3),  
46  
47 139 introducing breastfeeding support groups (4), using communities of practice to drive change  
48  
49 140 (5), involving patients and the public in research (6), how robotic surgery impacts on team  
50  
51 141 working and decision making within the operating theatre (7) and fines for delays in  
52  
53 142 discharge from hospitals (8).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1444  
5 145 **What is realist evaluation?**  
67  
8 1469  
10 147 Realist evaluation was developed by Pawson and Tilley in the 1990s, originally in the field of  
11  
12 148 criminology, to address the question “what works for whom in what circumstances and  
13  
14 149 how?” in criminal justice interventions (9). This early work made the following points:

- 15
- 
- 16
- 
- 17 150 • Social programmes (closely akin to what health services researchers call complex
- 
- 18
- 
- 19 151 interventions) are an attempt to address an existing social problem – i.e. to create
- 
- 20
- 
- 21 152 some level of social change.
- 
- 22
- 
- 23 153 • Programmes ‘work’ by enabling participants to make different choices (although
- 
- 24
- 
- 25 154 choice-making is always constrained by such things as participants’ previous
- 
- 26
- 
- 27 155 experiences, beliefs and attitudes, opportunities and access to resources).
- 
- 28
- 
- 29 156 • Making and sustaining different choices requires a change in a participant’s
- 
- 30
- 
- 31 157 reasoning (for example, in their values, beliefs, attitudes, or the logic they apply to a
- 
- 32
- 
- 33 158 particular situation) and/or the resources (e.g. information, skills, material resources,
- 
- 34
- 
- 35 159 support) they have available to them. This combination of ‘reasoning and resources’
- 
- 36
- 
- 37 160 is what enables the programme to ‘work’ and is known as a ‘mechanism’.
- 
- 38
- 
- 39 161 • Programmes ‘work’ in different ways for different people (that is, the contexts within
- 
- 40
- 
- 41 162 programmes can trigger different change mechanisms for different participants).
- 
- 42
- 
- 43 163 • The contexts in which programmes operate make a difference to the outcomes they
- 
- 44
- 
- 45 164 achieve. Programme contexts include features such as social, economic and political
- 
- 46
- 
- 47 165 structures, organizational context, programme participants, programme staffing,
- 
- 48
- 
- 49 166 geographical and historical context, and so on.
- 
- 50
- 
- 51
- 
- 52
- 
- 53
- 
- 54
- 
- 55
- 
- 56
- 
- 57
- 
- 58
- 
- 59
- 
- 60

- 1  
2  
3 167 • Some factors in the context may enable particular mechanisms to be triggered.  
4  
5 168 Other aspects of the context may prevent particular mechanisms from being  
6  
7 169 triggered. That is, there is always an interaction between context and mechanism,  
8  
9  
10 170 and that interaction is what creates the programme's impacts or outcomes: Context  
11  
12 171 + Mechanism = Outcome.  
13  
14  
15 172 • Because programmes work differently in different contexts and through different  
16  
17 173 change mechanisms, programmes cannot simply be replicated from one context to  
18  
19 174 another and automatically achieve the same outcomes. Theory-based  
20  
21 175 understandings about 'what works for whom, in what contexts, and how' are,  
22  
23 176 however, transferable.  
24  
25  
26 177 • Therefore, one of the tasks of evaluation is to learn more about 'what works for  
27  
28 178 whom', 'in which contexts particular programmes do and don't work', and 'what  
29  
30 179 mechanisms are triggered by what programmes in what contexts'.  
31  
32  
33  
34 180  
35  
36 181  
37  
38 182 A realist approach assumes that programmes are 'theories incarnate'. That is, whenever a  
39  
40 183 programme is implemented, it is testing a theory about what 'might cause change', even  
41  
42 184 though that theory may not be explicit. One of the tasks of a realist evaluation is therefore  
43  
44 185 to make the theories within a programme explicit, by developing clear hypotheses about  
45  
46 186 how, and for whom, programmes might 'work'. The implementation of the programme, and  
47  
48 187 the evaluation of it, then tests those hypotheses. This means collecting data, not just about  
49  
50 188 programme impacts, or the processes of programme implementation, but about the specific  
51  
52 189 aspects of programme context that might impact on programme outcomes, and about the  
53  
54 190 specific mechanisms that might be creating change.  
55  
56  
57  
58  
59  
60

1  
2  
3 191  
4  
5 192 Pawson and Tilley also argue that a realist approach has particular implications for the  
6  
7 193 design of an evaluation and the roles of participants. For example, rather than comparing  
8  
9  
10 194 changes for participants who have undertaken a programme with a group of people who  
11  
12 195 have not (as is done in randomised controlled or quasi-experimental designs), a realist  
13  
14 196 evaluation compares context-mechanism-outcome configurations within programmes. It  
15  
16  
17 197 may ask, for example, whether a programme works more or less well, and/or through  
18  
19 198 different mechanisms, in different localities (and if so, how and why); or for different  
20  
21 199 population groups (for example, men and women, or groups with differing socio-economic  
22  
23 200 status). Further, they argue that different stakeholders will have different information and  
24  
25 201 understandings about how programmes are supposed to work and whether they in fact do  
26  
27 202 so. Data collection processes (interviews, focus groups, questionnaires and so on) should be  
28  
29 203 constructed partly to identify the particular information that those stakeholder groups will  
30  
31 204 have, and thereby to refute or refine theories about how and for whom the programme  
32  
33 205 ‘works’. The philosophical underpinnings of realist evaluation maybe found in Box 1.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

#### 207 **Box 1: The philosophical underpinnings of realist evaluation**

208

44 “Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the  
45 philosophy of science and social science.” (10)  
46

47  
48 Philosophically speaking, realism can be thought of as sitting between positivism (‘there is a  
49 real external world which we can come to know directly through experiment and  
50 observation’) and constructivism (‘given that all we can know has been interpreted through  
51 human senses and the human brain, we cannot know for sure what the nature of reality is’).  
52 Realism holds that there is a real social world but that our knowledge of it is amassed and  
53 interpreted (sometimes partially and/or imperfectly) via our senses and brains, filtered  
54 through our language, culture and past experience.  
55

56  
57 In other words, realism sees the human agent as suspended in a wider social reality,  
58  
59  
60



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.

209

### 210 **The need for standards and training materials in realist evaluation**

211

212 The RAMESES JISCMail listserv ([www.jiscmail.ac.uk/RAMESES](http://www.jiscmail.ac.uk/RAMESES) - an email list for discussing  
213 realist approaches) postings suggest that enthusiasm for realist evaluation and belief in its  
214 potential for application in many fields have outstripped the development and application  
215 of robust quality standards in the field. Two recent publications have systematically shown  
216 that many so-called 'realist evaluations' were not applying the concepts appropriately and  
217 were (as a result) producing misleading findings and recommendations (11;12) .

218

219 Pawson and Manzano-Santaella in their paper 'A realist diagnostic workshop' used case  
220 examples of flawed realist evaluations to highlight three common errors in such studies  
221 (12). Their paper illustrates three common issues. Firstly, whilst it is possible to show  
222 associations and correlations in data from many types of evaluation, the focus of a realist  
223 evaluation is to explore and explain why such associations occur. Secondly, they explain  
224 what may constitute valid data for use in realist evaluation. Producing a realist explanation  
225 requires a mix of data types, not only qualitative data, to provide explanations and support  
226 for the relationships within and between context mechanisms outcome configurations.  
227 Thirdly, realist explanations require context-mechanism-outcome configurations to be  
228 produced. They note that some realist evaluations have become bogged down in finely

1  
2  
3 229 detailed lists of contexts, mechanisms and outcomes but failed to produce a coherent  
4  
5 230 explanation of how these Cs Ms and Os were linked and related (or not) to each other.  
6  
7  
8 231 Pawson and Manzano-Santaella call for greater emphasis on elucidating programme theory  
9  
10 232 (the theory about what a programme or intervention is expected to do and in some cases,  
11  
12 233 how it is expected to work) expressed as CMO configurations.  
13  
14 234  
15  
16  
17 235 Marchal et al. undertook a review of the realist evaluation literature to quantify and analyse  
18  
19 236 the field (11). They identified 18 realist evaluations and noted a range of challenges that  
20  
21 237 arose for researchers. Absence of prior theoretical and methodological guidance appeared  
22  
23 238 to have led to recurring problems in the realist evaluations they appraised. Firstly, "The  
24  
25 239 philosophical principles that underlie realist evaluation are variably interpreted and applied  
26  
27 240 to different degrees. Most authors only fleetingly refer to the philosophical foundation of  
28  
29 241 realist evaluation, which arguably is among its most distinctive features and provides much  
30  
31 242 of its explanatory power". In addition, they noted that different researchers had  
32  
33 243 conceptualised concepts used in realist evaluation, such as 'middle-range theory',  
34  
35 244 'mechanism' and 'context' differently. This, they concluded, was often related to  
36  
37 245 fundamental misunderstandings. Where misunderstandings occurred, rigour of the realist  
38  
39 246 evaluation undertaken often suffered.  
40  
41 247  
42  
43 248 These two papers show that realist evaluation is often an intellectually challenging task.  
44  
45 249 Both sets of authors point out that more guidance is needed to allay misunderstandings  
46  
47 250 about the purpose, underlying philosophical assumptions and analytic concepts and  
48  
49 251 processes of realist evaluation.  
50  
51  
52  
53  
54  
55  
56  
57 252  
58  
59  
60

1  
2  
3 2534  
5 254 **METHODS / DESIGN**6  
7  
8 2559  
10 25611  
12  
13 257 **Study design**14  
15 25816  
17 259 Mixed-methods study comprising literature review, online Delphi panel, real time  
18 260 engagement with teams undertaking realist evaluations and training workshops (Figure 1).19  
20 26121  
22 26223  
24 263 Figure 1: Study protocol25  
26 26427  
28  
29 265 **The online Delphi method**30  
31 26632  
33 267 To develop our quality and reporting standards we will use the online Delphi method. We  
34 268 had previously successfully used this method to develop quality and reporting standards and  
35 269 training materials for meta-narrative reviews and realist syntheses in the RAMESES I (Realist  
36 270 And Meta-narrative Evidence Syntheses: Evolving Standards) project (13).37  
38 27139 272 In brief, the essence of the Delphi technique is to engender reflection and discussion  
40 273 amongst a panel of experts with a view to getting as close as possible to consensus. Both the  
41 274 agreements reached and the nature and extent of residual disagreement are documented  
42 275 (14). It was used, for example, to set the original care standards which formed the basis of  
43 276 the Quality and Outcomes Framework for United Kingdom general practitioners (15). Our

1  
2  
3 277 experience and the evidence indicate that the online medium is more likely to improve than  
4  
5 278 jeopardise the quality of the development process. Delphi panels conducted at a distance  
6  
7 279 have been shown to be as reliable as face-to-face panels (16) and offer advantages, such as  
8  
9  
10 280 less cost, speed and greater flexibility for those involved (17). Our experiences of using the  
11  
12 281 online Delphi method chimes with that of others and indicate that it is the underlying design  
13  
14 282 and rigour of the Delphi process which is key to quality and not the medium through which  
15  
16 283 it happens (14;18).  
17  
18  
19  
20 284

## 21 285 **Study aims**

22 286

23  
24  
25  
26 287 This project sets out to:

- 27  
28  
29 288 • Develop quality standards, reporting guidance and training materials for realist  
30  
31 289 evaluation  
32  
33 290 • Build capacity for undertaking and critically evaluating realist evaluation in the  
34  
35 291 healthcare context  
36  
37 292 • Produce resources and training materials for lay participants, and those seeking to  
38  
39 293 involve them, in realist evaluations.  
40  
41  
42  
43 294

44  
45 295 The project has 10 operational objectives which are described in detail below. The project's  
46  
47 296 10 operational objectives will be delivered in three workstreams, underpinned by a  
48  
49 297 management and governance infrastructure. The detail is set out below.  
50  
51  
52  
53 298  
54  
55 299

1  
2  
3 300 *Objective 1 Establish a management and governance infrastructure, including a project*  
4  
5 301 *advisory group with lay representation and a patient/service user panel*  
6  
7 302  
8  
9  
10 303 A core working group will meet fortnightly, and the advisory group (with lay representation)  
11  
12 304 and a separate patient / service user panel will each meet 6 monthly. This infrastructure will  
13  
14 305 advise and support (but not replace) regular meetings among the researchers, as needed, to  
15  
16 306 execute the study, conduct the data analysis, discuss emerging findings and prepare  
17  
18 307 outputs.  
19  
20  
21 308  
22  
23 309 The project advisory group will have wide cross-sector representation (including experts in  
24  
25 310 realist evaluation, research support, NHS professionals and representatives from the patient  
26  
27 311 panel). It will monitor progress against milestones and spend against budget, provide  
28  
29 312 advice, promote the project, communicate with stakeholders and help maximise  
30  
31 313 dissemination and impact of findings. In addition, where needed it will act as a sounding  
32  
33 314 board and 'critical friend' to the project team.  
34  
35  
36 315  
37  
38 316 The patient panel will provide advice and feedback to the working group to on how to  
39  
40 317 present the study and findings in a way that is maximally accessible to lay people.  
41  
42 318 Representatives from it will attend the project advisory group (with training and support if  
43  
44 319 required). Where necessary we will provide induction and training to the group members  
45  
46 320 and ensure that they are made aware that their participation is entirely voluntary and may  
47  
48 321 withdraw at any time.  
49  
50  
51 322  
52  
53 323  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 324 Workstream 1 (Objectives 2, 3 and 4)  
4  
5  
6

7 325

8 326 *Objective 2 Recruit an interdisciplinary Delphi panel*  
9

10 327

11 328 For the online Delphi panel, we will apply the same successful approach as we did for the

12 329 RAMESES study (13). We will recruit 35 panellists the groups listed in the objective above

13 330 (including patient organisations). Recruitment will be done by the core working group,

14 331 drawing on our knowledge of the field, our different professional networks, the RAMESES

15 332 JISCmail listserv and our links to user organisations. Input from a wide range of experts in

16 333 relevant fields will be sought, consisting of researchers, people who support and help design

17 334 research studies, publishers, peer reviewers, policymakers, patient advocates and

18 335 practitioners with (various types of) experience relevant to realist evaluation. Those who

19 336 meet one or more criteria for expertise will be briefed on the project, what is expected from

20 337 them and informed that participation is voluntary and unpaid and that they may withdraw

21 338 at any time. We will ensure representation from all relevant stakeholder groups, if

22 339 necessary by asking existing panel members to nominate and invite others.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

40 340

41 341

42 342 *Objective 3 Summarise the current literature and expert opinion on best practice in realist*

43 343 *evaluation, to serve as a baseline / briefing document for the panel*  
44  
45  
46  
47  
48  
49  
50  
51

50 344

51 345 With expert librarian help, we will identify reviews, scholarly commentaries, models of good

52 346 practice and examples of (alleged) misapplication of realist evaluation (11;12). To identify

53 347 the relevant documents we will refine and develop the search used by Marchal et al for a  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 348 previous review on a similar topic (11), and also apply contemporary search methods  
4  
5 349 designed to identify 'richness' when exploring complex interventions (19;20). We will  
6  
7  
8 350 thematically summarise [a] what is considered by experts to be current best practice (and  
9  
10 351 the range and diversity of such practice); [b] what experts and other researchers believe  
11  
12 352 count as high quality and needs to be reported; and [c] what issues researchers struggle  
13  
14 353 with (based on thematic analysis of postings on the RAMESES JISCMail list archive as well as  
15  
16  
17 354 the published literature). The purpose of this step is not to produce definitive answers to  
18  
19 355 these questions but to prepare a baseline set of briefing materials for the Delphi panel, who  
20  
21 356 will deliberate on them and add to them in the next step.  
22  
23  
24 357  
25  
26 358

27  
28  
29 359 *Objective 4 Run three (and more if needed) rounds of the online Delphi panel to generate*  
30  
31 360 *and refine items for a set of quality and reporting standards*  
32

33 361  
34  
35 362 The Delphi panel will be run online using SurveyMonkey (Survey Monkey, Palo Alto, CA,  
36  
37 363 USA). Participants in Round 1 will be provided with briefing materials and invited to suggest  
38  
39 364 what might be included in the reporting standards. Responses will be analysed and fed into  
40  
41 365 the design of questionnaire items for round 2.  
42  
43  
44 366

45  
46  
47 367 In round 2 of the Delphi Panel participants will be asked to rank each potential item twice  
48  
49 368 on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree), once for relevance (i.e.  
50  
51 369 should an item on this theme/topic be included at all in the guidance?) and once for validity  
52  
53 370 (i.e. to what extent do you agree with this item as currently worded?). Those who agreed  
54  
55 371 that an item was relevant, but disagreed on its wording, will be invited to suggest changes  
56  
57  
58  
59  
60

1  
2  
3 372 to the wording via a free-text comments box. In this second round, participants will again be  
4  
5 373 invited to suggest additional topic areas and items.  
6

7  
8 374

9  
10 375 Each participant's responses will be collated and the numerical rankings entered onto an  
11  
12 376 Excel spreadsheet. The response rate, average, mode, median and interquartile range for  
13  
14 377 each participant's response to each item will be calculated. Items that score low on  
15  
16 378 relevance will be omitted from subsequent rounds. We will invite further online discussion  
17  
18 379 on items that score high on relevance but low on validity (indicating that a rephrased  
19  
20 380 version of the item may be needed) and on those where there was wide disagreement  
21  
22 381 about relevance or validity. The panel members' free text comments will also be collated  
23  
24 382 and analysed thematically.  
25  
26  
27  
28

29 383

30  
31 384 Following analysis and discussion within the project team we will then draw up a second list  
32  
33 385 of statements and will be circulated for ranking (round 3). Round 3 will only contain items  
34  
35 386 where consensus has not yet been reached. We plan that the process of collation of  
36  
37 387 responses, further e-mail discussion, and re-ranking will be repeated until a maximum  
38  
39 388 consensus is reached (round 4 et seq.). In practice, very few Delphi panels, online or face to  
40  
41 389 face, go beyond three rounds as participants tend to 'agree to differ' rather than move  
42  
43 390 towards further consensus. We will use email reminders to optimise our response rate from  
44  
45 391 Delphi panel members. We will consider consensus to have been achieved when the median  
46  
47 392 score is 6 or above.  
48  
49  
50

51 393

52  
53  
54 394 We plan to report residual non-consensus as such and the nature of the dissent described.  
55

56  
57 395 Making such dissent explicit tends to expose inherent ambiguities (which may be  
58  
59  
60



1  
2  
3 396 philosophical or practical) and acknowledges that not everything can be resolved; such  
4  
5 397 findings may be more use to those who use realist evaluation than a firm statement that  
6  
7  
8 398 implies that all tensions have been fixed.  
9

10 399

11 400

12 401

13  
14  
15 402 Workstream 2 (Objectives 5 and 6)

16 403

17  
18  
19  
20 404 *Objective 5 In parallel with the Delphi panel:*

21  
22  
23 405 *a) Provide ongoing advice and consultancy to up to ten realist evaluations,*  
24  
25 406 *thereby capturing the 'real world' problems and challenges of this*  
26  
27 407 *methodology.*

28  
29  
30 408 *b) Host the RAMESES JISCmail list on realist research, capturing relevant*  
31  
32 409 *discussions about theoretical, methodological and practical issues.*

33  
34 410 *c) Feed problems and insights from 5a and 5b into the deliberations of the*  
35  
36 411 *Delphi panel and the design of training resources and courses*

37 412

38  
39  
40 413 We will provide advice and or methodological support to up to 10 realist evaluations. To

41  
42 414 sample ten that unfold in parallel with our Delphi exercise, we will [i] ask NIHR to link us

43  
44 415 with planned evaluations funded by them that align with our own timeline; [ii] ask on the

45  
46 416 RAMESES list; [iii] capture unsolicited requests for help (of which we receive many). We will

47  
48 417 aim for maximum variety in experience of research teams, topics, settings and approach to

49  
50 418 patient and public involvement. We will work flexibly with teams, mostly by phone, Skype

51  
52 419 and email, to support them with methodological advice and troubleshooting. We will

53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 420 systematically capture the questions and issues from these ten primary studies and feed  
4  
5 421 them into the deliberations of the Delphi panel (where timings permit) and, if relevant, the  
6  
7 422 training materials and courses described below.  
8  
9

10 423  
11  
12 424 We provided a comparable service to realist review teams in the RAMESES I study, and plan  
13  
14 425 to follow a similar approach (13). In RAMESES I, there was considerable variation in the level  
15  
16 426 of expertise and confidence in the research teams. Some were highly skilled and used our  
17  
18 427 input mainly as 'sounding board' for their own developing ideas and methodology. Others  
19  
20 428 lacked basic understanding of realist concepts and methods; they were offered face-to-face  
21  
22 429 training workshops and bespoke support with data analysis and interpretation. We captured  
23  
24 430 numerous methodological issues that fed into the design of training materials and also  
25  
26 431 informed some methodological papers by our team and the teams we worked with (some of  
27  
28 432 whom have now joined this new collaborative bid) (21;22). We will aim for a similar set of  
29  
30 433 outputs in this work package in RAMESES II.  
31  
32  
33  
34  
35

36 434

37 435

38 436 *Objective 6 Write up the quality standards and reporting guidelines for an open-access*  
39  
40  
41  
42  
43 437 *journal*  
44

45 438

46  
47 439 We will follow the method applied successfully in RAMESES I to produce an account of the  
48  
49 440 background, methods, main findings and conclusions of the Delphi project, including  
50  
51 441 publishing a detailed protocol in an open access journal (23;24) and engaging the editors of  
52  
53 442 specialist journals in potential parallel publication to reach an extended range of readers  
54  
55 443 (25;26). We will also, as in RAMESES I, enter into dialogue with the EQUATOR network  
56  
57  
58  
59  
60

1  
2  
3 444 (<http://www.equator-network.org>), a clearinghouse for reporting standards which is used as  
4  
5 445 a first port of call by researchers seeking such standards, and which already lists the  
6  
7 446 RAMESES standards for secondary research.  
8  
9

10 447

11  
12 448 Achieving consensus on both quality standards and reporting guidelines may be more  
13  
14 449 difficult for realist evaluation than it was for realist review, since the former covers a huge  
15  
16 450 variety of settings, topics, approaches and configurations (1). Hence it is possible that, unlike  
17  
18 451 in RAMESES I, consensus among Delphi panel members may not be achieved for all items.  
19  
20 452 This is not inherently a problem: in a previous Delphi study to develop standards for  
21  
22 453 undertaking and reporting narrative research, we simply reported, and commented on, the  
23  
24 454 areas of residual disagreement between panel members, which were explained by their  
25  
26 455 different disciplinary and/or sectoral backgrounds (27).  
27  
28  
29  
30

31 456

32 457

33  
34  
35  
36 458 Workstream 3 (Objectives 7-10)

37 459

38  
39  
40 460 *Objective 7 Collate examples of learning / training needs for researchers, postgraduate*  
41  
42  
43 461 *students, reviewers and lay members in relation to realist evaluation*  
44

45 462

46  
47 463 We will seek examples of the kinds of requests that are made by researchers for support on  
48  
49 464 realist evaluation. We already have a rich archive of postings on the RAMESES JISCmail  
50  
51 465 listserv from both novice and highly experienced researchers, going back three years. We  
52  
53 466 will also proactively ask the list members for additional examples; use our empirical data  
54  
55 467 from workstream 2 on the real-world struggles of realist researchers (see Objective 5  
56  
57  
58  
59  
60

1  
2  
3 468 above); and draw on our literature review (Objective 3) and Delphi panel discussions  
4  
5 469 (Objective 4), to identify relevant examples. Finally, we will seek input from UK Research  
6  
7  
8 470 Design Service (RDS) staff, particularly with those who respond to an invitation sent out by  
9  
10 471 the RDS Steering Group on our behalf. We will ask such RDS staff (some of whom are  
11  
12 472 already members of the RAMESES list) to describe the kind of problems people bring to  
13  
14 473 them, and where they feel that further guidance, support and resources are needed.  
15  
16  
17 474

18  
19 475 We will use a thematic approach to classify examples into a coherent taxonomy of problems  
20  
21 476 and issues, each with a corresponding training need(s). This will be developed iteratively in  
22  
23 477 regular meetings of the research team. At least two researchers will independently classify  
24  
25 478 examples within this taxonomy and through subsequent discussion with the wider team,  
26  
27 479 both the taxonomy and the classification of examples within it will be refined. The goal of  
28  
29 480 this step will be to feed into a coherent and comprehensive curriculum for training realist  
30  
31 481 researchers and for 'training the trainers'.  
32  
33  
34  
35

36 482

37 483

38 484 *Objective 8 Develop, deliver and evaluate training materials for realist evaluation. Deliver 3*  
39  
40  
41  
42  
43 485 *x 2-day 'realist evaluation' workshops AND 3 x 2-day 'training the trainers'*  
44  
45 486 *workshops for a range of audiences (including interested NIHR Research Design*  
46  
47 487 *Service staff)*  
48  
49

50 488

51  
52 489 To develop training materials, we will analyse and take forward various problems, issues and  
53  
54 490 learning needs raised in the examples identified in Objective 7. Some will be philosophical or  
55  
56 491 theoretical, some methodological, some practical, some ethical, and so on. Different kinds  
57  
58  
59  
60

1  
2  
3 492 of learning need require different materials and resources and delivered by different media  
4  
5 493 (face-to-face, internet) and in different learning arrangements (self-study, online drill-and-  
6  
7  
8 494 practice, interactive group tasks and so on). Developing the resources will involve setting  
9  
10 495 specific learning objectives, preparing study notes (e.g. explanations, diagrams) and  
11  
12 496 developing and piloting exercises to engage learners. For each main challenge, we will  
13  
14 497 produce a menu of materials oriented to different audiences and learning styles. Several of  
15  
16 498 the applicants on this bid are experienced trainers and consultants on realist evaluation; we  
17  
18 499 will draw on, and refine, the existing training materials that we have developed and  
19  
20  
21  
22 500 acquired over the years.  
23

501

24  
25  
26 502 It is important to stress that realist evaluation cannot be achieved simply by following a  
27  
28 503 protocol in a technically correct manner. Rather, becoming competent at realist evaluation  
29  
30 504 involves acquiring the ability to think, reflect, and interpret data in a way that is resonant  
31  
32 505 with realist philosophy and principles. For this reason, much of the workshops will take the  
33  
34 506 form of “show and tell”, facilitated discussion and “apprenticeship” to experienced and  
35  
36 507 skilled realist researchers.  
37  
38  
39

508

40  
41  
42  
43 509 We will run 3 x 2-day ‘how to do a realist evaluation’ workshops for a main audience of  
44  
45 510 researchers and evaluators, and including research users – both lay and professional and 3 x  
46  
47 511 2-day ‘training the trainers in realist research’ workshops for a main audience of those who  
48  
49 512 train and support such work. In both sets of workshops, diversity of background will be used  
50  
51 513 productively in group-based case discussions and other hands-on, interactive formats.  
52  
53  
54

514

1  
2  
3 515 The training the trainers workshops in particular will be open to RDS staff who seek to  
4  
5 516 become confident in supporting realist studies; they will also seek interdisciplinary  
6  
7 517 participation from researchers, practitioners, policymakers and patient advocates. The  
8  
9 518 detailed curriculum for the workshops will emerge from our empirical work, but the training  
10  
11 519 the trainers programme will include all the steps needed to set up and run a responsive  
12  
13 520 service to support and evaluate realist reviews and evaluations, including costing different  
14  
15 521 components of support.  
16  
17  
18  
19  
20  
21  
22  
23

522

523

24 524 *Objective 9 Develop, deliver and evaluate information and resources for patients and other*  
25  
26 525 *lay participants in realist evaluation. In particular, draft template information*  
27  
28 526 *sheets and consent forms that could be adapted for ethics and governance*  
29  
30 527 *activity, and deliver up to six workshops for PPI organisations.*  
31  
32  
33

528

34  
35  
36 529 We will engage with our patient/service user panel to help us develop resources that are  
37  
38 530 relevant, understandable and useful to this group. Examples are: the quality and reporting  
39  
40 531 standards; some of the training resources, especially lay summaries of what a realist  
41  
42 532 evaluation is; template information sheets and consent forms for participants in realist  
43  
44 533 evaluations.  
45  
46  
47

534

48  
49  
50 535 As well as developing 'generic' patient/lay resources, we will offer up to six half-day  
51  
52 536 workshops on realist evaluation for patient organisations. We will work with each  
53  
54 537 organisation to develop a curriculum and format. Organisations for these workshops will not  
55  
56 538 be formally sampled as we have found in the past that we receive 'ad hoc' requests for such  
57  
58  
59  
60

1  
2  
3 539 input, which we often have to turn down because of lack of protected time. Hence this will  
4  
5 540 be a responsive component of the study, dependent on which organisations approach us.  
6  
7 541 Those who do so will probably hear about us from the following sources: [a] the RAMESES  
8  
9 542 listserv, whose membership includes a number of patient/lay advocates; [b] our patient  
10  
11 543 panel and their personal networks; [c] social media invitations (e.g. TG has an active  
12  
13 544 presence on Twitter and more than 10,000 followers, many of whom represent patient  
14  
15 545 organisations); and [d] newsletters and email feeds from organisations such as INVOLVE  
16  
17 546 (www.invo.org.uk).  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

551  
552 We will replicate the dissemination approach we used for the RAMESES I study, namely: [a]  
553 publish the standards in a peer-reviewed journal (in parallel if possible); [b] develop the  
554 existing RAMESES project website to host and facilitate open access to all resources; [c]  
555 continue to run the RAMESES JISmail list (on which we posted the links to the above); and  
556 [d] submit the reporting standards to the EQUATOR NETWORK (an international  
557 clearinghouse for peer-reviewed reporting standards, <http://www.equator-network.org>).  
558

559 In addition, we will emphasise the development, piloting and publishing of lay summaries of  
560 the key publications. Depending on the journal, it may be possible to publish these lay  
561 summaries alongside the academic papers (e.g. New England Journal of Medicine offers  
562 such an option). We will make lay summaries available on the RAMESES project website,

1  
2  
3 563 and will negotiate with COREC (research ethics) and INVOLVE to publish templates of  
4  
5 564 information sheets and consent forms for patient participants in realist evaluation. We will  
6  
7  
8 565 ask the Research Design Service to link to resources relevant to their staff and clients (and  
9  
10 566 have agreement from the RDS to do this in principle).

11  
12 567

13  
14 568

15  
16  
17 569 **DISCUSSION**

18  
19  
20 570

21  
22 571 Realist evaluation is a relatively new approach to evaluation, especially in health services  
23  
24 572 research. It potentially offers great promise in unpacking the 'black box' of the many  
25  
26 573 complex interventions that are increasingly being used to improve health and patient  
27  
28 574 outcomes. As relatively experienced users of this approach, we have noted a number of  
29  
30 575 common and recurrent challenges that face grant awarding bodies, peer-reviewers,  
31  
32 576 reviewers and users. These centre on two closely related questions, namely how to judge if  
33  
34 577 a realist evaluation, or a proposal for such an evaluation, is of 'high quality' (including, for  
35  
36 578 completed evaluations, how 'credible' and 'robust' findings are) and how to undertake such  
37  
38 579 evaluations. Our experience to date suggests that we can go a long way towards answering  
39  
40 580 these questions by giving due consideration to the theoretical and conceptual  
41  
42 581 underpinnings of realist evaluation, outlined briefly below.

43  
44 582

45  
46 583 Realist evaluation is based on a realist philosophy of science, which permeates and informs  
47  
48 584 its underlying epistemological assumptions, methodology and quality considerations. One of  
49  
50 585 the most common misapplications we have noted is that evaluators have not always  
51  
52 586 appreciated the underlying philosophical basis of this approach (and the implications of  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 587 these for how the evaluation should be conducted). Instead, they have based their  
4  
5 588 evaluations explicitly or implicitly on fundamentally different philosophical assumptions -  
6  
7  
8 589 most commonly the positivist notion that interventions in and of themselves cause  
9  
10 590 outcomes.

11  
12 591  
13  
14 592 Even when a realist philosophy of science has been adhered to in a realist evaluation,  
15  
16  
17 593 reviewers - ourselves included - often struggle with recurring conceptual and  
18  
19  
20 594 methodological issues. 'Mechanisms' present a particular challenge in realist evaluation -  
21  
22 595 how to define them, where to locate them, how to identify them and how to test and refine  
23  
24 596 them (28). Realist evaluation trades on the use of theoretical explanations to make sense of  
25  
26  
27 597 the observed data. Realist evaluators commonly grapple with how to define a theory (what,  
28  
29 598 for example, is the difference between a 'programme theory' and a 'middle-range theory'?)  
30  
31 599 and what level of abstraction is appropriate in what circumstances. On a more pragmatic  
32  
33  
34 600 level, those who seek to undertake realist evaluations wrestle with a broad range of 'how to'  
35  
36 601 issues: how to produce a programme theory; what type of data needs to be collected; how  
37  
38 602 to use collected data to refine a programme theory; how and to what extent to refine the  
39  
40  
41 603 scope as the evaluation as it unfolds; what changes can legitimately be made to data  
42  
43 604 collected methods; how to organise, analyse and synthesise the collected data; how to  
44  
45 605 make recommendations that are academically defensible and useful to policymakers and  
46  
47  
48 606 the research community; and so on.

49  
50 607  
51  
52 608 As we have mentioned above, realist evaluation is a relatively new approach and so we are  
53  
54  
55 609 aware that methodological development is very likely to occur. Realist evaluation as an  
56  
57 610 approach has also been used in a wide range of disciplines – both in and outside of health.  
58  
59  
60

1  
2  
3 611 These two issues will have a significant impact on the RAMESES II project and have already  
4  
5 612 been debated and discussed since the start of the project. We want to ensure that the  
6  
7 613 project's outputs do not stifle innovation and methodological development in realist  
8  
9 614 evaluation. To address this issue we can draw on our experience in developing similar  
10  
11 615 resources in the first RAMESES project (13). For example, in the quality and publication  
12  
13 616 standards we produced for the first RAMESES project, we deliberately stated that  
14  
15 617 researchers were able to make any changes they felt were needed to a review's processes,  
16  
17 618 but should explain what were the changes, where and why they had been made. To address  
18  
19 619 the issue of the wide range of disciplines that use realist evaluation, we will draw on realist  
20  
21 620 and/or evaluation expertise from a broad range of disciplines for our Delphi panel. This will  
22  
23 621 ensure that there is not just one dominant 'voice' (e.g. from health researchers) on the  
24  
25 622 panel, thus enabling any of the project's outputs to be suitable for use in a wide range of  
26  
27 623 circumstances.  
28  
29  
30  
31  
32  
33  
34  
35

## 625 **CONCLUSION**

36  
37  
38 626  
39  
40  
41 627 Whilst realist evaluation holds much promise for developing theory and informing policy in  
42  
43 628 many fields of research, misunderstandings and misapplications of this approach is  
44  
45 629 common. The time is ripe to start on the iterative journey of producing guidance on quality  
46  
47 630 and reporting standards as well as developing quality-assured learning resources to ensure  
48  
49 631 that funding decisions, execution, reporting and use of this evaluation approach is  
50  
51 632 optimised. Acknowledging that research is never static, the RAMESES II project does not  
52  
53 633 seek to produce the last word on this topic but to capture current expertise and establish an  
54  
55 634 agreed 'state of the science' on which future researchers will no doubt build.  
56  
57  
58  
59  
60

635

636 We anticipate that the Delphi panel will commence in September 2015 (at the latest) and

637 that a paper describing the guidance will be submitted by April 2016. The online discussion

638 forum is open to anyone with an interest in realist evaluation and may be found at

639 <http://www.jiscmail.ac.uk/RAMESES>.

640

641

642

643

644

645

646

647

For peer review only

1  
2  
3 648 **Competing interests**

4 649 All the authors provide, or may provide, training in realist research and evaluation methods.  
5 650 All the authors work in organisations which tender to undertake realist evaluations or apply  
6 651 for research funding for realist research projects. All the authors have agreed to contribute  
7 652 intellectual property to the project. Products from the project will be provided for free to  
8 653 the international community.

9 654 The views and opinions expressed therein are those of the authors and do not necessarily  
10 655 reflect those of the United Kingdom's National Institute of Health Research Health Services  
11 656 and Deliver Research (NIHR HS&DR), NIHR, National Health Service (NHS) or the Department  
12 657 of Health.

13 658  
14 659 **Authors' contributions**

15 660 TG conceptualised the study with input from GWo, JJ, AMS, JG, GWe and RP. TG wrote the  
16 661 first draft and GWo, JJ, AMS, JG, GWe and RP critically contributed to and refined this  
17 662 manuscript. TG, GWo, JJ, AMS, JG, GWe and RP have read and approved the final  
18 663 manuscript.

19 664  
20 665 **Acknowledgements and funding**

21 666 This project was funded by the United Kingdom's National Institute of Health Research  
22 667 Health Services and Deliver Research (NIHR HS&DR) Programme (14/19/19).

23 668 The views and opinions expressed therein are those of the authors and do not necessarily  
24 669 reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.

25 670  
26 671 **Ethics approval**

27 672 Ethics clearance has been granted by the relevant committee at the University of Oxford.  
28 673  
29 674  
30 675  
31 676  
32 677

33 678 **REFERENCES**

- 34 679  
35 680  
36 681  
37 682  
38 683 (1) Pawson R. The Science of Evaluation: A Realist Manifesto. London: Sage, 2013.  
39 684  
40 685 (2) Greenhalgh T, Kristjansson E, Robinson V. Realist review to understand the efficacy  
41 686 of school feeding programmes. *BMJ* 2007; 335(7625):858-861.  
42 687  
43 688 (3) Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C, Pawson R. How do you  
44 689 modernize a health service? A realist evaluation of whole-scale transformation in  
45 690 London. *The Milbank Quarterly* 2009; 87(2):391-416.  
46 691  
47 689 (4) Hoddinott P, Britten J, Pill R. Why do interventions work in some places and not  
48 690 others: a breastfeeding support group trial. *Social Science & Medicine* 2010;  
49 691 70(5):769-778.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 692 (5) Ranmuthugala G, Cunningham F, Plumb J, Long J, Georgiou A, Weatbrook J et al. A  
4 693 realist evaluation of the role of communities of practice in changing healthcare  
5 694 practice. *Implementation Science* 2011; 6(49).
- 7 695 (6) RAPPORT: ReseArch with Patient and Public invOLvement: A Realist evaluaTion.  
8 696 NIHR INVOLVE conference; Leeds: 2012.
- 11 697 (7) Randell R, Greenhalgh J, Hindmarch J, Dowding D, Jayne D, Pearman A et al.  
12 698 Integration of robotic surgery into routine practice and impacts on communication,  
13 699 collaboration, and decision making: a realist process evaluation protocol.  
14 700 *Implementation Science* 2014; 9(52).
- 17 701 (8) Manzano-Santaella A. A realistic evaluation of fines for hospital discharges:  
18 702 Incorporating the history of programme evaluations in the analysis. *Evaluation* 2011;  
19 703 17(1):21-36.
- 21 704 (9) Pawson R, Tilley N. *Realistic evaluation*. London: Sage, 1997.
- 23 705 (10) Pawson R. *Evidence-based policy: A realist perspective*. London: Sage, 2006.
- 25 706 (11) Marchal B, van Belle S, van Olmen J, Hoérée T, Kegels G. Is realist evaluation keeping  
26 707 its promise? A review of published empirical studies in the field of health systems  
27 708 research. *Evaluation* 2012; 18(2):192-212.
- 30 709 (12) Pawson R, Manzano-Santaella A. A realist diagnostic workshop. *Evaluation* 2012;  
31 710 18(2):176-191.
- 33 711 (13) Wong G, Greenhalgh T, Westhorp G, Pawson R. Development of methodological  
34 712 guidance, publication standards and training materials for realist and meta-narrative  
35 713 reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses - Evolving  
36 714 Standards) project. *Health Serv Deliv Res* 2014; 2(30).
- 39 715 (14) Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey  
40 716 technique. *J Adv Nurs* 2000; 32(4):1008-1015.
- 42 717 (15) Campbell S, Roland M, Shekelle P, Cantrill J, Buetow S, Cragg D. Development of  
43 718 review criteria for assessing the quality of management of stable angina, adult  
44 719 asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual*  
45 720 *Health Care* 1999; 8:6-15.
- 48 721 (16) Washington D, Bernstein S, Kahan J, Leape L, Kamberg C, Shekelle P. Reliability of  
49 722 clinical guideline development using mail-only versus inperson expert panels. *Med*  
50 723 *Care* 2003; 41:1374-1381.
- 52 724 (17) Holliday C, Robotin M. The Delphi process: a solution for reviewing novel grant  
53 725 applications. *Int J Gen Med* 2010; 3:225-230.
- 55 726 (18) Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the  
56 727 Delphi technique in nursing research. *J Adv Nurs* 2006; 53(2):205-212.

- 1  
2  
3 728 (19) Booth A, Harris J, Crott E, Springett J, Campbell F, Wilkins E. Towards a methodology  
4 729 for cluster searching to provide conceptual and contextual "richness" for systematic  
5 730 reviews of complex interventions: case study (CLUSTER). BMC Medical Research  
6 731 Methodology 2013; 13(118).
- 8  
9 732 (20) Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in  
10 733 systematic reviews of complex evidence: audit of primary sources. BMJ 2005;  
11 734 331(7524):1064-1065.
- 13 735 (21) Jagosh J, Pluye P, Wong G, Cargo M, Salsberg J, Bush P et al. Critical reflections on  
14 736 realist review: insights from customizing the methodology to the needs of  
15 737 participatory research assessment. Research Synthesis Methods 2014; 5(2):131-141.
- 17 738 (22) Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Pawson R. Realist RCTs  
18 739 of complex interventions - an oxymoron. Social Science & Medicine 2013; 94:124-  
19 740 128.
- 21 741 (23) Greenhalgh T, Wong G, Westhorp G, Pawson R. Protocol - Realist and Meta-narrative  
22 742 Evidence Synthesis: Evolving Standards(RAMESES). BMC Medical Research  
23 743 Methodology 2011; 11:115.
- 25 744 (24) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
26 745 realist syntheses. BMC Medicine 2013; 11:21.
- 28 746 (25) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
29 747 meta-narrative reviews. Journal of Advanced Nursing 2013; 69(5):987-1004.
- 31 748 (26) Wong G, Greenhalgh T, Westhorp G, Pawson R. RAMESES publication standards:  
32 749 realist syntheses. Journal of Advanced Nursing 2013; 69(5):1005-1022.
- 34 750 (27) Greenhalgh T, Wengraf T. Collecting stories: is it research? Is it good research?  
35 751 Preliminary guidance based on a Delphi study. Med Educ 2008; 42(3):242-247.
- 37 752 (28) Dalkin S, Greenhalgh J, Jones D, Cunningham B, Lhussier M. What's in a mechanism?  
38 753 Development of a key concept in realist evaluation. Implementation Science 2015;  
39 754 10(49).

40  
41  
42  
43  
44 755  
45 756  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

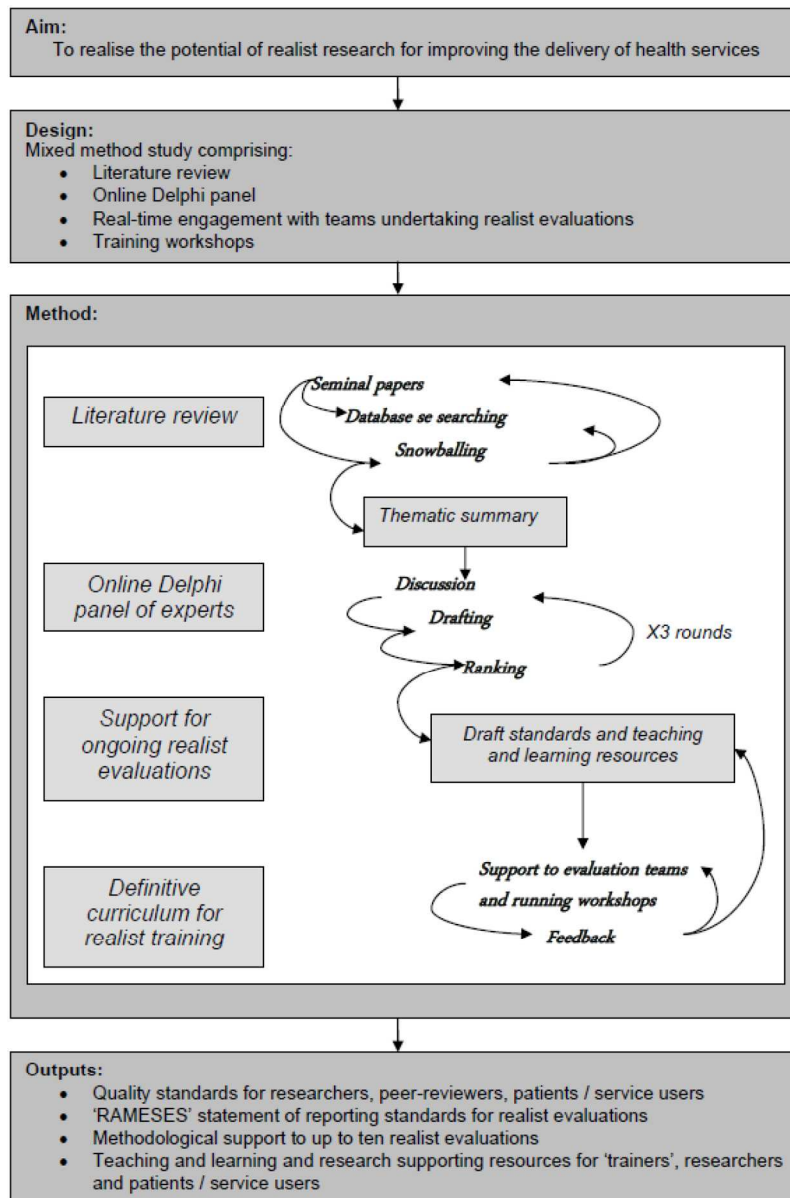


Figure 1: Study protocol  
132x198mm (300 x 300 DPI)