

Developing consensus based quality and reporting standards and resources for realist evaluation: The RAMESES II project

Geoff Wong* [1], Gill Westhorp [2], Joanne Greenhalgh [3], Ana Manzano [3], Justin Jagosh [4], Trisha Greenhalgh [1]

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

[2] Realist Research Evaluation and Learning Initiative, Charles Darwin University, Darwin, Australia

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

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

Important

A 'first look' scientific summary is created from the original author-supplied summary once the normal NIHR Journals Library peer and editorial review processes are complete. The summary has undergone full peer and editorial review as documented at NIHR Journals Library website and may undergo rewrite during the publication process. The order of authors was correct at editorial sign-off stage.

A final version (which has undergone a rigorous copy-edit and proofreading) will publish as part of a fuller account of the research in a forthcoming issue of the Health Services and Delivery Research journal.

Any queries about this 'first look' version of the scientific summary should be addressed to the NIHR Journals Library Editorial Office – journals.library@nihr.ac.uk

The research reported in this 'first look' scientific summary was funded by the HS&DR programme or one of its predecessor programmes (NIHR Service Delivery and Organisation programme, or Health Services Research programme) as project number 14/19/19. For more information visit <https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/141919/#/>

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors have tried to ensure the accuracy of the authors' work and would like to thank the reviewers for their constructive comments however; they do not accept liability for damages or losses arising from material published in this scientific summary.

This 'first look' scientific summary presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Scientific summary

Background

Many of the problems confronting policy and decision makers, evaluators and researchers today are complex. For example, much health need results from the effects of smoking, suboptimal diets (including obesity), alcohol excess, inactivity or adverse family circumstances (e.g. partner violence) – all of which in turn have multiple causes operating at both individual and societal level. Interventions or programmes designed to tackle such problems are themselves complex, having multiple, interconnected components delivered individually or targeted at communities or populations. Their success depends both on individuals' responses and on the wider context in which people strive (or not) to live meaningful and healthy lives. What works in one family, or one organisation, or one city may not work in another.

Designing and evaluating complex interventions is challenging. Randomised trials that compare 'intervention on' with 'intervention off', and their secondary research equivalent, meta-analyses of such trials, may produce statistically accurate statements (e.g. that the intervention works 'on average') but which leave us none the wiser about where to target resources or how to maximise impact.

Realist evaluation seeks to address these problems. It is a form of theory-driven evaluation, based on realist philosophy, it aims to advance understanding of why these complex interventions work, how, for whom, in what context and to what extent – and also to explain the many situations in which a programme fails to achieve the anticipated benefit.

Realist evaluation assumes both that social systems and structures are 'real' (because they have real effects) and also that human actors respond differently to interventions in different circumstances. To understand how an intervention might generate different outcomes in different circumstances, realism introduces the concept of *mechanisms* – which may be helpfully conceptualised as underlying changes in the reasoning and behaviour of participants that are triggered in particular contexts.

This project aims to: develop quality and reporting standards; resources and training materials; build research capacity and; develop materials for lay participants involved in realist evaluations.

Objectives

1. Recruit an interdisciplinary Delphi panel of (e.g.) researchers, support staff, policymakers, patient advocates and practitioners with (various types of) experience relevant to realist evaluation.
2. Summarise the current literature and expert opinion on best practice in realist evaluation, to serve as a baseline / briefing document for the panel.
3. Run three (and more if needed) rounds of the online Delphi panel to generate and refine items for a set of quality standards and reporting guidance.
4. In parallel with the Delphi panel:
 - a. Provide ongoing advice and consultancy to up to ten realist evaluations, including any funded by NIHR, thereby capturing the 'real world' problems and challenges of this methodology.
 - b. Host the RAMESES JISCMail list on realist research, capturing relevant discussions about theoretical, methodological and practical issues.
 - c. Feed problems and insights from 4a and 4b into the deliberations of the Delphi panel.
5. Write up the quality standards and guidance for reporting in an open-access journal.
6. Collate examples of learning/training needs for researchers, postgraduate students, and peer-reviewers in relation to realist evaluation.
7. Develop, deliver and refine resources and training materials for realist evaluation. Deliver 3 x 2-day 'realist evaluation' workshops AND 3 x 2-day 'training the trainers' workshops for a range of audiences (including interested NIHR Research Design Service staff).
8. Develop, deliver and refine information and resources for patients and other lay participants in realist evaluation. In particular, to draft template information sheets and consent forms that could be adapted for ethics and governance activity.
9. Disseminate resources and training materials and other resources – e.g. via public access websites.

Methods

In this project we used a range of methods to meet the objectives set out above. To fulfil objectives 1 and 2 we undertook a thematic review of the literature that was supplemented by our content expertise and from collating feedback from presentations and workshops. We synthesised our findings into briefing materials for realist evaluations. We recruited members to the Delphi panel, which had wide representation from researchers, students, policy-makers, theorists and research sponsors. We used the briefing materials to brief the Delphi panel, so they could help us in fulfilling objective 3. For the advice and consultancy in objective 4, we drew not only on our experience in developing and delivering education materials, but also relevant feedback from the Delphi panel, the RAMESES JISCMail e-mail list on realist research approaches (www.jiscmail.ac.uk/RAMESES), training workshops and the evaluations teams we had supported methodologically in the past. To help us refine our publication standards (objective 5) we captured methodological and other challenges that arose within the realist evaluation projects we provided methodological support to. To produce the definitive reporting standards, quality standards and resources and training materials (objective 5), we synthesised expert input (from the Delphi panel), literature review and real-time problem analysis (e.g. feedback from the e-mail list, training sessions and workshops, and presentations).

Throughout this project we did not set specific time points when we would refine the drafts of our project outputs. Instead, we iteratively and contemporaneously fed any data we captured into our draft reporting standards, quality standards and resources and training materials, making changes gradually. Only our Delphi panel ran within a specific time frame. The definitive guidance and standards were, therefore, the product of continuous refinements. To understand and develop information and resources for patients and other lay participants in realist evaluation (objective 8) we convened a group consisting of patients and the public. We addressed objective 9 through academic publications, online resources and delivery of presentations and workshops.

Results

Our literature review identified 152 realist evaluations and when we had analysed 37 of these we had reached thematic saturation. Our analysis and discussion within the project team produced a summary of the published literature, and common questions and challenges in briefing materials for the Delphi panel. The Delphi panel comprised of 35 members from 27 organisations across six countries and five disciplines. Within three rounds the panels had reached a consensus on 20 key reporting standards, with an overall response rate of 76% and 80% for rounds 2 and 3 respectively. The RAMESES II reporting standards for realist evaluations have been published in an open access journal and the EQUATOR (Enhancing the QUALity and Transparency Of health Research) network (www.equator-network.org).

The quality standards and resources and training materials drew on the following sources of data: (1) personal expertise as researchers and trainers; (2) data from the Delphi panels; (3) feedback from participants at training sessions we ran; and (4) comments made on RAMESES JISCMail mailing list. We developed eight quality criteria for realist evaluations with different versions for evaluators, researchers, peer reviewers and funders/commissioners of research. For our resources and training materials we used the data we captured to identify the methodological topics that were identified by the majority of realist evaluators as most challenging. We developed training materials for 15 theoretical and methodological topics in realist evaluations. The quality standards and training materials are freely available online (www.ramesesproject.org).

We provided methodological support to 17 projects and presentations or workshops to help build research capacity in realist evaluations to 29 organisations – nationally and internationally. This training included two ‘training the trainers’ workshops run in conjunction with the National Institute of Health Research Research Design Service East Midlands. Finally we produced a generic patient information leaflet for lay participants in realist evaluations.

Conclusions

In conclusion, while realist evaluation hold much promise for developing theory and informing policy in some of the health and other sectors' most pressing questions, misunderstandings and misapplications of it is common. To try to address these problems we have produced reporting and quality standards, and resources and training materials. In addition we provided methodological support and advice to realist evaluation projects, ran training workshops for fellow realist evaluators and developed information and resources for patients and other lay participants in realist evaluation. However, for the quality of realist evaluations to improve, evaluators who wish to use realist evaluation will have to develop the necessary skills and use the materials we have developed.

We hope that our resources will be the start of an iterative journey of refinement and development of better resources for realist evaluations. Acknowledging that the science of evaluation should never be static, the RAMESES II project does not seek to produce the last word on these issues but to capture current expertise and establish an agreed state of the science on which future researchers will no doubt build. Much methodological development is needed in realist evaluation (for example work on appropriate quantitative methods, implications for research ethics, development of realist approaches in particular sectors and adaptation of existing evaluation tools for realist approaches). But this can only take place if there is a sufficient pool of highly skilled realist evaluators. Capacity building through (for example) training and 'apprenticeships' of less experienced evaluators with more experienced ones is the next key step in realist evaluation,

Funding: This project was funded by the National Institute for Health Research Health Services and Delivery Research Programme (project number 14/19/19).