Quality Standards for Realist Evaluation

For funders or commissioners of realist evaluations

The RAMESES II Project

Funded by NIHR

1

1. The evaluation purpose

Realist evaluation is a theory-driven approach, rooted in a realist philosophy of science, which emphasises an understanding of causation and how causal mechanisms are shaped and constrained by context. This makes it particularly suitable for evaluations of certain topics and questions – for example, complex interventions and programmes that involve human decisions and actions. A realist evaluation question contains some or all of the elements of 'What works, how, why, for whom, to what extent and in what circumstances, in what respect and over what duration?' and applies a realist logic to address the question(s). Above all realist evaluation seeks to answer 'how' and 'why?' questions. Realist evaluation always seeks to explain. It assumes that programme effectiveness will always be conditional and is oriented to improve understanding of the key contexts and mechanisms contributing to how and why programmes work.

Criterion	Inadequate	Adequate	Good	Excellent
A realist approach is suitable for the purposes of the evaluation.	There is no statement of the purpose of the evaluation *and/or* The evaluation does not seek to explain how and why the evaluand* works.	There is a clear statement of purpose for the evaluation. *and/or* The evaluation seeks to explain how and why the evaluand works.	Adequate plus: The evaluation seeks to explain how and why the evaluand works differently in different contexts and for different sub-groups.	Good plus: Stated purpose clearly explains how the findings are intended to be used. There is a coherent argument as to why a realist approach is appropriate.
The evaluation question(s) are framed in such a way as to be suitable for a realist evaluation.	The evaluation question(s) are not structured to reflect the elements of realist explanation. For example, answering the questions: • requires only description; and/or • requires only a numerical aggregation of outcomes; and/or • requires only summary of processes; and/or • relies exclusively on methods that are inadequate to generate realist understanding (e.g. 'a thematic analysis of …')	The evaluation question(s) include a focus on how and why outcomes are likely to be generated, and contain at least some of the additional elements, "for whom, in what contexts, in what respects, to what extent and over what durations".	Adequate plus: The rationale for excluding any elements of 'the realist question' from the evaluation question(s) are explicit. The question(s) are sufficiently focused to be managed within a realist evaluation.	Good plus: The evaluation question(s) are clear and as simple as possible. They can be understood by people without specialist methodological or content expertise.

*evaluand: is defined as 'that which is being evaluated', for example an intervention, programme, policy, product or initiative, or in some cases, sets of programs, policies or initiatives

2. Understanding and applying a realist principle of generative causation in realist evaluations

Realist evaluations are underpinned by a realist principle of generative causation. That is, underlying causal processes (called 'mechanisms') operate (or not) in certain contexts to generate outcomes. The explanatory framework is Context + Mechanism = Outcome (CMO). Realist evaluations aim to understand how different mechanisms generate different outcomes in different contexts. This intent influences everything from the type of evaluation question(s) to an evaluation's design (e.g. the construction of a realist programme theory, recruitment process and sampling strategy, data collection methods, data analysis, to recommendations).

Criterion	Inadequate	Adequate	Good	Excellent
A realist principle of generative causation is applied.	Significant misunderstandings of realist generative causation are evident. Common misunderstandings include: • programme activities or strategies are mislabelled as mechanisms • contexts are assumed to cause outcomes directly, rather than affecting whether and how mechanisms operate • outcomes are assumed to be caused directly by the programme/ intervention (rather than by underlying mechanisms) • no attempts are made to understand underlying mechanisms • relationships between an outcome, its causal mechanism(s) and context(s) are not explained • if theory is provided, this is not explicitly linked to context-mechanism-outcome configurations	Some misunderstandings of realist generative causation exist, but the overall approach is consistent enough that a recognisably realist analysis results from the process.	Assumptions and methods used throughout the evaluation are consistent with a realist generative causation.	Good plus: The evaluation strategy demonstrates exemplary understanding of a principle of realist generative causation, and application of methods consistent with that understanding throughout (for example, in question(s), design and the evaluations outputs). Emerging challenges arising as the evaluation unfolds are dealt with in ways that are consistent with realist generative causation.

3. Constructing and refining a realist programme theory or theories

At an early stage in the evaluation, the main ideas that went into the making of an intervention, programme or policy (the programme theory or theories – which may or may not be realist in nature) are surfaced and made explicit. An initial tentative programme theory (or theories) is constructed which sets out how and why an intervention, programme or policy is thought to 'work' to generate the outcome(s) of interest. Where possible, this initial tentative theory (or theories) will be progressively refined over the course of the evaluation. Over the course of the evaluation, if needed, programme theory (or theories) are 're-cast' in realist terms (describing the contexts in which, populations for which, and main mechanisms by which, particular outcomes are, or are expected to be, achieved.) Ideally, the programme theory is articulated in realist terms prior to data collection, in order to guide the selection of data sources about context, mechanism and outcome. However, in some cases, this will not be possible and the product of the evaluation will be an initial realist programme theory.

Criterion	Inadequate	Adequate	Good	Excellent
An initial tentative programme theory (or theories) is, or will be, identified and developed. Programme theory is or will be 're-cast' and refined as realist programme theory.	Programme theory (or theories): are not or will not be developed; *or* • are described but it is not clear how they were or will be used in the evaluation; *or* • are offered but it is not clear how they were or will be refined as realist programme theory during the evaluation.	Initial tentative programme theory (or theories) are or will be identified and (as far as possible) described in realist terms (that is, in terms of the causal relationship between contexts, mechanisms and outcomes). These are or will be refined as the evaluation progresses. Where possible, aspects of theory to be 'tested' are: • specified and justified in the evaluation design. • appropriate to the purpose of the evaluation. • aspects that will not be tested are identified and explanation is provided as to why.	Adequate plus: Programme theory (or theories) are described in realist terms and used to inform all aspects of the evaluation (e.g. focus an evaluation, identify questions, determine what types of data need to be collected and from whom and where). Where relevant, the programme theory or theories take into account the physical/ material (e.g. environmental) and social aspects of systems necessary to answer evaluation questions.	Good plus: The relationships between the programme theory (or theories) and relevant formal theory (or theories) will be sought. Where relevant, contexts which are not included in the evaluation are expressly addressed. The final realist programme theory (or theories) comprise one or more context-mechanism-outcome configurations, describing how and why different mechanisms are triggered (or not) in different contexts to generate different outcomes.

4. Evaluation design

Descriptions and justifications of what is planned in the evaluation design, in what order, and why should be clearly articulated. Realist evaluations are ideally adaptive – that is, the evaluation question(s), scope and/or design may be adapted over the course of the evaluation to 'test' (confirm, refute or refine) aspects of the programme theory as it evolves. If changes are made to the evaluation design, these should be clearly described and justified. At the start of an evaluation, where possible, any changes that might be needed should be anticipated and contingencies planned.

Criterion	Inadequate	Adequate	Good	Excellent
The evaluation design is described and justified.	The evaluation design is not clearly described or is not coherent. There is a lack of clarity as to what is planned in the evaluation design, in what order and why. The evaluation design does not clearly relate to or test the programme theory. The analyses are inconsistent with the assumptions underpinning realist evaluation.	What is planned in the evaluation design, in what order and why is described and justified in detail. The evaluation design is informed by an initial programme theory or theories, and sets out 'tests' important or priority aspects of these. The design is coherent, with a logical flow from purpose through focus, questions, data collection and analysis methods.	Adequate plus: The design tests multiple aspects of programme theory. The design enables alternative explanations to be investigated.	Good plus: The design is efficient, adding value by (for example) maximising use of existing data or increasing portability of findings. The design identifies or will identify the extent to which the interventions contribute to overall outcomes, and/or identifies other aspects of the context (e.g. other policies or programmes) which are likely to contribute to outcomes.
Ethical clearance is or will be obtained if required.	No consideration is given to whether the evaluation requires ethical approval.	Protocols for ethics approval are considered and approval sought if required.	Proposals for ethical approval clearly distinguish the implications of the evaluation for different groups and different contexts.	Where relevant, specific implications of realist methodology are explained in the proposal for ethical approval and specific strategies to address those implications are provided.

5. Data collection methods

In a realist evaluation, a broad range of data increases the robustness of the theory 'testing' process and a range of methods used to collect them. Data will be required for all of context, mechanism and outcome and to inform the relationships between them. Data collection methods should be adequate to capture not only intended but also (as far as possible) unintended outcomes (both positive and negative), and the context-mechanism interactions that generated them. Realist evaluation is usually multi-method (i.e. uses more than one method to gather data). Where possible, data about outcomes should be triangulated (at least using different sources, if not different types, of information).

Criterion	Inadequate	Adequate	Good	Excellent
Data collection methods are suitable for capturing the data needed in a realist evaluation.	Within the realist evaluation project: • it is unclear which data collection methods are used *and/or* • data collection methods are not informed by the need to find data to confirm, refute or refine the programme theory *and/or* • methods used are unlikely to capture necessary data to test the programme theory	Methods for collecting and documenting data are driven by the programme theory (or theories) and: • will capture the necessary data; *and* • will capture intended and unintended outcomes. *and will consider:* • the sampling needed to 'test' programme theory; *and* • the evaluation questions. The rationale for the methods and their implications for data analysis are explained.	Adequate plus: Data collection methods are explicitly consistent with realist methodology (e.g. realist interviewing). Quality control processes ensure that data collection methods are applied rigorously and consistently. Allowance is made to collect additional data for further refinement of programme theory (or theories) and/or CMO configurations as the evaluation unfolds. Data management processes (e.g. data bases, use of participant identifiers) are or will be constructed to enable intended analyses (e.g. sub-group analyses, tracking participants over time).	New data collection methods, tools and processes are adapted and/or developed where required and are consistent with realist principles. Any specific techniques used or adaptations made to instruments or sampling processes are justified.

6. Sample recruitment strategy

In a realist evaluation data are required for all of C, M and O. One key source is respondents or key informants. Data are used to develop and refine theory about how, for whom, and in what circumstances programmes generate their outcomes. This implies that any processes used to invite or recruit individuals need to identify an adequate sample of individuals who are able to provide information about contexts, mechanisms, outcomes and/or programme theory.

Criterion	Inadequate	Adequate	Good	Excellent
The respondents or key informants recruited are likely to be able to provide sufficient data needed for a realist evaluation.	Recruitment is or was ad hoc, opportunistic and/or not informed by the programme theory. Random samples are or will be used to generalise to whole populations (as distinct from sampling within theory-specified sub-groups). Convenience samples not related to programme theory are or will be used to test programme theories.	Recruitment is: • designed to find an appropriate sample of respondents who can provide information about contexts, mechanisms and/or outcomes for the programme theory. • purposive, with samples selected to test specific aspects of programme theory.	Adequate plus: Where needed, further recruitment is or will be undertaken to collect the data needed for further refinement of programme theory.	Sampling follows a rigorous and sequenced process of theory testing. A sufficiently large and diverse sample of relevant respondents is or will be recruited to provide evidence across contexts. When needed respondents will be approached again as new evidence emerges, to explore context and mechanism more extensively. Where applicable, sampling will involve sensitive strategies to successfully recruit respondents from disenfranchised communities or other ‘hard to reach’ groups.

7. Data analysis

Data analysis in realist evaluation is not a specific method but a way of interrogating programme theory (or theories) with data and a way of using theory to understand patterns in data. In other words, data analysis is a way of teasing out what works, for whom, in what contexts, in what respects, over what duration and so on.

In a realist evaluation, where possible, the analysis process should occur iteratively. The overall approach to data analysis is retroductive* (i.e. it moves between inductive and deductive processes, includes and tests researcher 'hunches' and aims to provide the best possible explanation of acknowledged-to-be-incomplete data). The processes used to analyse the data and integrate them into one or more realist programme theories should be consistent with a central principle of realism - namely generative causation. How these data are then used to further develop, confirm, refute or refine one or more programme theories should be clearly described and justified.

Criterion	Inadequate	Adequate	Good	Excellent
The overall approach to analysis is or will be retroductive.*	The approach to analysis is not retroductive. *or:* The overall approach to analysis is not clear.	The approach to analysis moves or will move between theory and data, data and theory, appropriate to the stage of theory development.	Adequate plus: Any theory (or theories) are developed and refined through the use of retroductive reasoning. Evaluators' 'hunches' are clearly described. Theories that remain untested at the end of the evaluation are identified.	Good plus: The analysis clearly links data, programme theory and formal theory.
Data analyses processes are consistent with a realist principle of generative causation.	Analytic processes are not described. Analysis is not or will not be disaggregated by sub-groups (i.e. 'for whom') or contexts. Subgroup analyses are planned without reference to programme theory (for example, disaggregating by demographic sub-groups rather than theory-relevant groupings).	Qualitative analysis identifies and explains the relationships between contexts, mechanisms and outcomes. Quantitative analysis 'tests' differences between sub-groups or contexts, in relation to programme theory. Findings from analysis are aligned against programme theory.	Adequate plus: Specific analyses are or will be conducted to 'test' the relationships within and between CMOCs. That is, evidence is not just aligned against programme theory: the linkages within the programme theory are 'tested'.	Good plus: When iterations in evaluation design and/or programme theory require additional analytic methods to be employed, those used are consistent with realist principles.

7. Data analysis continued

Criterion	Inadequate	Adequate	Good	Excellent
A realist logic of analysis is used to develop and refine theory.	The analyses used or planned do not: • identify contexts, mechanisms or outcomes; • identify the relationships between contexts, mechanisms and outcomes *and/or* • explain how the programme theory (or theories) are or will be further developed, confirmed, refuted and refined	Data is or will be analysed to develop and refine initial programme theory (or theories) into realist programme theory (or theories). The realist analysis has or will: • assign conceptual labels of C, M or O to each data element or finding within a Context-Mechanism-Outcome configuration (CMOC) – (e.g. 'in this aspect of the analysis, this item of data is functioning as context within this CMOC'). • identify the relationship of contexts, mechanisms and outcomes within particular CMOCs. • identify relationships across CMOCs – i.e. the location and interactions between CMOCs within a programme theory (or theories)	Adequate plus: Analysis: • integrates a range of data sources (e.g. qualitative and quantitative, primary and secondary data) *and* • describes how the multiple data types were or will be integrated to support inferences.	Data analysis is iterative over the course of the evaluation, with earlier stages of analysis being used to refine programme theory and/or refine evaluation design for subsequent stages.

* For more details on retroduction see:'Retroduction in realist evaluation' which may be found in the Standards and Training materials section of The RAMESES Projects website (www.ramesesproject.org).

8. Reporting

Realist evaluations may be reported in multiple formats – detailed reports, summary reports, articles, websites and so on. Reports should be consistent with the RAMESES II reporting standards for realist evaluations (see https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0643-1).

Criterion	Inadequate	Adequate	Good	Excellent
The realist evaluation is or will be reported using the items listed in the RAMESES II reporting standards for realist evaluations.	No information is provided on whether the RAMESES II reporting standard for realist evaluations will be used.	The RAMESES II reporting standard for realist evaluations is or will be used.	A firm commitment is made to: • use the RAMESES II reporting standard for realist evaluations. • provide justifications where items will not been reported.	Good plus: The report is well written, transparent, and easy to understand. Various reporting formats are used to present relevant findings to different audiences.

The RAMESES II Project
Funder: NIHR HS&DR

Project team:

Professor Trish Greenhalgh UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson

Dr Geoff Wong UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano

Dr Justin Jagosh UNIVERSITY OF LIVERPOOL
www.liverpool.ac.uk/psychology-health-and-society/staff/justin-jagosh

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

Professor Trish Greenhalgh's salary is part-funded by the Oxford Biomedical Research Centre, NIHR grant number BRC-1215-20008.

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

For information on how these quality standards were developed and how to use them please see the full project report which may be found on The RAMESES Project website (www.ramesesproject.org).

UNIVERSITY OF OXFORD
NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES Medical Sciences Division
CHARLES DARWIN UNIVERSITY
UNIVERSITY OF LEEDS
UNIVERSITY OF LIVERPOOL
NHS National Institute for Health Research