Introduction

The Medicines for Human Use (Clinical Trials) Regulations 2004 defines a protocol as “A document that describes the objectives, design, methodology, statistical considerations, and organisation of a clinical trial”. In the last 20 years, the production and publication of protocols for randomised controlled trials (RCTs) and systematic reviews (SRs) has increased exponentially. More recently, the enthusiasm for protocols has been extended to other study designs, including realist evaluation and realist synthesis. Thus, in the last 5 years, there have been a number of protocols published for realist evaluations (REs) (e.g. Dalkin et al, 2012; Randell et al, 2014; Mukumbang et al 2016; Mirzoev et al 2016). Unlike the RCT literature, relatively little has been written concerning the role of protocols in realist evaluations and to date there has been no research to compare protocols of realist evaluations with subsequent papers reporting their findings. Here we consider the extent to which the assumptions underlying the production of protocols for RCTs can be applied to protocols for realist evaluation and suggest the amendments that are necessary for realist evaluations.

Why write a protocol?

There are several ideas and assumptions underlying the publication of protocols. For RCTs, these include reducing the potential for observer and researcher bias by ensuring that methods are appropriately designed and holding authors to account for following the protocol to reduce outcome reporting bias (McNamee, 1997; Jones and Abbasi 2004). Protocols are also argued to enable the replication of the study, which in some quarters is considered to be the hallmark of ‘good science’ (Loscalzo, 2012). Protocols are also hypothesised to act as a planning and risk assessment tool, enabling teams to identify potential risks to the study and put in place plans to address these prior to the beginning of the study (Shamseer et al, 2015). The publication of a protocol is expected to stimulate debate amongst the academic community regarding the validity and value of specific studies or of research methodologies more generally, and thus advance knowledge. Finally, they are also hypothesised to serve as a means of reducing duplication of effort by alerting other research groups to ongoing research (Shamseer et al, 2015).

Protocols and validity: Closed systems vs open systems

The validity and credibility of RCTs rests on making decisions about key aspects of trial design to avoid influence from stakeholders on these decisions during the trial. Such influences are assumed to threaten the creation of a closed system, in which the effects of an intervention can (at least in theory) be isolated from its context in order to understand the contribution that the intervention alone makes to outcomes. However, realists argue that it is not possible to create closed social systems based on experimental manipulation. Rather than treat contextual influences as contamination or bias, realist evaluators seek to understand how context shapes the mechanisms through which an intervention produces outcomes. As such, validity in realist evaluation does not rest on the achievement of a closed system. Instead, the validity of a realist evaluation rests on the application of realist logic to identify, test and refine theories which explain how the context in which interventions are implemented shape the mechanisms through which they work and thus their outcomes (although this idea of validity is still contested (Greenhalgh et al 2015)).
Planning for Emergence: Flexibility and not Deviation

It is not always desirable or possible to make decisions about every aspect of study design in advance for a realist evaluation. Some aspects are expected to emerge as the theory develops, making their specification within a protocol challenging. Even where programme theories can be identified at the outset (which is not always the case), these theories are always provisional and subject to change. Like the programmes realist evaluation seeks to study, the methodology is emergent and iterative; new programme theories and ‘sub theories’ may surface during the evaluation and may require changes to data collection or analysis in order to test them (Manzano 2016). Therefore, protocols in realist evaluation cannot be written with the same degree of detail and precision as those for RCTs. Instead, flexibility needs to be written into realist evaluation protocols. For example, the number of interviews and specific participant groups may not be specified rigidly but described as a range (e.g. 20–35) with the option to reduce or extend if required (“these numbers may decrease or increase dependent of the process of theory testing”).

Deviation from protocol as a quality control process?

For RCTs, there is a clear consensus that deviation from the protocol on particular aspects of trial design and analysis, such as inclusion criteria and outcome reporting, constitutes bias (Schulz et al, 2010; McNamee 1997; Goldacre et al 2016). The flexibility built into a realist protocol poses challenges to this idea. Given that modes of data collection and analysis in realist evaluation are iterative and emergent, deviation from the protocol cannot be assumed to constitute bias. Instead, if deviations from the protocol are acknowledged and explained, then this enhances the transparency of reporting and readers can judge for themselves the extent to which this may compromise the study’s validity.

“…there are many different ways in which to conduct a robust and valid realist evaluation.”

However, lurking underneath this logic is a more fundamental problem with the idea that comparisons between protocols and reported findings enable the detection of bias in realist evaluations. Realist evaluation is not ‘a method’ but an approach, variation in the practice of realist evaluation is evident (Pawson and Manzano 2012; Marchal et al 2012) and there are many different ways in which to conduct a robust and valid realist evaluation. This makes it difficult to pinpoint the precise ways in which deviation from a realist evaluation protocol may constitute a threat to validity. The reporting and quality standards for realist evaluation produced by the RAMESES II study (www.ramesesproject.org) include a rubric against which the quality of realist evaluation study can be assessed.

The underlying assumptions of ‘what constitutes validity’ are to some extent different in realist as compared to other studies. In most approaches, validity refers to the ability of a study to accurately measure or assess what it sets out to assess. This remains a concern in realist evaluation. However, realist approaches acknowledge that some things (particularly, some mechanisms) may not be either observable or measurable; but nonetheless must be incorporated for a study to be considered a valid realist investigation.

Protocols for realist evaluations

Given the need for flexibility in conducting realist evaluations and accepting that there are many different ways in which to carry out a robust and rigorous realist evaluation, is there value in writing a protocol for a realist evaluation?

We argue that there can be, because:

• Writing a protocol requires rigour in planning, and realist evaluations (like all other methods) require good planning. This does not imply that the plans are rigid and cannot be changed, but that the plans provide the basis from which to adapt as required;

• Describing the methods to be used can identify ‘non-realist’ assumptions (for example, describing sampling methods that are not theory-based, or describing statistical methods that do not allow for disaggregation of outcomes by sub-groups or contexts);

• Realist evaluations do not expect to be replicable in the traditional sense of the word, but they do expect to be transparent. A protocol enables others to assess and critique the quality of the work, with the assumption that critique contributes to quality;

• A well-written protocol can provide the basis for other aspects of an evaluation project, including developing research ethics applications (where required) and writing the methods chapter of a report.

However, just as standards for realist evaluations are different from standards for other evaluation or research designs, the standards for realist protocols must also be different from those for other designs. The protocol, like the evaluation itself, should be consistent with the standards for realist evaluation (www.ramesesproject.org).
References


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.liv.ac.uk/cares

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.