Realist RCTs of complex interventions – an oxymoron

Bruno Marchal  
Department of Public Health, Institute of Tropical Medicine, Antwerp (bmarchal@itg.be)

Gill Westhorp  
Community Matters Pty Ltd, Australia (gill.westhorp@communitymatters.com.au)

Geoff Wong  
Centre for Primary Care and Population Health, Queen Mary, University of London, London (grckwong@gmail.com)

Sara Van Belle  
Politics and Policy Group, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine (Sara.VanBelle@lshtm.ac.uk)

Trisha Greenhalgh  
Global Health, Policy and Innovation Unit, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry (p.greenhalgh@qmul.ac.uk)

Guy Kegels  
Department of Public Health, Institute of Tropical Medicine, Antwerp (gkegels@itg.be)

Ray Pawson  
School of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT, UK (r.d.pawson@leeds.ac.uk)

Corresponding author
Bruno Marchal  
Department of Public Health, Institute of Tropical Medicine, Antwerp  
Nationalestraat 155, B-2000, Antwerp, Belgium  
(bmarchal@itg.be)  
Tel.: +32.3.247.63.84
Realist RCTs of complex interventions – an oxymoron

Abstract

Bonell et al. discuss the challenges of carrying out randomised controlled trials (RCTs) to evaluate complex interventions in public health, and consider the role of realist evaluation in enhancing this design (Bonell et al., 2012). They argue for a “synergistic, rather than oppositional relationship between realist and randomised evaluation” and that “it is possible to benefit from the insights provided by realist evaluation without relinquishing the RCT as the best means of examining intervention causality.” We present counter-arguments to their analysis of realist evaluation and their recommendations for realist RCTs.

Bonell et al. are right to question whether and how (quasi-)experimental designs can be improved to better evaluate complex public health interventions. However, the paper does not explain how a research design that is fundamentally built upon a positivist ontological and epistemological position can be meaningfully adapted to allow it to be used from within a realist paradigm. The recommendations for “realist RCTs” do not sufficiently take into account important elements of complexity that pose major challenges for the RCT design. They also ignore key tenets of the realist evaluation approach.

We propose that the adjective ‘realist’ should continue to be used only for studies based on a realist philosophy and whose analytic approach follows the established...
principles of realist analysis. It seems more correct to call the approach proposed by Bonell and colleagues ‘theory informed RCT’, which indeed can help in enhancing RCTs.

**Key words:** Research design; complexity; randomised controlled trials; realist evaluation; methodology

**Introduction**

Bonell et al. discuss the challenges of carrying out randomised controlled trials (RCTs) to evaluate complex interventions in public health, and consider how realist evaluation could help (Bonell et al., 2012). The authors agree with the realist call for examining not only what works, but also how, for whom and in what conditions. They agree that careful attention should be paid to the influence of context and to programme theory development. They then go on to claim that RCTs are the best design to assess mechanisms and the influence of context, “because randomized control groups actually take proper account of rather than bracket out the complexity of social causation” (page 1). They argue for a “synergistic, rather than oppositional relationship between realist and randomised evaluation”, by which they mean that the RCT can – and indeed should – be used as part of a realist evaluation of a complex intervention. In this paper, we present counter-arguments to their analysis of realist evaluation and their recommendations for realist RCTs.

The authors are right to question whether and how (quasi-)experimental designs may
be improved to better evaluate complex public health interventions. All these
questions are important current concerns in public health (Forss et al., 2011;
McDaniel et al., 2009; Webster et al., 2010; Zimmerman et al., 2012). However, we
believe the paper lacks clarity on the authors’ central premise, which holds that a
research design that is fundamentally built upon a positivist ontological and
epistemological position can be meaningfully adapted to allow it to be used from
within a realist paradigm. For instance, Bonell et al. do not explain how the use of
control groups in trials may allow researchers to take context into account instead of
merely controlling for it. The statement “RCT’s use of control groups actually reflects
the opposite: how interventions interact with contextual factors in order to produce
an outcome” (page 15) surely begs the question of how RCTs identify how
interventions interact with contextual factors. Their proposal for “realist RCTs” seems
to be based on what we consider to be flawed interpretations of key elements of both
complexity theory and realist evaluation that present major challenges for the RCT
design.

The problem with RCTs of complex interventions

RCTs and experimental methods are widely considered to be the gold standard
strategy for causal investigation in medicine. They are increasingly proposed in
evaluations of public health interventions. It is generally agreed that the RCT and
other quasi-experimental designs are built upon objectivist (or ‘positivist’)
assumptions, which hold that causality cannot be observed and that the best we can do
is to demonstrate regularity between a particular intervention and a particular
outcome. This refers to the Humean concept of constant conjunction. Because its adherents assume that causation cannot be observed, all that an RCT attempts to do is address attribution: can we attribute the observed outcome to the intervention? Methodologically, this is translated in various quasi-experimental designs, which focus on the use of counterfactuals in order to demonstrate attribution (‘What would have happened in cases without intervention?’) and a preference for quantitative data collection and statistical techniques like linear regression and cluster analysis (Fiss, 2010).

Bonell et al. (page 14) indicate that they do not necessarily agree that RCTs are based on a positivist ontological and epistemological foundation, but they opt not to discuss this further in this paper. This is a pity, because besides the practical challenges of RCTs that Bonell and colleagues mention (stakeholder resistance, contamination effects, blinding and information bias), it is the ontological position and its epistemological consequences that limit the usefulness of RCTs when applied to complex interventions. The resulting methodological choices regarding causality make it difficult to infuse realist evaluation principles in RCT designs. As Sanderson (2000) states, “Approaches founded upon the assumptions of stability and equilibrium, of linearity in the relationship between variables, and of proportionality of change in response to causal influences are not appropriate in seeking to understand social systems that exhibit complexity”. RCT designs make just these assumptions.

Currently, some RCTs go beyond a ‘pure’ experiment and include various measures to take account of contextual variables that cannot be controlled for by the researcher.
The influence of context will be levelled out by, for example, including study sites whose contexts are broadly comparable. Randomised designs will also typically define a stratified sample and include statistical correction for baseline differences in gender, ethnicity and age between groups. Indeed, much of the progress in RCT methodology in recent years has been in the refinement of such techniques. However, applying such techniques to rigorously maintain internal validity leads to a situation where it is not possible to determine in which conditions and through which configuration of factors the outcome of interest is reached.

In their discussion of RCTs, Bonell and colleagues downplay the consequences of complexity for research and evaluation. They seem to consider complex interventions as merely consisting of multiple components and are mainly concerned with ascertaining the contribution of the intervention components to the observed outcome. This view ignores other elements of complexity, such as [a] the principle of non-linearity (i.e. that a small change in input may, under certain conditions but not others, produce a large change in outcome); [b] the key contribution of local adaptiveness and feedback loops; [c] the phenomenon of emergence; [d] the importance of path dependence (i.e. that a particular context has arisen for complex historical reasons and interventions need to take account of the path of history rather than view current reality as a ‘freeze frame’); and [e] the role of human agency (that complex interventions are introduced, delivered and at times resisted by people who have identities, values, skills, beliefs, goals and so on).

Experimental designs, especially RCTs, consider human desires, motives and behaviour as things that need to be controlled for (Fulop et al., 2001, Pawson, 2006).
Furthermore, its analytical techniques, like linear regression, typically attempt to isolate the effect of each variable on the outcome. To do this, linear regression holds all other variables constant “instead of showing how the variables combine to create outcomes” (Fiss, 2007, p. 1182). Such designs “purport to control an infinite number of rival hypotheses without specifying what any of them are” by rendering them implausible through statistics (Campbell, 2009), and do not provide a means to examine causal mechanisms (Mingers, 2000). Non-realist scholars present similar arguments. They argue that RCTs and similar designs are inadequate when considering social change over time and that they cannot elucidate theories of change in relation to such change (Barnes et al., 2003; Berwick, 2008; Victora et al., 2004). This is in sharp contrast with realist evaluation. Resonating with complexity theory, realist approaches view human agency and social interactions as the very core of the change. “It is through the workings of entire systems of social relationships that any changes in behaviours, events and social conditions are effected. A key requirement of realist evaluation is thus to take heed of the different layers of social reality which make up and surround programmes ” (Pawson & Tilley, 2004). The realist evaluation of the long-term effectiveness of a cardiac rehabilitation programme by Clark et al. (2005) is an example of how this interplay between intervention, individuals and context can be examined. The authors aimed at identifying the mechanisms and context elements that keep cardiac patients attending the preventive activities of the programme in order to clarify not only whether but also how such programmes work. On the basis of focus group discussions with patients, they identified psychological mechanisms that are triggered by the programme, such as patients’ regaining trust in
the capacity of their body, and social mechanisms, including camaraderie and building up of social capital. They investigated in which context this programme worked and found that it operates in community-based settings where the patients feel safe (e.g. because competent physical trainers are permanently available). “The positive effects of CR [cardiac rehabilitation] on health arose predominantly from increased social confidence and ability to interpret and judge the body’s physical boundaries. The perceived safety of community exercise settings was the most formative contextual factor affecting health behaviours over the longer-term” (Clark et al. 2005). The results of this study indicate how CR programmes of this kind work and help programme designers to develop interventions that trigger such responses among the target group so as to bring about changes in behaviour.

Whereas a RCT would be able to compare the effectiveness of different variations of such a programme – e.g. differences in duration, frequency of contacts, target groups, etc. – the methodological requirements to do so exclude picking up the interplay between programme implementation, the individuals who are targeted, the programme’s context and the wider social context. Even if evaluations of implementation, process and context are added, they can elucidate just that – the intensity, fidelity and actual process of implementation, and the context in which the intervention took place. In practice, such additional studies do not focus on the mechanisms, the reasons why individuals continue to participate. Such information is lost in the aggregation process required to give RCTs their power. RCTs and derivate designs come fully into their own when used to compare the effects of simple, single interventions (Grimshaw et al., 2001; Eccles et al., 2003; Berwick, 2008), but they
themselves cannot prove the underlying causal mechanism.

Interpreting realist principles

We also believe that Bonell and colleagues misinterpret some important principles of realist evaluation and its critique of RCTs and positivist approaches in general.

First, they misinterpret Pawson and Tilley (1997) when they state on page 15 that realists would not be interested in whether interventions have an effect. Indeed, it is not possible to undertake a realist analysis (“context, mechanism, outcome”) without considering outcomes. They also depict the conventional realist approach in a somewhat extreme way, for instance stating (incorrectly) on page 25 that realists would flatly reject experimental methods. They appear to conflate the terms ‘control’ and ‘comparison’ (for example on page 17), thereby creating the impression that Pawson and Tilley (1997) do not believe that control of the context in a trial is helpful.

Second, Bonell et al. appear to underestimate the potential of realist evaluation, in particular when they say that “By neglecting the counterfactual, and therefore failing to test hypotheses about what might have happened in the absence of intervention, the model of ‘realist’ evaluation proposed by Pawson and Tilley is extremely limited.” (page 16).

Our own view is that the realist approach is essentially about hypothesis testing. However, both the nature of the hypothesis and the way of testing the hypothesis is different from experimental designs, because of the different philosophical
assumptions about the nature of reality (ontology) and about how we might know that reality (epistemology). Realist approaches, whether in evaluation or synthesis, begin by developing “a realist hypothesis” about the question at hand. That hypothesis proposes that particular mechanisms will operate in particular contexts to generate particular outcomes. Empirical work then tests, refines and further specifies those hypotheses, resulting in context-mechanism-outcome configurations – more detailed understandings of the contexts in which particular mechanisms generate particular outcomes.

We refer to the study by Byng et al. (2008) as an illustration of this process. The authors carried out a realist evaluation that aimed at understanding the results of a cluster RCT of a complex intervention for shared care for patients with long-term mental illness. The intervention consisted of (1) primary care-based systems for registering patients, recall and review; (2) education and audits and (3) development of a liaison relationship between primary care teams and specialists. Facilitators were trained for delivering the intervention, a toolkit was introduced and small financial incentives were created. While the training, the toolkits and the incentives were fixed elements, the actual work of the facilitators was allowed to be flexible – they were supposed to liaise the various actors according to the needs of the actors, the setting and the health needs. To assess the effectiveness of this multi-faceted intervention, a cluster RCT was carried out. It showed that relapse rates were reduced and practitioner satisfaction improved. However, the records for physical and mental health care processes documented no improvement, patient unmet need did not reduce and patient satisfaction did not improve. The realist evaluation aimed at making sense
of these findings. It started with the development of a theoretical model of the intervention (a realist hypothesis, also called ‘programme theory’ by realist evaluators). The researchers found that principles of ‘shared care’ (need for co-ordination between providers at different levels to limit duplication and respond to unmet needs) and ‘chronic disease management’ (service redesign including timely review, expert input, patient involvement and information systems) were underlying the design of the intervention (see Byng & Jones, 2004 for more details).

Acknowledging the central role of the implementers, the researchers zoomed in on the service managers and the providers in both arms of the intervention. Twelve case studies were developed, focusing each on a single primary care setting. Interviews were carried out with these cadres in order to identify the essential elements of the intervention and the interaction between programme and context.

The difference with a classic process / context evaluation lies in the way the data were analysed. Following the realist approach, the analysis was informed by the initial programme theory. During the initial case analysis, the researchers assessed the actual implementation of the intervention, the organisational context of the primary care practice and its general environment, significant external events and the perceptions of the respondents of the conditions of success. This led to case descriptions that clarified how the planned intervention actually played out in each setting.

Subsequent cross case analysis led to finding regularities. For instance, most practitioners considered the toolkit as not important and the financial incentive as a mere token. In cases with positive results, the integration of link workers into the primary care teams was found to have initiated a positive feedback cycle: by enabling
face-to-face contact between providers of different services and initiating critical review of ineffective practices eventually contributed, a climate of trust and further liaison opportunities were created. In the cases with negative results, link workers were found not to participate in reviewing patients records. The context was not favourable: the mental health trust management teams were not supporting the work of the link workers.

During the last stage, the results of the realist evaluation were compared with the RCT results. Similar results were found regarding GP satisfaction and different results concerning intervention facilitation, practice systems, liaison, mental health outcome and relapse (see table in Byng et al. 2008, page 7). The realist study not only allowed to identify the patterns through which the fixed components of the intervention (the toolkit, the training of facilitator, the incentives) combined with the flexible part (the facilitation) to produce the observed outcomes (including changes in process of care, patient satisfaction and relapse). It also identified the critical elements of the facilitators work (liaison, case discussion and review) and the necessary context conditions (for instance support by senior management and stable personnel).

The above discussion shows that Bonell et al.’s example of how realists would do an assessment of the IMAGE project is an overly simplistic and fundamentally incorrect representation of realist evaluation’s approach.

Third, Bonell et al. claim that realist evaluation “can only develop a sense of the plausibility of the effects of an intervention, not their probability” (page 16). We agree: realists argue that prediction of specific outcomes of complex interventions is
out of bounds, and necessarily so because of the only semi-predictable nature of complex interventions. Therefore, the best we can offer is plausible explanations.

Rather than presenting these differences in polarised terms (‘experimental studies are probabilistic; realist explanations are plausible’), we suggest a more nuanced version: experimental studies generate probabilities but engage only minimally with external validity (the extent to which those probabilities hold across contexts). Realist studies are correspondingly restrained in making probabilistic statements because realists assume that truly complex issues cannot be fully known to the extent that probabilistic laws can be generated.

Flawed recommendations

The flawed interpretations of both complexity theory and the work of Pawson and Tilley (1997) outlined above underpin the recommendations that Bonell et al. make for ‘realist RCTs’. The fourth and fifth recommendations are somehow compatible with the realist logic, but the first three are not.

We will start with the fourth proposition that advocates the mixed use of quantitative and qualitative methods. This is in line with realist evaluation, which is method neutral: the study design and the methods should allow collection of data that permit testing the hypothesis. We also generally agree with the fifth recommendation: to build and validate theories is entirely consistent with a realist approach – which indeed builds and validates theories. However, the authors combine theory-driven evaluation concepts (see their reference to Weiss) with concepts of theories of change (see reference to Connell & Kubisch) and realist evaluation. In other places, we noted
a similar conflation of terms, like ‘programme theory’ or ‘mechanism of change’.
Since such terms have different meanings in the different schools of theory-driven
inquiry, precise definitions of these terms are required.

We have more problems with the first three recommendations. The first is to have
designs with two interventions and four groups in order to understand “the effects of
intervention components separately as well as in combination.” Such an approach
only works if the contexts for all four groups are identical in all important respects.
This implies that the researchers know in advance what it is about the context that
really matters, which from a complexity theory point of view is often impossible. This
assumption would thus limit the domain of application of ‘realist RCTs’ to
interventions for which such knowledge is already available and which could then be
built into the study design as part of the programme theory. This rules out its use for
complex interventions.

The second recommendation is that ‘realist RCTs’ would demonstrate mechanisms of
change, for which they propose what is essentially a logic model and mediation
analysis. These methods attempt to identify key elements of the intervention and
estimate the contribution of intervention elements on intermediate outputs and
observed outcomes. While that could work in complicated interventions, which
consist of several, non-interacting components, they are inadequate for complex
interventions, where human agency and non-linear interaction between intervention
components and with context is highly likely to influence the outcomes. Mediation
analysis does nothing to elucidate mechanisms as defined in a realist approach. For
realists, elements of the intervention are not mechanisms. Realists see programme
mechanisms as involving a change of reasoning on the part of actors in particular contexts – sometimes described as an interaction between the resources provided by the programme and the reasoning of participants – and this cannot be tested simply through mediation analysis.

From a realist perspective, we could not agree more with their third recommendation to pay more attention to context, both during the hypothesis development and the empirical phase. However, if the intervention is truly complex, the interaction between the intervention and its context may be hard to predict, and thus difficult to examine. Bonell and colleagues are not clear on how a ‘realist RCT’ would do that.

Conclusion

Whilst we believe Bonell et al. have made a poor case for integrating realist evaluation in RCTs, we would like to suggest a compromise with more potential to deliver insights in the realm of complex public health interventions. We suggest that the adjective ‘realist’ should continue to be used only for studies based on a realist philosophy of science and whose analytic approach follows established principles of realist analysis. We further propose that the term ‘realist RCT’ should be replaced by ‘theory informed RCT’, which could include using methods such as the logic model and mediation analysis proposed by Bonell et al., entirely consistent with an positivist philosophy of science. Such an approach would resonate with that of the International Initiative for Impact Evaluation (3ie) (White, 2009), which proposes theory based impact evaluations for complex interventions. Also the latest MRC guidelines for developing and evaluating complex interventions are, broadly speaking, proposing
this kind of approach (Craig et al., 2008). But the MRC guidelines, rightly, make no claim to being a realist approach.

The science of using RCTs to evaluate complex interventions is progressing apace, and more scholarly debate is needed on the place of different approaches. Whilst we are cautiously supportive of quasi-experimental designs, we believe there are inherent limitations as well as strengths to these designs. When interventions are truly complex, the extent to which impacts might be predicted with any consistency is limited. Instead, realist evaluation provides a methodology to develop plausible explanations for what has happened using accumulation and specification. Although this may not match the ambition of those who wish to seek out ‘hard’ predictive probabilities, it requires – paradoxically - the courage to discover hidden explanation (Astbury & Leeuw, 2010), a far more challenging undertaking.

References


Fiss, P. (2010). Case studies and the configurational analysis of organisational


