



LEEDS
BECKETT
UNIVERSITY

Citation:

Morgan, JA (2019) A Realist Alternative to Randomised Control Trials: A Bridge Not a Barrier? The European Journal of Development Research, 31 (2). pp. 180-188. ISSN 0957-8811 DOI: <https://doi.org/10.1057/s41287-019-00200-y>

Link to Leeds Beckett Repository record:

<http://eprints.leedsbeckett.ac.uk/5802/>

Document Version:

Article

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

Published in *European Journal of Development Research* 2019
DOI:10.1057/s41287-019-00200-y

A Realist Alternative to Randomised Control Trials: A bridge not a barrier?

Jamie Morgan*

Keywords: randomised control trials, realism, mixed methods, QCA

Wendy Olsen’s “Bridging to action” is a timely intervention that seeks to reorient development research (Olsen, 2019a). Olsen’s position takes as its point of departure the growing schism between action research and the use of randomised control trials (RCTs) and specifically argues for a realist alternative approach as one way to underpin and operationalize development research. In the following short essay, I set out the terms of the underpinning warrant for the argument and in the conclusion suggest additional ways to elaborate.

The problem to be addressed

RCTs focus on an observable *specific* feature of a situation, that can, through design, “in principle” be *isolated*, seek *equivalent* groups and apply an intervention to a treatment group, which can then be contrasted with a control group; the intervention is subject to a hypothesis test and an analytical statistical *measure* of the existence and *strength* of an effect. For advocates of RCTs the approach encourages a focus on “what works”, is pragmatic in so far as it is not overburdened by received theory or ideology, is transparent in procedure and claims, and these are fundamentally empirical (which others can replicate); the dominant inference is that RCTs provide a substantive means to justify generalization of the intervention from the specific case, which *ought to* lead to better outcomes in an instrumental sense. That is, one replicates the measurable effect for the population and does so based on policy that has been derived from a “scientific” method – so policy is rooted in “rigour” and this in combination with transparency means intervention is more “credible” and is likely to be “efficient” in the economic sense of that term when implemented by a donor agency, NGO, government etc (see, for example, Banerjee and Duflo, 2017).

However, RCTs are increasingly recognized as problematic. This is in several fundamental ways (see, for example, Ravallion, 2012; Deaton and Cartwright, 2018; Sindzingre, 2019). Perhaps most importantly, the application of RCTs is not just adoption of an approach, but rather the creation of a socio-political process with consequences for its object of study and field of intervention (“development”). RCTs have become dominant in development research and the effects are cumulative and self-reinforcing. The claim of scientific rigour and the production of measurable (costed) effects appeals to donors, since they must justify the use of funds to some constituency. A growing preference for RCTs means that interventions that are conducive to an initial RCT format tend to be ones that become development programmes. As a result, the assessment of development programmes increasingly requires

* Jamie Morgan: School of Economics, Analytics and International Business; Room-520 The Rose Bowl, Leeds Beckett University, Portland Place, Leeds, LS1 3HB, UK.

✉ j.a.morgan@leedsbeckett.ac.uk

an RCT skillset and so the field of recognized development expertise is gradually captured by RCT proponents, whilst the dominance of the RCT skillset simultaneously works to disempower or marginalize the skills and voices of other experts who might place a brake on the cumulative tendency towards RCT adoption; and so the development of development increasingly becomes one focused on and through RCTs. Recipient countries, meanwhile, are also potentially disempowered, since socio-economic policy can become weighted towards interventions based on RCTs funded by donors, to which the recipient may lack the political will, resources or expertise to object.

There is, therefore, a political economy to the sociology of knowledge effects of RCTs. Though RCT proponents are well-intentioned, they have captured the broader field of development in ways that reduce the real plurality, and inter and multi-disciplinary basis of applied development research. This has reduced the capacity of critique to operate effectively and for the actual practice of development to cope with its own cumulative power asymmetries. Martin Ravallion has been particularly vocal in claiming that development has become skewed towards individual interventions of some measurable kind and that these reduce the scope of development to some given alleviation rather than more broad-based transformation, which may in the end be self-limiting.

There is a further normative issue here in so far as a measurably better outcome raises the questions of better than what and under what conditions and for whom? RCTs both by methodological focus and by the consequence of their dominance create problems of equality of participation in the deliberation that affects the dynamics of answers to these questions. Moreover, it does so because of the precedence RCTs have acquired through claims to be rigorously scientific in orientation. However, this status as adequate science too is open to dispute. It can be argued that RCTs have captured the field despite fundamental problems of modelling and questions regarding efficacy – which may indicate that it is because of capture of the field rather than by virtue of adequacy that RCTs continue to dominate -- the approach has achieved a “lock-in” creating a path dependence in development research. The implication is that the dominance of RCTs expertise is self-confirming, since any failure is positioned as relative success according to scientific method: problems are expressed as limitations and are grounds for iteration rather than repudiation of the approach. However, if one looks beyond the dynamics of this self-confirming position then the adequacy of RCTs depends on three sets of conditions:

1. The *efficacy* of design depends on it being *possible* to identify similar groups and to isolate a significant feature for a treatment to be applied to one group and where the measurable effect can be assumed to be contrastive to a control group; this requires significant subsequent difference to be reducible to what is isolatable, which in turn requires consistency for similarity in both what is isolated from and in what is initially isolated (for the difference to subsequently apply as a group contrast), and this requires a threshold of homogeneity within both groups and between both groups for the isolation to be practical.¹ This is an order of *design* which cannot be achieved for *human social subjects*, it can only be asserted based on the attempt -- based on claims regarding some observation that licenses *assumption* regarding the constituting

¹ Unless difference is irrelevant, which is equally odd (see 3).

components (in the end similarity must mainly be found rather than designed, for the assumptions to be warranted). Multiple critiques demonstrate that what proponents claim to be viable and simple is in actuality a high (and thus rare or forbiddingly difficult) benchmark. However, even allowing that the procedure can *to some degree* be reasonably attempted in some circumstances (approximating its ideal form):

2. The *relevance* of design *for* the control and treatment groups in the RCT depends on it being not just *possible* but also *meaningful* to isolate a particular feature to introduce an intervention to produce a measurable effect. That is, for the isolation to be a relevant exercise with meaningful results, it must be assumed that the production of the effect is not significantly different in its operation and consequences when produced in isolation than it could or would otherwise be for the members of the group. Multiple critiques demonstrate that model misspecification is technically difficult to identify (because some statistical tests can be passed irrespective and because some problems are allowed to persist under such categorisations as “observational equivalence”) and yet misspecification is a continual possibility. The restriction of an RCT design distorts the complex constitutive nature of many social phenomena, such that the conditions that produce the feature are *not* meaningfully expressed via isolation. As such, though “internal validity” may be claimed it cannot genuinely be established – and, concomitantly, the possibility of model misspecification may actually imply the impossibility of specification, subject to the guiding methodology that informs the understanding of the RCT procedure. However, even allowing that, in *some* circumstances, a phenomenon can be identified for which a given and simply produced feature exists, thereby facilitating *and* warranting an isolating design procedure, the problem remains that:
 3. The *relevance* of RCTs as a procedure *beyond* the treatment and control groups depends on it being meaningful to export to *other* groups and to society at large, the intervention to produce a measurable effect. That is, for the procedure to be relevant, the measurable effect must be reproducible elsewhere according to the same set of conditions as the initial RCT or on the assumption that the conditions of the uncontrolled situation elsewhere would not constitute the effects differently or would not confound the effects or would not lead to further adverse consequences (and these are neither the same nor are they mutually exclusive possibilities). Multiple critiques demonstrate that exporting the intervention depends on transference of what is essentially an aggregated input-output manipulation that is immune to difference of constitution, such that the procedure is context independent (over and above the terms of the intervention, differences make insufficient difference to affect the effect). This is an extremely stringent criteria, and, as such, “external validity” cannot be presupposed; it, therefore, follows that RCTs cannot provide a substantive means to justify generalization of the intervention from the specific case, in so far as RCTs cannot provide *a priori*-type grounds for claims and this remains the case even if (and this is a major *if*) the RCT procedure can be seen to work in the particular case -- the extension is indeterminate as to whether it would work elsewhere, undermining the force of argument that RCTs focus on “what works” or provide any special status of “efficiency”: what works requires a more basic understanding of why *it* works and this is a universal problem, but not likely one that can or could express universal invariant conditions (the explanation of “why” for “what” matters precisely because difference matters).

In combination, 1-3 have created problems in development research for the construction of RCT models, the replication of research findings of RCT approaches and the reproduction of the measurable effect based on the generalization of the intervention. In one sense, model specification, internal validity, and external validity are technical issues regarding method, and, as such, critique is epistemic. However, the underlying critique is ontological, in so far as the problems of the approach are rooted in how social reality is treated. Possible differences of cause and consequence operative in terms of persons, the social relations of individuals, the group, groups, social structure, system, systems and society at large are reduced to the cause-effect mechanics of the RCT (group similarity, isolation, implementation, effect...). The underlying procedure requires the transformation of complexity into simplicity, suppressing significant difference, as though homogeneity rather than heterogeneity were pervasive.

RCTs treat social reality as though some simulacrum of laboratory conditions was a feasible and appropriate scientific method to apply, but in development research, unlike laboratory condition treatments, interventions are *not* manipulations of individuated and additive or simply combinable material components (the inducement of a molecular or chemical reaction etc), but rather intervention into material *social relations*. Whilst for the former, assuming away or stripping away everything other than a given effect-focus can reveal the underlying invariant mechanics of that effect, in the latter one cannot take it as given that there is an underlying invariant mechanics that will continue to apply and one is just as liable to be assuming or stripping away what is important to the constitution of the material social relations. Exploration of complex constitution, contingency and variation may be intrinsic to adequate explanation, and thus to appropriately constructed interventions that work in different contexts and work at a societal level (since this is a different order of effect and intervention). This sits awkwardly with RCTs. Moreover, though proponents of RCTs are concerned by the ethics of approval, acquiescence and ownership of interventions, the practical dynamics of this fits awkwardly with the methodological mapping of RCTs – the procedure treats the human as merely consistently manipulable in some significant facet – decomposing the human as though a stimulus-response metaphor was compatible with fully realized humans as centers of ultimate concern; this raises basic issues about what it means for “transparency” to apply and for credibility to be a reasonable claim. As such, RCTs may make for poor *social science*, because the approach is based on a mismatch between the RCT procedure and the constitution of reality under investigation – including the treatment of humans as deliberative centers of ultimate concern. In any case, technical sophistication is no guarantor of appropriately conceived “rigour” if the orientation of methods is inappropriate.

Olsen’s position

The fundamental problems of RCTs all involve a mismatch between social reality and the projection of a science of society for development purposes. The *dominance* of RCTs commits development research to managing this mismatch as though RCTs was *the* scientific approach and as though there was no alternative. This is self-perpetuating. Olsen’s point of departure is simple, why not adopt a methodology that works with (claimed) characteristics of social reality? For realistically posed development research one can look to a (non-foundational, “immanent critique” derived) philosophical realism that is compatible with social construction and, based on its principles, embraces action research (Olsen, 2019a). Situated

social subjects produce, reproduce and transform social reality, subject to the potentials and limitations of the materiality their activity is embedded in and subject to norms and practices (beliefs, habits, ethics etc), according to social position (that may be class based, gendered, and affected by multiple other factors such as ethnicity, caste, religion etc), and subject to irreducible personhood (I and we are more than the sum of socialization or mere roles), all of which varies in place and through time. Social reality is a complex multi-faceted process where the past affects the present and helps to shape the future, but does so subject to historical time, rather than periodised, atomised and regularised, theoretical time (see also Morgan, 2015).

Given this account, one can still *systematically* explore social reality working with rather than against its value base and variation. Social reality has duration and grounds -- events occur according to contexts, and we can explore and explain why events occurred in the way they did, why there may be reasons for the reproduction of similar outcomes and why outcomes can vary or be changed. That is, suitably conceived, we can explore what “causes” events to happen as they have, whilst realizing that this idea of cause is not some simple stimulus-response or repetitive input-output situation analogous to the geometric collision of billiard balls, rather it involves conscious reflexive beings who can flourish or suffer. To flourish or suffer as a human being is something we can reflect upon with reference to the conditions in which we live (so there is a contrastive possibility based on why things have happened as they have and subject to what is deemed possible under the values we hold about what we desire; see Morgan and Patomäki, 2017); furthermore, our involvement in the terms of our own existence is part of what it means to flourish as a fully realized human.

It is with all of the above as background that Olsen argues for a matching methodology in development research: a staged plural and inclusive iterative learning approach that is multi- and inter-disciplinary, mixed-method and collaborative. It converges around explanations of outcomes and around evidence use, but does so based fundamentally on the claim that evidence provides grounds for reasoned disagreement and agreement through “multi-logue” (communication), which in turn, requires participants to be capable of reasoned discussion of values and to be appropriately committed to the process. A sceptic, of course, might respond that this assumes that differentials of power, bias, and interest do not affect standpoints and thus both what is preferred and any possibility of agreement expressed through “learning”.

However, such skepticism misapprehends Olsen’s realist argument. Olsen’s position is not an assumption of something unreal, rather it is an acknowledgement that progress must work with difference according to some meta-ethic that must be adopted for progress to be possible. This is self-referential insofar as the meta-ethic is no more-or-less than a commitment to progress through learning (a good faith criterion).² However, the commitment is not unreal, but rather something we must choose to apply to reality (to realise) based on who and what we are. It would actually be a constructivist paradox to argue we are unable to adopt such a position, since this would be to assume we are unable to shape our own collective situation and unable to adopt others’ points of view to change our own and others’ terms of existence. This meta-ethic does not make agreement and progress

² An RCT proponent might reasonably ask how this resolves or differs from the problem of failure expressed as success in regard of scientific progress (context and reference points would be the short answer).

neutral or simple. As such, it has similar strengths and weaknesses to Habermas's ideal speech – one either adopts a good faith approach or one does not; but any version of progress must seemingly begin from a meta-ethic of good faith and to some degree once one starts to think seriously about an objection, then the objection becomes self-cannibalizing, since no development researcher will position themselves as a carrier of bad-faith, and so cannot start from the assumption that good faith is impossible or that learning, in a position of dispute and difference is impossible (it is merely fallible within what Olsen terms structured or strategic pluralism, shaped by objectivity as open-mindedness, see also Olsen, 2019b). In any case, action research would make little sense without a meta-ethic of good faith. Thereafter, a realist approach is highly attractive because it makes sense of and works to realise this good faith through the informed use of methods – a methodology. As a prominent and highly competent social statistician, Olsen is well placed to make this case – for example, the issue of “equifinality” as an issue of multiple pathways to a given outcome that RCTs can only express in one reductive way (see Olsen, 2012, 2014; also Freedman, 2010; Gillies, 2000).

Conclusion: evidence use and logical forms

The Latin derivation of data is “given”, but data are never “given” if we mean by this that data are passive, data are acquired and processed, and both of these are activities according to some procedure. The collection of data cuts into the social world, presupposing things about it. This cannot be neutral. Data is simply a linguistic formulation that privileges one form of evidence based on the connotations it carries for given methods and how we think about what they deliver. Data carries status, but we need to think again about this and realise that evidence is Catholic. What matters is insight and it is counter-productive to create hierarchies that *a priori* privilege one set of means to acquire insight over others *before* we engage in investigation. This is different than having standards or placing a value on expertise, since one can become expert based on open-minded pluralistic learning. Olsen makes the case for mixed methods use along these lines.

A central theme she articulates is evidence as “traces” and research as “tracing”. Here, how one positions method use and findings is as important as the definitive expression of the procedure. Most datasets start life as surveys or categorized counts conducted somewhere, and it requires a degree of distance from the act to license the amnesia that typically accompanies the use of some datasets. More generally, an analytic statistic has grounds of construction and limits on what it identifies regarding some relation or outcome. However, we are all familiar with research where modelers, advocates of quants, statisticians etc. progress no further than construction of a model and the reporting of a hypothesis and test of a relation (with comment on limits of the model and test). Yet this when done well is still merely indicative in a generalized way of some underlying contingent relation, and so provides no more than a point of departure for an investigation of something that can be explained. Pursuing the explanation necessarily takes the researcher into qualitative territory. However, as Olsen argues, this does not entail a radical transition in evidential terms, since there are ways to systematically explore qualitative evidence and no simple dichotomy of methods according to quantitative and qualitative approaches.

She, for example, is an advocate of Qualitative Comparative Analysis (QCA). QCA provides a means for researchers to explore configurations of factors that affect outcomes in terms of

whether the factors can be inferred to be sufficient or necessary.³ The evidence source is real cases and the method provides specific countable expression of structured sets of different and varied relations. This provides quite a different basis for insights than the sophisticated averaging of a regression, with its frequency based probability expression. For Olsen, both can contribute to development research along realist lines. They are part of “bridging to action” when appropriately positioned as methods that provide evidence within an evaluative process.

However, to justify the use of many methods requires an approach to the logic applied to such methods that reconciles rather than undermines their use. It is with this in mind that Olsen introduces retroduction and highlights that deduction and induction can be nested within retroduction. The logic of methods is not merely additive of modes of expression that might otherwise be deemed incompatible. Consider, for example, Condorcet’s paradox – a person may make decisions that do not follow the logic of transitivity because the significant features that affect what is decided are not identical from one decision to the next – as such, I might prefer A to B and B to C but C to A. What are we to make of this? It is not illogical but rather a reasonable process of reflexive preference, based on the way evidence is positioned and used. Analogically, there are a variety of forms of logic, logical expression can be internal to the construction of a method but also situating of the use of methods and their logical forms. It is implicit to Olsen’s argument that advocates of quants and statistics would do well to keep this distinction in mind, without its use of methods and exclusion of others can vandalize the world in the name of preserving (inviolable) logics of methods. RCTs, arguably, commit this error. However, the point is merely one facet of the mechanics of Olsen’s argument; what is most important is the emphasis on “bridging” that the argument for retroduction facilitates, and yet, ultimately, bridging *to* action is a (good faith) practice, and if we take an action research perspective to this, it is bridging *in* action that demonstrates its value.

References

- Banerjee, A. and Duflo, E. editors (2017) *Handbook of Economic Field Experiments* (two volumes) Oxford: North Holland-Elsevier
- Deaton, A. and Cartwright, N. (2018) ‘Understanding and misunderstanding randomized control trials’, *Social Science & Medicine* 210: 2-21
- Freedman, D. (2010) *Statistical Models and Causal Inference* Cambridge: Cambridge University Press
- Gillies, D. (2000) *Philosophical Theories of Probability* London: Routledge

³ In an ordinary language sense, necessary implies cannot happen *without* this (unless x no y) but need not mean if *only* x then y, since it need not unless stated exhaust the conditions for y to be the case (unless I am a man I cannot be a king); sufficient implies enough for (if x then y) but does not imply “no x then no y”, (one can have no x but still y, unless it is also necessary). Nuance of statement matters for the conditionals (if I am a man then I may be a king, but if I am a man I need not be a king, however if I am a king I must be a man etc) and there is a great deal more to context and argument for what is necessary and what is sufficient and there is a long tradition of discussion of this in philosophy (including the difference between analytical truths of statements and matters of real causal conditions). For example, J. L Mackie’s INUS: something may be an insufficient but non-redundant part of an unnecessary yet sufficient condition (e.g. x is necessary within a sufficient condition, but is not itself sufficient for y – a spark and concentration of flammable gas in a confined space produces an explosion, but a spark alone does not).

- Morgan, J. (2015) 'Seeing the potential of realism in economics', *Philosophy of the Social Sciences* 45(2): 176-201
- Morgan, J. and Patomäki, H. (2017) 'Contrast explanation in economics: its context, meaning, and potential', *Cambridge Journal of Economics* 41(5): 1391-1418
- Olsen, W. (2019a) 'Bridging to action requires mixed methods not randomized control trials', *European Journal of Development Research*, this issue
- Olsen, W. (2019b) 'Social statistics using strategic structuralism and pluralism', Chapter 8, in Nagatsu, M and Ruzzene, A. editors, *Contemporary Philosophy and Social Science: An interdisciplinary dialogue* London: Bloomsbury
- Olsen, W. (2014) 'Comment: The usefulness of QCA under realist assumptions' *Sociological Methodology* 44(1): 101-107
- Olsen W. (2012) *Data Collection* London: Sage
- Ravallion, M. (2012) 'Fighting poverty one experiment at a time: A review of Abhijit Banerjee and Esther Duflo's "Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty"', *Journal of Economic Literature* 50(1): 103-114
- Sindzingre, A. (2019) 'Experiments in economics and their ethical dimensions: The case of developing countries' in Dolfsma, W. and Negru, I. editor *The Ethical Formation of Economists* London: Routledge

Jamie Morgan is Professor of Economic Sociology at Leeds Beckett University. He co-edits the *Real-World Economics Review* with Edward Fullbrook. He has published widely in the fields of economics, political economy, philosophy, sociology, and international politics. His recent books include *Realist Responses to Post-Human Society: Ex Machina* (ed. with I. Al-Amoudi, Routledge, 2018); *Brexit and the Political Economy of Fragmentation: Things Fall Apart* (ed. with H. Patomäki, Routledge, 2018); *Trumponomics: Causes and consequences* (ed. with E. Fullbrook, College Publications, 2017); *What is neoclassical economics?* (ed., Routledge, 2015); and *Piketty's capital in the twenty-first century* (ed. with E. Fullbrook, College Publications, 2014).