This is a repository copy of *Moving beyond randomized controlled trials in the evaluation of compulsory community treatment*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/154156/

Version: Published Version

Article:

Duncan, C., Weich, S. orcid.org/0000-0002-7552-7697, Moon, G. et al. (10 more authors) (2020) Moving beyond randomized controlled trials in the evaluation of compulsory community treatment. Journal of Evaluation in Clinical Practice, 26 (3). pp. 812-818. ISSN 1356-1294

https://doi.org/10.1111/jep.13245

This is the peer reviewed version of the following article: Duncan, C, Weich, S, Moon, G, et al. Moving beyond randomized controlled trials in the evaluation of compulsory community treatment. J Eval Clin Pract. 2019, which has been published in final form at https://doi.org/10.1111/jep.13245. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Moving beyond randomised controlled trials in the evaluation of compulsory community treatment

Short running title: Beyond trials in community treatment evaluation

Craig Duncan[a*], Scott Weich[b], Graham Moon[c], Liz Twigg[a], Sarah-Jane Fenton[d], Kamaldeep Bhui[e], Alastair Canaway[f], David Crepaz-Keay[g], Patrick Keown[h], Jason Madan[f], Orla McBride[i], Helen Parsons[f], Swaran Singh[f]

[a] Department of Geography,
University of Portsmouth,
Portsmouth PO1 3HE, UK
** Corresponding author (+44 23 92842493; craig.duncan@port.ac.uk)*

[b] School for Health and Related Research,
University of Sheffield, UK

[c] Geography and Environment,
University of Southampton, UK

[d] Institute for Mental Health,
University of Birmingham, UK

[e] Centre for Psychiatry,
Queen Mary University of London, UK

[f] Warwick Medical School,
University of Warwick, UK

[g] Mental Health Foundation, UK

[h] Institute of Neuroscience
Newcastle University, UK

[i] School of Psychology
Ulster University, UK

Abstract

Compulsory community treatment for people with severe mental illness remains controversial due to conflicting research evidence. Recently, there have been challenges to the conventional view that trials-based evidence should take precedence. This paper adds to these challenges in three ways. First, it emphasises the need for critiques of trials to engage with conceptual and not just technical issues. Second, it develops a critique of trials centred on both how we can have knowledge and what it is we can have knowledge of. Third, it uses this critique to develop a research strategy that capitalises on the information in large-scale datasets.

Keywords

compulsory community treatment; mental health policy; realist evaluation; randomised controlled trials; clinical effectiveness

Introduction

Although widely adopted around the world, compulsory community treatment (CCT) for people with severe mental illness remains highly controversial.[1] Stemming from intrinsic concerns about its coercive basis, the controversy has been heightened considerably by conflicting evidence about CCT's effectiveness.

Of key significance has been a division between findings from research based on randomised controlled trials (RCTs) with those based on a variety of non-experimental designs. While the latter have shown both positive (e.g. increased follow-up with mental health services and improved forensic outcomes) and negative outcomes (e.g. no reduction in hospitalisation), the results of the former have overwhelmingly been negative.

Following conventional biomedical thinking, the findings from trials have tended to have been given priority. Recently, however, some workers have challenged this position, arguing that while RCTs have advantages for discrete, single component clinical interventions, they may not be intrinsically superior for the evaluation of complex multi-component interventions such as CCT.[2,3]

This paper aims to contribute to this challenge in three ways. First, it draws out, and emphasises the need, for critiques of randomised trials of CCT to move beyond technical problems and consider conceptual/theoretical issues. Second, it draws on realist notions of open and closed systems to develop a conceptual critique of trials that considers both how we can have knowledge (epistemology) and what it is we can have knowledge of (ontology).

Third, it outlines the key implications of this critique before using them to develop a novel and innovative research methodology capable of capitalising, but not relying, on the wealth of information in increasingly available large-scale datasets.

Randomised trials: the technical critique

There is now a large body of evidence on the effectiveness of CCT including a number of reviews.[4-6] While the vast majority of this evidence has been non-experimental in design, three randomised trials have been conducted: the first two in America; the third, and most recent, in England.[7-9]

Although much fewer in number, adherence to the traditional hierarchy of evidence has meant the findings and conclusions of the three randomised trials have often been given much greater credence. As would be expected, attachment to this way of thinking has permeated a Cochrane review process (and associated meta-analyses) with only a tiny fraction of observational studies being included.[10-12] It has, however, also been prominent in less structured reviews where ‘further RCTs’ have been seen as the way of providing definitive evidence.[6] Although a similar more recent review avoids such a conclusion, it still placed greater emphasis on the problems associated with non-experimental studies.[13]

Despite their higher standing, several workers have drawn attention to weaknesses with the randomised trials of CCT. Given the well-recognised ethical problems associated with randomising patients with severe mental health problems to experimental intervention and control groups, this is, perhaps, not surprising. Indeed, it is evident in the history of the trials themselves. The early American trials were widely recognised as having a number of difficulties including the exclusion of patients with a history of violence, high attrition rates, small sample sizes and protocol violations. In light of this, the later English trial was seen as an opportunity to put things right and provide ‘the missing link – a convincing RCT’.[14]

Since publication, however, a number of commentators have also drawn attention to problems with this later trial. For Kisely, the concerns raised crystallise around two main issues – potential bias, especially selection bias, and the nature of the control group.[15] Factors contributing to the first include: the exclusion of patients not capable of giving consent, the inability to include patients viewed by clinicians as clear candidates for CCT, the decision by patients not to participate, and the switching of patients between arms of the study after randomisation. In terms of the second, legal and ethical approval could only be obtained for the comparison of people on CCT with those on section 17 leave which is, itself, a form of CCT. As a result, it has been argued that the trial did not in fact compare CCT against no CCT, thus narrowing the chances of finding significant differences on the selected outcomes.[16]

There is, therefore, a well-established and clearly de-lineated technical critique of randomised trials within the CCT literature. Significantly, as already noted, these problems, although widely recognised, are often not sufficient (or, at least, not perceived as being a more significant than those associated with other research designs) to unsettle a belief in the relevance of the hierarchy of evidence for research on CCT.

Randomised trials – a conceptual critique

The difficulty with trials in CCT research is not, however, simply technical: it consists of more than practical or ethical problems. Put slightly differently, while technical fixes (i.e. better designed trials) might be useful, they would not eliminate all of the difficulty. Rather, there is a more fundamental problem due to the conceptual bases and assumptions (the 'world view') on which trials are based. Put most simply, there is a mismatch between the object of study and the methods of study. While one way in which this manifests itself is through technical difficulties, there is also a more significant conceptual non-alignment. One fruitful way to think of this is through the notion of closed and open systems as developed within realist philosophies of science.[17]

Closed systems are those in which objects with causal powers do not change in themselves (the intrinsic condition of closure), nor vary according to the external conditions in which they occur (the extrinsic condition of closure). When there is closure, it is reasonable to expect causal mechanisms to manifest themselves as regularities. While closed systems can occur naturally, they can also be created by intervention. In fact, this is the aim and purpose of randomised trials. Thus, by isolating a mechanism (the intervention), and triggering its outcome in a stable context (equivalent treatment and control groups), regularities in trials (succession of events) come to stand as causality.

Open systems, meanwhile, are those in which either, or both, of the conditions of closure are not met. As the social world in which CCT occurs is undoubtedly open – the actors involved are capable of reflection and change, they are affected by the contextual settings in which they act and, indeed, the system of CCT itself can change and alter – it is unreasonable to expect regularities and causality must be found in deeper levels of explanation.

One useful approach arising from the challenge of open and closed systems - realist evaluation (RE) – has been recognised in recent work on CCT.[3] RE emphasises that interventions in and of themselves do not 'work', rather they are shaped by social actors - it is the people involved in them, and the circumstances in which those people find themselves, that determine an intervention's capacity to work (or not).[18] Thus, while the content of interventions may act as generative causal mechanisms, these mechanisms can be activated ('fired') or lie dormant, depending on the salience of different components to those applying interventions, and the context in which the social actors involved are situated. This thinking is summarised in RE's key formula:

Context (C) + Mechanism (M) = Outcome (O)

As O'Reilly and Vingilis argue, therefore, a key part of the difficulty with trials for research on CCT is epistemological: different approaches are needed to get knowledge of the world.

Notions of closed and open systems as outlined here, however, remind us that there is a further challenge that is more ontological in nature - what is the (social) world like that we want to have knowledge of? Implicit within the positivist-inspired thinking that shapes RCTs is a view of the world as stable, regular and typified by universal rules or laws. In contrast, realist notions of open and closed systems suggest a world which, although not random, is characterised by patterned variations.[19] Crucially, such patterns, or 'demi-regularities' as they have been called,[20] can provide an ideal starting point for uncovering the mechanisms generating them. Identifying and examining 'demi-regs' offers, therefore, an important, but so far neglected, way for realist-inspired research on CCT to get a handle on causal mechanisms.[21] In this way, they can help secure a switch from the focus on overall 'average' effects advanced by trials-dominated approaches to the alternative RE goal of investigating and taking seriously diverse, contingent, situated outcomes or, as it's often summarised, establishing *what works, for whom, under what circumstances and why*.

Realist principles for the evaluation of CCT

Before outlining one way that the analysis of demi-regs may proceed and how it can connect with the approach outlined by O'Reilly and Vingilis, it is worth briefly drawing out the general lessons from RE and its philosophical antecedents so as to construct some general principles for future empirical research on CCT. Again, these principles aim to move research on CCT beyond technically dominated discussions of RCTs.

First, it comes as no surprise from a realist-inspired perspective that the research findings on CCT have been mixed. Indeed, one worker describes this as the 'iron law of evaluation': 'the expected value of any net impact assessment of any large-scales social program is zero'.[22] When interventions are seen as centring on people behaving in an open-system social world, and when the contexts comprising that social world are taken seriously, then the workings of mechanisms, and the outcomes arising from them, will only ever be contingent, multiple and varied. CCT will not work for all people, in all places, at all times - no human, social-based intervention does. Outcome variation is to be expected and the findings from research will be inconsistent. From a realist perspective, the job is to find out for which types of patients, in which types of circumstances CCT works best, not whether it works 'as a whole'. Research needs to be able to anticipate, handle and explore the variation – the demi-regs - that this implies.

Second, the complexity of the open-system social world means no single research method has privileged access to it. Accordingly, realist-inspired evaluation does not champion or disparage any particular research design but argues for a plurality of methods and sensitivity to sources of variation and causation. It is important to note that that this may include RCTs, but it would not be restricted to them, nor privilege them. Extensive quantitative research capable of detecting the variation and heterogeneity that RCTs struggle to capture would be an essential counterpart. Equally valuable, however, is intensive qualitative research, particularly when it moves beyond simply recording the views and experiences of participants to uncover the capacities of human agents to respond to the resources embodied in mechanisms.[23]

Lastly, accepting the need for a plurality of methods necessitates a more pragmatic attitude to evidence. If there are no 'crucial experiments', or the only ones possible are far removed from the ideal type necessary for the advantages of experimental design to apply, then the application of traditional evidence hierarchies is no longer appropriate. Furthermore, treating interventions as the product of situated human action and interaction means a ‘formula and rulebook’ approach is not as important as one that follows processes of implementation and ‘learns as it goes’.

Beyond randomised trials: another way forward for research on CCT

As already noted, when the critique of trials is restricted to the technical, the necessary solution is better-designed trials. Triallists have, of course, made considerable advances in this area, developing a number of extensions to the basic parallel group design such as stepped wedge and patient preference designs. Indeed, the UK Medical Research Council have argued these types of trials are more appropriate for evaluating complex interventions such as CCT.[24] Proponents of realist-inspired thinking are, of course, more sceptical. From their perspective, and as argued here, experimental designs have intrinsic ‘blind spots’, rather than just technical problems to which a fix or extension can be applied. As one of the originators of RE has succinctly put it, “one cannot design out complexity”.[25]

O’Reilly and Vingilis outline a typical RE approach to working with this complexity (as well as a Theory of Change approach) consisting of four stages: (i) talking to program implementers; (ii) drawing up mini-theories based on the formula outlined above; (iii) assessing the mini-theories in light of outcomes, (iv) developing and refining the mini-theories. In this final section, we add to their work in light of our conceptual critique and the principles we have drawn from it. In brief, we respond to the specific need for a research strategy capable of detecting and unravelling multiple outcomes in a heterogeneous, open system social world.

To start, we would like to emphasise the importance of recognising different ways of thinking of context. For O'Reilly and Vingilis, context is seen in terms of social processes/structures: in their illustrative example, for example, the context is homelessness. As Sayer's seminal work emphasised, however, social processes/structures always have geographies (and histories for that matter).[19] Not only do these geographies provide a setting or back-drop, but they also make a difference to the social processes/structures themselves. Homelessness is not singular or uniform, instead it is multiple and complex. Homelessness in one place will not be the same as homelessness in another. Thus, the actual workings and effects of mechanisms will depend on the specific geographical settings in which they operate. Although not without problems of its own, by treating context in more explicitly geographical terms, (i.e. as spatially-based locales, for example, areas of residence or specific health service provider settings), we have a way of producing bounded ecological systems that can help provide a way-in to the operation of causal mechanisms in empirical research. In short, interventions come together in particular places in specific arrangements with specific effects.

When context is seen in terms of geographical-based entities, it becomes possible to outline a research strategy in tune with open social systems that combines the strengths of both extensive and intensive research methods. Importantly, this strategy is also able to exploit the richness of increasingly available large-scale administrative routine datasets without being entirely dependent on them. The strategy consists of three phases.

In the first phase, extensive quantitative research based on large-scale datasets could be used to identify 'outliers' – contexts where CCT outcomes are unusual, either particularly good or particularly poor. Returning to the conceptual ideas covered earlier, this phase is concerned with getting to grips with the demi-regularities of open social systems. The profile of outcomes could, of course, be different for different types of people such that there is an interaction between patients and the contexts, or levels, through which, and at which, CCT operates.

Multilevel modelling approaches provide one particularly effective way of identifying anomalous settings. Traditional forms of statistical analysis focus on finding an average effect based on a large number of patients in one, or few, settings and assume it can be generalised elsewhere. Multilevel modelling approaches, in contrast, focus on the variation in the average effect that exists across a large number of settings and assume it applies nowhere.[26] Such approaches provide, therefore, a way of seeing both the general and the specific – estimates of the relationship between variables across all places, and estimates of the extent to which places differ from this relationship. By explicitly recognising the way in which patients nest within a variety of real-world contexts, some of which may potentially overlap, the method allows a series of technical benefits to be harnessed – standard errors can be adjusted for autocorrelation and estimates of the differences between contexts can be precision-weighted. In this way, extensive research can be both geographically more realistic and statistically more accurate.

As anticipated almost two decades ago now, the setting is the real unit of analysis in this type of work and research designs need to be powered according to the number of these, not the number of patients.[27] Given this, being able to use large-scale, representative administrative datasets is an intrinsic feature of this approach and a considerable strength. As others have noted, RCT-based approaches usually attempt to factor out the effect of context by randomising or controlling the settings that are included.[28] Here, in contrast, the full range of real-world settings captured in the data is worked with purposefully and willingly so that any variation that exists can be brought to light.

The unusual contexts, or ‘contrastive demi-regularities’, identified in the first phase become the focus for the second phase of research: intensive qualitative research in the places where outcomes are most different. As has been noted by others, adopting a ‘most different’ comparative approach maximises the scope for theory development.[29] Thus, we have a way of targeting the places where it is most likely that more conventional RE methods could be used most productively.

It is important to emphasise that this second phase of work must move beyond simply recording the perceptions and views of those involved in CCT to concentrating on how causal mechanisms are constituted and operate in different places. In broad terms, CCT seeks to bring about behavioural control through mandatory programmes. Often, though, these programmes are reduced simply to ‘coercion’. Patients will, however, respond to coercion in different ways depending on their capacities, as will the clinicians responsible for determining and enforcing it. This last point is key, for while it is often said that CCT binds patients to clinicians, it also binds clinicians to patients. Intensive research is, therefore, necessary to open the ‘black box’ constituting the actual mechanisms underlying CCT. Whether CCT will be effective will depend on: the workloads of clinicians, the background of patients, levels of familial support, access to community mental-health services, the range of additional support services, the availability of suitable accommodation. The role and balance of these factors will, of course, depend on the specific, local geographical context in which they exist. Useful work in this vein has started to be produced and has begun to illuminate the intricacies involved in the human-based social intervention that is CCT.[23] It is interesting to note, however, that only two settings were studied in this work, with no attempt seeming to be made to utilise a 'most different' comparative design so as to maximise theory development.

The final phase of research would, using Pawson's words, be to 'scavenge' amongst the range of evidence developed through this plurality of methods and subject it to 'organised scepticism'.[25] This would involve a collegiate group of stakeholders coming together to assess carefully whether the evidence supports the inferences drawn from it. Crucially, all those engaged in this would have to put aside any tribal loyalties they might have, including the rigid adherence to traditional hierarchies of evidence. Significantly, it will produce results and outputs very different from those produced by trials: rather than average treatment effects (or sub-group components, thereof), it will consist of rich, complex pictures more like narratives or field descriptions.[30]

It is in this final phase that we would see again the value of seeing context through a more geographically-based lens. As others have emphasised, evidence-based policy is not simply about finding out if an intervention works *somewhere*.[31] It is also about finding out if it will work *here*: a new location where the intervention hasn't been tried or a location where it is clear that it is not working but could be made to do so if suitable modifications were made. For interventions to travel successfully it is necessary to know how they work and the factors needed to support their working. Since interventions are always implemented in a specific place, *here,* not just *somewhere*, viewing context in terms of geographical-based locations offers an important way of helping ensure research moves from '*it worked somewhere*' to the more useful conclusion of, '*it will work here'.* Applicability and transferability to and from particular settings is more relevant than notions of simple generalisibility and these are likely to be fostered to a greater extent when context is treated in a more explicitly geographically based fashion.[32]

Discussion

The technical limitations of randomised trials for research on CCT are well-recognised. Indeed, some workers have suggested that the difficulties of conducting trials mean it is unlikely that many others will be attempted.[15] This has not, nevertheless, diminished their standing or a sense that other methods are, in comparison, an inferior fall-back option. This paper emphasises that the problem with trials is, however, never simply technical. It is, instead, always conceptual. The importance of this distinction is considerable, not least because of the tendency for trials to be seen as a way of delivering atheoretical, assumption-free impact evaluation.[33]

In contrast to other recent work, this paper articulates arguments as much on the basis of what we can expect the social world to be like as how we can have knowledge of it. In this way, it develops and extends these attempts to move research on CCT beyond trials. The research strategy outlined offers a means of integrating different methodological approaches to mental health services research in the manner imagined by Slade and Priebe: quantitative analysis centred on finding differences in outcomes dovetailing with qualitative, ethnographic research targeted at the differences found. The former finding where and for whom things are different; the latter establishing how and why things are different.

Importantly, this approach is able to capitalise on the ever-increasing investment in 'big data' without being over-exposed to its limitations or losing sight of the need to complement its version of the truth with that obtained from other methods.[34]

The approach outlined here is, of course, not without its own difficulties. Operationalising context in geographical based terms (either in terms of residential areas or in terms of service providers) will, itself, be partial in some way. Geographical/institutional boundaries are never fixed, impermeable or all-embracing. As is well-known, study results can be prone to the modifiable area unit problem.[35] While multilevel models help provide a way of reflecting the layered-nature of social reality, notions of nested hierarchies can imply that causal processes run downwards and that mechanisms associated with one level can be neatly separated from those at other levels.[36] It is, therefore, in part for these reasons that the use of such models is seen only as a first step in the approach outlined here. Imperfect though they are, we believe such models can still provide a valuable starting point for empirical research on complex interventions in complex open systems.

Using administrative data also brings with it its own challenges. Although increasingly available, they are still not present across many parts of health systems. Perhaps of more significance, however, is the fact that when data is routinely collected it is often of low quality such that it can be characterised as FUPS: Flawed, Uncertain, Proximate and Sparse.[37] As the workers who have coined this acronym show, however, such data can still be used to instigate useful, important conversations in the 'swampy lowlands' of everyday health care practice. It is, therefore, very much in this spirit that we conceive of its use here: as opening up a way of directing focused intensive qualitative-based research. In contrast to the data from RCTs, the fact that it is continuously collected, long-lasting, large-scale and grounded in real world settings emphasise a need to find ways of using it. Making use of it in the way suggested here – as a first-step guide – means the dangers of non-use can be avoided while the dangers of over-interpretation are minimised.

As we have noted, other work has sought to move research on CCT beyond RCTs. While this work has usefully emphasised the significance of CCT being a complex intervention, it has not so clearly drawn out (or on) the significance of the systems in which such interventions occur as being complex.[38] We have drawn on realist perspectives of science to make this point. It should be noted, however, that there are other ways of approaching this. Complexity theory is one alternative which is receiving much attention at the moment, though it has a longer pedigree, even within health research.[39] While there are important differences between realist and complexity-based thinking, leading to proponents of the former having serious reservations about the latter,[25] it is likely that there will be increasing, productive engagement between them.[40] If nothing else, both approaches unite in throwing into light the limitations of context-free RCTs.[28]

Finally, it should be emphasised that by moving beyond trials we do not mean abandoning them or dismissing them outright. The real villain of the piece is not the trial but the traditional hierarchy of evidence. Like those who have worked so hard on one of the trials of CCT, we believe wholeheartedly that mental health care needs to be evidence-based.[1] For us, though, evidence cannot be thought of in such clearly self-evident, unambiguous hierarchical terms.[41] All evidence has blind spots and prejudices and, as we hope to have shown here, these are not just technical in nature. As workers in other areas have emphasised, holding on to this hierarchy will mean research on CCT remains unbalanced, impartial and unable to capture what is required for effective real world practice.[33]

References

1 Rugkåsa J, Dawson J, Burns T. CTOs: what is the state of the evidence? *Social Psychiatry and Psychiatric Epidemiology.* 2014;49:1861-1871.

2. Mustafa FA. Notes on the use of randomised controlled trials to evaluate complex interventions: Community treatment orders as an illustrative case. *Journal of Evaluation in Clinical Practice*. 2017;23:185-192.

3. O'Reilly R, Vingilis E. Are randomized control trials the best method to assess the effectiveness of community treatment orders?. *Administration and Policy in Mental Health and Mental Health Services Research*. 2018;45:565-574.

4. Churchill R, Owen G, Singh S, & Hotopf M. *International experiences of using community treatment orders.* London: Institute of Psychiatry; 2007.

5. Maughan D, Molodynski A, Rugkåsa J, & Burns T. A systematic review of the effect of community treatment orders on service use. *Social Psychiatry and Psychiatric Epidemiology*. 2014;49:651-663.

6. Rugkåsa J. Effectiveness of community treatment orders: the international evidence. *Canadian Journal of Psychiatry.* 2016;61:15-24.

7. Swartz MS, Swanson JW, Wagner HR, Burns BJ, Hiday VA, & Borum R. Can involuntary outpatient commitment reduce hospital recidivism?: findings from a randomized trial with severely mentally ill individuals. *American Journal of Psychiatry.* 1999;156:1968-1975.

8. Steadman HJ, Gounis K, Dennis D, Hopper K, Roche B, Swartz M, & Robbins PC. Assessing the New York City involuntary outpatient commitment pilot program. *Psychiatric Services.* 2001;52:330-336.

9. Burns T, Rugkåsa J, Molodynski A, Dawson J, Yeeles K & Vazquez-Montes M, et al. Community treatment orders for patients with psychosis (OCTET): A randomised controlled trial. *Lancet.* 2013;381:1627–1633.

10. Kisely SR, Campbell LA, & O'Reilly R. Compulsory community and involuntary outpatient treatment for people with severe mental disorders. *Cochrane Database of Systematic Reviews.* 2017;3.

11. Kisely S, Campbell LA, Scott A, Preston NJ, & Xiao J. Randomized and non-randomized evidence for the effect of compulsory community and involuntary out-patient treatment on health service use: systematic review and meta-analysis. *Psychological Medicine*. 2007;37:3-14.

12. Kisely S, & Hall K. An updated meta-analysis of randomized controlled evidence for the effectiveness of community treatment orders. *Canadian Journal of Psychiatry*. 2014;59:561-564.

13. Barnett P, Matthews H, Lloyd-Evans B, Mackay E, Pilling S, Johnson S. Compulsory community treatment to reduce readmission to hospital and increase engagement with community care in people with mental illness: a systematic review and meta-analysis. *The Lancet Psychiatry.* 2018;5:1013-1022.

14. Burns T, & Dawson J. Community treatment orders: how ethical without experimental evidence? *Psychological Medicine.* 2009;39:1583-1586.

15. Kisely S. Assessing the effectiveness of compulsory community treatment. In A Molodynski, J Rugkåsa & T Burns (Eds.), *Coercion in community mental health care* (pp.75-94). Oxford: Oxford University Press; 2016.

16. Hastings TJ, & Gray JE. Community treatment orders disconnect. *Canadian Journal of Psychiatry*. 2016;61:435-436.

17. Bhaskar R. *A realist theory of science*. Brighton: Harvester Press; 1978.

18. Pawson R & Tilley N. *Realistic Evaluation*. London: Sage; 1997.

19. Sayer A. *Method in social science: A realist approach*. London: Routledge; 1992.

20. Lawson T. Abstraction, tendencies and stylised facts: a realist approach to economic analysis. *Cambridge Journal of Economics.* 1989;13:59-78.

21. Jones K. The practice of quantitative methods. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences.* (pp. 201-211). London: Sage; 2010.

22. Rossi PH. The iron law of evaluation and other metallic rules. *Research in Social Problems and Public Policy.* 1987;4:3-20.

23. Jobling H. Using ethnography to explore causality in mental health policy and practice. *Qualitative Social Work*. 2014;13:49-68.

24. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, & Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal.* 2008;337:a1655.

25. Pawson R. *The Science of Evaluation*. London: Sage; 2013.

26. Merlo J, Ohlsson H, Lynch KF, Chaix B & Subramanian SV. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of Epidemiology & Community Health*. 2009;63:1043-1048.

27. Slade M, and Priebe S. Conceptual limitations of randomised controlled trials. In S. Priebe & M. Slade (Eds.), *Evidence in Mental Health Care* (pp.101-108). Hove: Brunner-Routledge; 2002.

28. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. *The Lancet*, 2017;390:2602-2604.

29. Monaghan M, Pawson R, & Wicker K. The precautionary principle and evidence-based policy. *Evidence & Policy: A Journal of Research, Debate and Practice.* 2012;8: 171-191.

30. Greenhalgh T, & Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Medicine*. 2018;16:95.

31. Cartwright N, & Hardie J. *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press; 2012.

32. Craig P, Di Ruggiero E, Frolich KL, Mykhalovskiy E, White M, Campbell R. et al. *Taking account of context in population health intervention research: guidance for producers, users and funders of research.* Southampton: NIHR Journals Library; 2018.

33. Ravallion M. Should the randomistas (continue to) rule? *Centre for Global Development Working Paper*, *492; 2018*.

34. Schofield P. Big data in mental health research–do the ns justify the means? Using large data-sets of electronic health records for mental health research. *BJPsych Bulletin*. 2017;41:129-132.

35. Pearce JR. Complexity and uncertainty in geography of health research: Incorporating life-course perspectives. *Annals of the American Association of Geographers*. 2018;108:1491-1498.

36. Byrne D. Complex realist and configurational approaches to cases: A radical synthesis. In D. Byrne & C.C. Ragin (Eds.), *The Sage handbook of case-based methods* (pp.101-112). London: Sage; 2009.

37. Wolpert M, & Rutter H. Using flawed, uncertain, proximate and sparse (FUPS) data in the context of complexity: learning from the case of child mental health. *BMC Medicine*. 2018;16:82.

38. Moore GF, Evans RE, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C, & Murphy S. From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation.* 2019;25:23-45.

39. Plsek PE, & Greenhalgh T. The challenge of complexity in health care. *British Medical Journal.* 2001;323:625-628.

40. Williams M, & Dyer W. Complex realism in social research. *Methodological Innovations*. 2017;10:1-8.

41. Mercuri M, Baigrie BS. What counts as evidence in an evidence-based world? *Journal of evaluation in clinical practice*. 2019 Jul 3.

Acknowledgements

This work was supported by the National Institute for Health Research (NIHR) Health Services and Delivery Research Programme (project number 14/52/40). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, NIHR, NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC), the Health Services and Delivery Research (HS&DR) programme or the Department of Health.

Conflict of interest

All authors received financial support from the National Institute for Health Research Health Services and Delivery Programme for the submitted work.