Testing and refining middle-range theory in evaluations of public-health interventions: evidence from recent systematic reviews and trials

Christopher Bonell[1*]
Ruth Ponsford[1]
Rebecca Meiksin[1]
G.J. Melendez-Torres[2] (N.B. This individual wishes to be known by initials not name)

1 Department of Public Health, Environments and Society, London School of Hygiene & Tropical Medicine, London, UK.
2 Medical School, University of Exeter, Exeter, UK.
* corresponding author: Professor Chris Bonell, Department of Public Health, Environments and Society, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. chris.bonell@lshtm.ac.uk. Tel. +44 (0)20 7612 7918. Fax +44 (0)20 7436 5389.

Word count: 3000

Contribution: CB drafted the paper based on a discussion among all authors. GJMT, RP and RM edited the paper before CB finalised it. CB is responsible overall for the paper's contents.

Testing and refining middle-range theory in evaluations of public-health interventions: evidence from recent systematic reviews and trials

Abstract

Evaluations of public-health interventions might potentially be used to test and refine middle-range theory (i.e. theory about the mechanisms which generate outcomes that is analytically generalizable enough to span a range of contexts, interventions or outcomes, but specific enough to be salient in a given application). This approach has been suggested as one means of developing more informed assessments of how different interventions work and whether mechanisms might transfer across contexts. However, studies included in some of our recent systematic reviews are not oriented towards helping test middle-range theory because interventions draw on multiple middle-range theories (so that it is difficult to draw any conclusions about each middle-range theory based on their results) and these middle-range theories are insufficiently clear (with vague constructs) or parsimonious (with too many constructs) to be readily testable. Some studies might in future better contribute to testing and refining middle-range theory via focusing on interventions informed by one middle-range theory and focused on one mechanism at a time. Such 'proof-of-principle' studies should draw on middle-range theory that is sufficiently clear and parsimonious to allow such testing. These evaluations might facilitate more rigorous testing of middle-range theory and hence refinement of scientific knowledge. They might inform broader assessments of how mechanisms transfer across contexts aiding the development of future public-health interventions. Such studies would be a complement not an alternative to pragmatic studies of scalable complex interventions, often informed by more than one middle-range theory.

240 words

What is already known on this topic
In public-health, intervention theories of change are commonly informed by middle-range theory.

What this study adds
Evaluations of public-health interventions are not oriented towards helping test middle-range theory because interventions draw on multiple middle-range theories and these theories are insufficiently clear or parsimonious to be testable.

How this study might affect research, practice or policy
'Proof-of-principle' studies might contribute to testing and refining middle-range theory via focusing on interventions informed by one middle-range theory and focused on one mechanism at a time, and choosing middle-range theories sufficiently clear and parsimonious to allow testing. Such evaluations might facilitate refinement of scientific knowledge and inform broader assessments of how mechanisms transfer across context.

Testing and refining middle-range theory in evaluations of public-health interventions: evidence from recent systematic reviews and trials

Introduction

According to Popper's hypothetic-deductive approach, experiments should aim to falsify hypotheses derived from theory.[1] Theories are models for explaining or predicting causal mechanisms generally or in defined circumstances. Hypotheses are empirically-testable predictions about specific mechanisms in specific circumstances informed by theories. Knowledge is said to consist of theories subject to experimental tests which have not (yet) been falsified. Where hypotheses are falsified, theories should be abandoned, or refined so they can then be further tested.

Critical realists and realist evaluators accept that science should test hypotheses derived from theory but contend that such testing often does not straightforwardly determine falsification.[2][3] Realists argue that empirically-unobservable but nonetheless real causal mechanisms generate observable events but do so differently between contexts because of variations in the other mechanisms operating. Therefore, if a mechanism does not generate observed events as hypothesised, this may necessitate refinement rather than rejection of the theory so that it better specifies the contexts in which the mechanism might generate outcomes.[3] This position is not very different to Popper's idea that scientific theories need to state the contingencies that underpin causal mechanisms.[4]

In public health, evaluations employing various quantitative designs examine intervention effectiveness. Such evaluations can test whether intervention theories of change appear valid i.e. whether use of intervention resources appears to trigger expected mechanisms generating outcomes.[5] First popularised by Weiss, theories of change are intervention-specific, defining how implementation of intervention activities generates particular outcomes.[6][7] These are distinct from theories of implementation which describe the processes whereby interventions are enacted.[8] Weiss popularised theory-based evaluation, which focuses evaluation on testing theories of change.[9] Realist evaluators have argued that theories of change should define how mechanisms interact with context to generate outcomes (termed context-mechanism-outcome configurations), defining mechanisms as responses, actions and interactions triggered by introduction of intervention resources into a social setting which potentially generate outcomes.[3][10]

Intervention theories of change are often informed by middle-range theory (MRT).[11] This is theory about the general mechanisms (i.e. not necessarily arising from a particular intervention) which generate outcomes that is analytically generalizable enough to span different contexts, interventions or outcomes, but specific enough to be salient in a given application. Basing intervention theories of change on MRT should be good for the intervention and for refining the MRT. It should be good for the intervention because assumptions are made explicit and may be critically considered before an intervention is finalised. If the MRT is itself supported by empirical evidence, it is a means by which interventions may be informed by scientific knowledge.[12] It should also be good for the MRT because evaluation provides an opportunity for empirically testing not only the intervention theory of change (as is the case in theory-based evaluation[9]) but also the MRT. This has not been a focus previously for evaluation. Realist evaluators recommend using MRT to help define context-mechanism-outcome configurations.[3] However, neither they nor those advocating theory-based evaluation have argued that evaluation can test MRTs.

This is a missed opportunity. Consider the example of an intervention subject to rigorous evaluation where the intervention is delivered and reaches participants as planned . The evaluation reports the intervention does not achieve the intended outcomes described by its theory of change. This would suggest that the specific theory of change for the intervention might be incorrect, at least as applied

to the particular setting, population, intervention, comparison and outcome in question. If the theory of change is informed by a single, specific MRT, it might also be used to draw conclusions about the MRT in question. A single evaluation would not definitively falsify the MRT but it might suggest refinements. It might suggest the hypothesised mechanism does not generate the expected outcomes in the context in question. This might inform gradual refinement of the MRT based on this and subsequent studies, so that it included specific contextual contingencies, in line with the suggestions of both realists and Popperians.[2][4]. To those wanting to know what interventions are effective to address a particular outcome in a particular population and setting, the evaluation null result might suggest the choice of MRT to inform the intervention theory of change was wrong. But, for those wanting to examine the validity of the MRT, the null result might offer useful evidence to inform MRT refinement.

An evaluation is more likely to provide evidence useful for refining MRTs if the choice of MRT is apt. The MRT should not be retro-fitted onto an existing intervention.[13] It should have supportive empirical evidence, ideally from populations and settings similar to those the intervention will target. It should describe mechanisms which could plausibly be triggered by intervention activities. And the theorised mechanisms should also align with influences on the health of the target population. In a systematic review of e-health interventions for men who have sex with men to reduce sexual risk and substance use and improve mental health,[14] we identified a number of interventions informed by MRTs. For example, the 'Hot and Safe M4M' intervention drew on the information-motivation-behavioural skills (IMB) model.[15] The IMB model proposes that behaviour change can come about via development of information, skills and motivation related to the behaviour in question.[16] The IMB model had some supporting evidence though not specifically for addressing the population and outcomes in question.[17] The online intervention activities might plausibly develop information and skills and perhaps motivation. But evidence suggests that broader social and early-life developmental factors are more important influences on substance use risk than deficits in knowledge or skills.[18] So, the IMB model may not have been the best choice for informing this intervention.

Evaluations testing MRT might be useful in assessments of intervention transferability. There is increasing consensus that evaluations should not examine the effectiveness of an intervention in situ but should provide empirically-informed insights into whether effects are likely to transfer to other contexts (settings or populations).[19] One means of doing so is for evaluations to inform testing and refinement of intervention theories of change so that these suggest in which settings and populations and for which outcomes the intervention mechanism might be expected to generate effects.[19][20] But such an approach only offers predictions about the precise intervention that has been evaluated. Another approach would be for evaluations to inform refinement of MRTs. This could then be used to make predictions about the likely effects and transferability of other interventions aiming to trigger the same broad mechanisms.[21] This would be particularly powerful where such refinements to MRT draw on multiple evaluations in different contexts.

Testing of MRT in public-health interventions: evidence from systematic reviews

We contend, however, that the way evaluations currently proceed in the field of public health means evaluations are rarely configured to allow testing of MRT. The first problem is that few studies test hypotheses linked to a single MRT. As a result, evaluations do not provide a very pure means of testing MRT. Many intervention studies focus on multi-component, multi-mechanism interventions with a theory of change influenced by multiple MRTs. In one sense, of course, this is positive, in that these interventions are more likely to be effective and to transfer successfully across a broad range of contexts.[12] Furthermore, such studies might also be useful in exploring how different intervention mechanisms interact with one another.[22]

Such pragmatic studies should therefore be useful in telling us about the effectiveness of scalable interventions. However, with such studies of interventions informed by multiple MRTs, it is often not clear what hypotheses are being tested beyond the very specific hypothesis that a specific intervention works for a specific population, setting and outcome. If such an intervention is found to be ineffective, it is not always clear what this evidence implies about the different MRTs informing the intervention.

In the systematic review of e-health interventions referred to above,[14] the theories of change for some interventions drew on multiple MRTs.[14] For example, as well as drawing on the IMB model, the 'Hot and Safe M4M' intervention drew on motivational interviewing theory.[15] Motivational interviewing theory proposes that behaviour change is promoted by acknowledging that individuals will be at different stages of motivation for behaviour change and interventions needs to help the individual resolve their uncertainties to build motivation.[23] Similarly, the 'TXT-Auto' intervention was informed by social cognitive theory and the health belief model.[24] Both of these interventions were found to be ineffective in reducing sexual risk behaviours.[25] But it is not clear what the trials are telling us about the validity of the various MRTs informing them. For example, the 'Hot and Safe M4M' intervention might have been ineffective because, for the population and outcome in question, the key influences on behaviour change were not information, motivation and/or behavioural skills. Alternatively, the intervention might have been ineffective because, for this population and outcome, behaviour change was actually hindered by consideration of men's stages of motivation or by the provision of support to resolve uncertainties.

The second problem in the field of public-health evaluations is that some MRTs used to inform intervention theories of change do not describe their constructs with sufficient clarity to be testable. For example, we undertook a systematic review of school-based interventions to prevent dating and relationship violence and gender violence.[26] The synthesis of intervention theories of change in this review was informed by an MRT called the theory of human functioning and school organisation.[27] This proposes that schools can promote the health of students by increasing their commitment to school and that one way that schools can achieve this is by eroding 'boundaries', for example between students and their teachers and between the school and its community.

However, in the course of the synthesis, we realised that it was not clear which boundaries should be eroded to promote student commitment and health and which should be maintained. Some studies suggested that some boundaries contribute to reducing dating and relationship violence. For example, some boundaries between schools and communities appeared to help prevent violence at school when schools were in neighbourhoods with social norms supportive of gender-based violence but school rules and norms strongly opposed to these. We came to see that our MRT was insufficiently clear as to what constituted an 'unhealthy' boundary and so it was difficult to test it in the face of the empirical research.

Another problem is that many theories are difficult to test because it is not clear exactly what relationships between constructs the theory is postulating. For example, it is not clear whether the IMB model described earlier is proposing that for an individual to change their behaviour, they need to increase their level of information, their skills *and* their motivation or only *one* of these. Or the model may be proposing that increasing any one of those three will change behaviour *as long as* the other two have reached a certain threshold. It is also not clear whether the theory is suggesting that, to achieve significant changes in overall behaviour, an intervention needs to increase *every individual's* information, skills *and* motivation or whether population changes in behaviour can be achieved by increasing some individuals' information, some individuals' skills and some individuals' motivation (because different attributes will be limiting factors for behaviour change for different

individuals). This may seem like needless pedantry but these questions are important if we are to design studies to test such a theory and then interpret their results.

A final problem is that some MRTs include so many constructs theorised to be influencing an outcome as to render this difficult to test. For example, we found in a systematic review of interventions aiming to modify school environments to reduce substance use and violence that the theories of change for some interventions drew on MRTs proposing that human health or development is influenced by multiple mechanisms operating at different levels.[28] The 'Cyber Friendly Schools' intervention, for example, was informed by ecological systems theory.[29] This MRT suggests that human development is influenced by mechanisms concerning the: microsystem (institutions and groups with which the child interacts); mesosystems (interconnections between microsystems); exosystem (links between social settings not directly involving the child with those that do); macrosystem (the overarching culture); and chronosystem (broader environmental events and transitions, and changing socio-historical circumstances).[30] The 'Cyber Friendly Schools' intervention aimed to transform school and family (as microsystem factors) to reduce violence perpetration and victimisation by increasing family awareness of cyberbullying and modifying school organisational structures to reduce cyberbullying. A trial of the intervention reported no effects on violence perpetration or victimisation.[29] Such evidence is not useful in testing the MRT when the empirical study cannot examine the breadth of factors required.

The more explanatory constructs a theory includes, the more difficult it will be to test its validity in predicting an outcome. In the case of the IMB model, i is possible to imagine a study examining whether increases in information, skills and motivation are associated with changes in behaviour. This might be done within a population-level study with an intervention versus control design where the intervention aimed to increase information, skills and motivation and the evaluation assessed changes in these as well as changes in behaviour. Even better (in theory) would be a factorial design in which there were different intervention arms aiming to increase various combinations of information, skills and motivation. However, such a study would be unwieldy and a factorial design would get even more challenging with theories with more explanatory constructs. We conclude that, although broad models such as ecological systems theory might be useful in summarising our knowledge about the overall range of causes of health, they are not readily open to empirical testing.

Recommendations for testing MRT

Our first suggestion is that to be of wider scientific value in refining MRTs, there could be a limited investment in 'proof-of-principle' studies. These would focus on interventions theorised from single MRT and involving a single mechanism. For example, a proof-of-principle study might test whether men who have sex with men participating in an online intervention to promote motivation to reduce substance use and sexual risk report increased motivation and reduced substance use and/or sexual risk. The study could occur across different settings and populations. As such studies are focused on assessing proof of principle rather than determining the scaleability of pragmatic interventions, they may not need to focus on such distal outcomes. It might be sufficient to determine that a mechanism generates intermediate outcomes, such as improvements in motivation, rather than in the health outcome in question.

Such proof-of-principle evaluations would sometimes, but not always, be done in preparation for pragmatic evaluations of scalable interventions so that we are clearer about how specific intervention mechanisms generate outcomes before putting them together into a pragmatic, potentially-scalable intervention. There would be continuing value in pragmatic evaluations focused on scalable 'multi-mechanism' interventions both in order to understand how multiple mechanisms

interact with one another and in order to identify which interventions to scale up. Such a complementary programme of proof-of-principle and pragmatic evaluations would help us determine whether impacts can be achieved by parsimonious, single-theory interventions or only arise through the action of multiple mechanisms in complex, multi-theory interventions.

Even where such proof-of-principle studies are not run, it might be possible for some pragmatic evaluations to provide some insights in the validity of MRT. This could, and indeed does, occur, for example, by evaluations examining mediation pathways which are specific to one particular MRT. For example, Gardner et al undertook a trial of a parenting intervention; mediation analysis suggested that intervention mechanisms which generated improvements in parenting worked via increasing parent knowledge and skills rather social support.[31]

Our second suggestion is that those developing intervention theories of change should draw on MRTs whose constructs are defined clearly enough that it would be possible to determine in empirical intervention studies whether the theorised mechanisms have been triggered. Some MRTs, such as the IMB model cited above, clearly define their constructs and how these relate to one another. This might be commoner, for example, within psychology and economics than sociology partly because in the latter discipline many theories aim not to predict causal mechanisms but to define the rules of, or simply describe, social institutions or interactions.[32 33] It would be useful for public-health intervention if there were more MRTs describing sociological mechanisms and there have been some attempts to systematise sociological theories to achieve this.[34] Where MRTs include constructs that are currently too vague, these should, if possible, be refined to make them clearer. Theories could also be refined so that they express how their constructs relate to one another more precisely. They should state whether the causal links between constructs apply at the individual and/or population level. They might sometimes use Boolean logical operators to define what combinations of contexts and mechanisms might be expected to generate outcomes.

Our final suggestion is that theories of change should generally be informed by MRTs that are not so broad in the number of different mechanisms described as to render them empirically untestable. This could help move the field forward by providing empirical evidence from experimental and quasi-experimental evaluation to test and refine MRTs. In turn, this would improve the validity of the stock of MRT we possess, which should prove more useful in making informed predictions about what sorts of intervention approaches might prove effective in generating which outcomes for which populations in which settings.

Conclusion

Evaluations could usefully contribute towards testing MRTs. Some evaluation studies could be designed as proof-of-principle assessments oriented towards testing and refining MRT. These would be a complement rather than an alternative to pragmatic studies of scalable complex interventions. Such studies should draw on MRT that is sufficiently clear and parsimonious to allow testing. These studies could facilitate more rigorous testing of MRT and hence refinement of scientific knowledge, which could then inform the development of future interventions. Without a greater focus on using evaluation studies to test and refine MRT, it is not obvious how these can be empirically assessed other than via laboratory experiments conducted under controlled conditions bearing little resemblance to everyday contexts.

References

1. Popper K. The Logic of Scientific Discovery. London: Routledge & Kegan Paul 1959.
2. Bhaskar R. A Realist Theory of Science. Leeds: Leeds Books 1975.
3. Pawson R, Tilley N. Realistic evaluation: Sage 1997.

4. Popper K. The Poverty of Historicism. London: Routledge & Kegan Paul 1957.

5. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj* 2008;337:a1655.

6. Moore G, Audrey S, Barker M, et al. Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance. London: MRC Population Health Science Research Network 2014.

7. Weiss C. Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In: Connell JP, ed. New Approaches to Evaluating Community Initiatives Concepts, Methods, and Contexts Roundtable on Comperhensive Community Initiatives for Children and Families. Washington DC: Aspen Institute 1995.

8. May C. Towards a general theory of implementation. *Implementation Science* 2013;8:18.

9. Weiss CH. How can theory-based evaluation make greater headway? *Evaluation Review* 1997;21:501–24.

10. Lemire S, Kwako AJ, Nielsen SB, et al. What is this thing called a mechanism? Findings from a review of published realist evaluations. *New Directions for Evaluation* 2020;167:73-86.

11. Merton RK. Social Theory and Social Structure. New York: Free Press 1968.

12. Campbell R, Bonell C. Development and evaluation of complex interventions in public health. In: Detels R, Beaglehole MA, Lansang M, et al., eds. Oxford Textbook of Public Health, 7th edition. Oxford: Oxford University Press 2014.

13. van Urk F, Grant S, Bonell C. Involving stakeholders in programme theory specification: discussion of a systematic, consensus-based approach. *Prevention Science, in preparation* 2016;12(14):541-57.

14. Meiksin R, Melendez-Torres GJ, Miners A, et al. E-health interventions targeting HIV/STIs, sexual risk, substance use and mental ill-health among men who have sex with men: systematic review. *Public Health Research* 2022;10(4)

15. Carpenter K, Stoner S, Mikko A, et al. Efficacy of a Web-Based Intervention to Reduce Sexual Risk in Men Who Have Sex with Men. *AIDS Behav* 2010;14(3):9. doi: 10.1007/s10461-009-9578-2 [published Online First: 2009]

16. Fisher J, Fisher W, Misovich S, et al. Changing AIDS risk behavior: Effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioral skills in a college student population. *Health Psychology Review* 1996;15:114–23.

17. Fisher WA, Fisher JD, Harman J. The Information-Motivation-Behavioral Skills Model: A General Social Psychological Approach to Understanding and Promoting Health Behavior. In: Suls J, Wallston KA, eds. Social Psychological Foundations of Health and Illness. Hoboken NJ: Blackwell 2003:82-106.

18. Stall R, Greenwood PG, Pollack LM, et al. Alcohol use, drug use and alcohol-related problems among men who have sex with men: the Urban Men's Health Study. *Addiction* 2001;96(11):1589-601.

19. Burchett HED, Kneale D, Blanchard L, et al. When assessing generalisability, focusing on differences in population or setting alone is insufficient. *Trials* 2020;21(1):286. doi: 10.1186/s13063-020-4178-6

20. Bonell C, Prost A, Melendez-Torres GJ, et al. Will it work here? A realist approach to local decisions about implementing interventions evaluated as effective elsewhere. *Journal of Epidemiology and Community Health* 2021;75(1):46-50.

21. Davey C, Hassan S, Cartwright N, et al. Designing evaluations to provide evidence to inform action in new settings. London: Department for International Development 2019.

22. Bonell C, Jamal F, Melendez-Torres GJ, et al. "Dark logic" – theorising the harmful consequences of public health interventions. *Journal of Epidemiology and Community Health* 2015;69(1):95-8.

23. Miller WR, Rollnick S. Motivational Interviewing: Preparing People to Change. New York NY: Guilford Press 2002.
24. Reback C, Fletcher J, Swendeman D, et al. Theory-Based Text-Messaging to Reduce Methamphetamine Use and HIV Sexual Risk Behaviors Among Men Who Have Sex with Men: Automated Unidirectional Delivery Outperforms Bidirectional Peer Interactive Delivery. *AIDS and Behavior* 2019;23(1):11. doi: 10.1007/s10461-018-2225-z [published Online First: 2018]
25. Melendez-Torres GJ, Meiksin R, Witzel TC, et al. eHealth interventions to address HIV and other Sexually transmitted infections, sexual risk behavior, substance use, and mental ill-health in men who have sex with men: systematic review and meta-analysis. *JMIR Public Health Surveillance* 2022;8(4):e27061.
26. Melendez-Torres GJ, Orr N, Farmer C, et al. School-based interventions To Prevent Dating and Relationship Violence and Gender-Based Violence (STOP-DRV-GBV): systematic review to understand characteristics, mechanisms, implementation and effectiveness. *Public Health Research (in press)* 2022
27. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Social Science & Medicine* 2003;56(6):1209-20.
28. Ponsford R, Falconer J, Melendez-Torres GJ, et al. Whole-school interventions promoting student commitment to school to prevent substance use and violence: synthesis of theories of change. *Health Education (Online first)* https://doiorg/101177/00178969221100892 2022
29. Cross D, Barnes A, Cardoso P, et al. Cyber-Friendly Schools. Final report to Healthway. Reducing Cyberbullying in Schools: International Evidence-Based Best Practices2018:95-108.
30. Bronfenbrenner U. The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press 1979.
31. Gardner F, Burton J, Klimes I. Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry* 2006;47(11):1123-32.
32. Boudon R. What middle-range theories are. *Contemporary Sociology* 1991;20(4):519-22.
33. Giddens A. Sociology: A Brief but Critical Introduction (2nd Edition). Basingstoke: Palgrave 1986.
34. Turner JH. Must sociological theory and sociological practice be so far apart? *Sociological Perspectives* 1998;41(2):243–58.