Theory driven inquiry for health systems research

Wanted: better research and evaluation approaches to deal with complexity in health systems research

Since a few years, complexity has emerged as a key issue in health systems research. It is now generally acknowledged that health interventions often take place in settings or health care systems that are complex and that this requires adapted research and evaluation methods.

At the same time, dissatisfaction with ‘traditional’ quasi-experimental designs has emerged, because they focus on testing effectiveness but overlook the underlying processes of change and the context conditions that are needed for success.

Under the label ‘theory-driven inquiry’, we present theory-driven evaluation, theories of change and realist evaluation, 3 approaches that share a strong focus on the theory on which an intervention is based. They all aim at explaining how an intervention works and in which conditions. These approaches have their foundations in realism and make extensive use of theory with the aim of opening the black box and analysing how interventions work, in addition to whether they work. Theory-driven inquiry thus seems promising for evaluation and research of complex issues in health.

The website and the document were developed and written by Bruno Marchal, Sara Van Belle and Guy Kegels with inputs from Josefien Van Olmen and Tom Hoerée (Health Care Management Unit of the Department of Public Health, Institute of Tropical Medicine, Antwerp) and in collaboration with the Development Policy & Practice Unit of the Royal Tropical Institute, Amsterdam.

ITM and KIT collaborate on the development and implementation of theory-driven inquiry in health systems research, and aim at stimulating a better understanding of the advantages of this approach among both researchers and policymakers. To this end, we are engaged in a number of research and evaluation studies that use TD inquiry principles.
Weakenes of commonly used research methods

Most current research and evaluation methods in health systems research do not sufficiently take into account complexity.

- Quasi-experimental methods focus on outcomes and impact
- They do not take enough into account the influence of the context in which the intervention took place
- They rarely expose the mechanisms that make the intervention work
- Case studies, action research and process evaluations provide context-specific information, but their findings are difficult to generalise

Such studies are not very useful for policymakers. Research should indicate whether interventions work (or not), but they ideally also provide information on the conditions that are required and the mechanisms that underlie success. This is what policymakers and planners need in order to design the appropriate intervention for their context. Theory-driven inquiry seeks to provide this information.

Theory-driven inquiry

The domain of theory-based evaluation has seen quite some evolution since the 1970s. Under the label Theory-driven inquiry, we present three main evaluation approaches that share a strong focus on the theory on which an intervention is based. They all aim at explaining how an intervention works and in which conditions.

Theory-driven evaluation

Theory-driven evaluation emerged in the discipline of evaluation in the 1980s. Chen and Rossi developed it as an answer to policy and programme evaluation approaches that remained limited to before-after and input-output designs, or that focused narrowly on methodological issues.

Realist evaluation and synthesis

Realist evaluation intends to answer the question: “Why does this intervention work, for whom and in what circumstances?” In other words: which mechanism(s) cause which outcome(s) in which conditions? Realist inquiry has an explanatory focus and aims to unravel mechanisms of change. It is applied in research and evaluation, and in synthesis of evidence.

Theories of change

The Theory of Change approach (ToC) was developed by the Roundtable on Community Change (Aspen Institute, USA).

Pragmatic in its approach, ToC is used to evaluate community-based programmes that typically involve many agencies and actors, have several levels and strands of activities, objectives and strategies that shift in time, and outcomes that are difficult to measure.

Nothing as practical as a good theory (Weiss 1995)
Defining ‘complexity’

“Complexity” is a term understood in different ways. It has a long history in fields like systems theory, artificial intelligence and mathematics. It has been taken up much more slowly in medicine, public health and health systems research. In these domains, there is still some important conceptual confusion around the definition of ‘complexity’.

Complexity in health care
In health systems, complexity can be applied to (1) organisations, (2) problems and (3) policies, management interventions and programmes.

Health care organisations are complicated systems
• They are staffed by cadres having their specific competences, cultures and shared mental models
• They are made up by different units and levels
• Staff and units relate and interact to each other through formal and informal channels
• Each unit is dealing with different tasks and specialised functions, but all units need to be working towards the shared goal of providing care to patients.

Health care organisations are also complex systems
It is the actors’ “agency” at work in social relationships and feedback mechanisms that make health care organisations complex.
• Interactions lead to new unpredictable actions of the actors (emergence).
• But these emergent behaviours are somehow restricted by the history (path dependency) and the “structure” (formal and informal rules) of the organisation.

Problems or situations can be complex
Problems or situations are complex if they are determined by multiple, interacting factors. The people involved, their interactions and the specific context they live and work in shape webs of connections and relations. The interactions may involve positive or negative feedback loops. Understanding such complex situations requires the disentangling of intervention, actors, mechanisms of change, outcome and context.

Some policies, management interventions and programmes are complex
Health policies, management interventions and programmes are complex if
• they consist of various parts that interact with each other in non-linear, unpredictable ways
• the underlying mechanism of change has a complex nature or is unknown

Further reading
Theory driven inquiry in the field of evaluation and research

1. Theory-driven evaluation
During the 1980s, theory-driven evaluation emerged in the discipline of evaluation. Chen & Rossi developed it as an answer to policy and programme evaluation approaches that remained limited to before-after and input-output designs or that focused narrowly on methodological issues (method-driven evaluation).

Theory-driven evaluation aims at assessing not only the effectiveness of an intervention but also its causal mechanisms, taking into account the context of the intervention.

The programme theory
The central element of this approach is the programme theory.

- For any intervention, a programme theory can be described.
- This theory explains how the planners expect the intervention to reach its objective. Describing this often implicit set of assumptions allows to understand what is being implemented and why (not).
- The programme theory represents a hypothesis that can be tested and further developed.

Theory should in this case be understood as middle-range theories or theories of the middle range as defined by R.K. Merton. These are “theories that lie between the minor but necessary working hypotheses (…) and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organisation and social change”.

The 2 parts of a programme theory
For Chen & Rossi, the programme theory consists of 2 parts: the normative theory and the causative theory. Later, Chen called this respectively the action model and the causal model.

The normative theory (the action model)
The normative theory presents the theoretical background that informs the design and implementation of the intervention and spells out the objectives and the implementation procedure.

- It describes how the actual intervention was different or not from the planned intervention and the way the intervention was actually implemented.
- It also describes the actual outcomes compared to the intended outcomes, and seeks not only to document positive but also negative outcomes.

- Assessment of the normative theory can help decision makers, since it evaluates not only the effectiveness, but also the consistency of the implementation of the programme. This allows distinguishing programme theory-failure from implementation failure and provides feedback for improving the intervention.
The causative theory (the causal model)
The causal theory specifies the underlying causal mechanisms in terms of relationships between intervention and outcome, influence of context and intervening factors.

Unearthing the causal theory opens the box between the intervention and its outcomes. It allows judging on the value of the intervention and assessing whether the assumed causal processes actually did take place. It allows unpacking the intervention into its components and to check which element was most important and which intervening factors are most critical.

2. Theories of change
The Theories of Change (ToC) approach was developed by the Roundtable on Community Change (the Aspen Institute, USA). More pragmatic in its approach, ToC was developed to evaluate community-based programmes that typically involve many agencies and actors, have several levels and strands of activities, objectives and strategies that shift in time, and outcomes that are difficult to measure.

It is similar to theory-driven evaluation in that it seeks to establish the links between interventions, contexts and outcomes.

This is done by developing and testing ‘logic models’ that describe the populations that are targeted, the indicators used to monitor change, the thresholds of indicators that indicate significant change and the time lines. They specify short-, medium- and long-term outcomes to map how the intervention leads to the expected outcome and to enable attributing change to (parts of) the intervention.

These models are developed in collaboration with the stakeholders, ideally during the planning phase of a project during so-called ‘surfacing exercises’.

Further reading
3. Realist evaluation

Realist evaluation can be considered as an approach within theory-driven evaluation although it has stronger philosophical underpinnings. It proposes conceptual tools to apply the principles of theory-based evaluation. Realist inquiry intends to answer the question: “What is it about this programme that makes it work, for whom and in what circumstances?”, or in other words: which mechanisms cause which outcome under which conditions?

Realist inquiry has an explanatory focus and aims in essence at unravelling mechanisms of change. Based on critical realism, it considers that interventions work (or not) because actors take up what is offered by the intervention (or not). The interaction with the actors and with specific context elements triggers mechanisms, which cause certain outcomes to occur.

In this approach, the theory that provides the starting point is called the “Middle Range Theory”.

The CMO configuration

In realist evaluation, the “Context-Mechanism-Outcome” configuration describes how outcome patterns emerge from the interaction between intervention, context and mechanisms.

- The context is made up by the circumstances within which public health interventions are implemented. It includes the stakeholders, their interests and convictions regarding change and the process of implementation, and the organisational, socio-economic, cultural and political conditions.

- Mechanisms are the drivers of the reactions of the target group, which lead to change and which are triggered by the intervention within a certain context.

Realist evaluation in practice

Evidence building using a realist perspective implies searching for CMO-configurations:

1. At the start, the middle range theory that underlies the intervention is made explicit: these are the assumptions held by the people involved regarding the expected outcome of an intervention and how - through which mechanisms or reactions - this will be achieved. This is complemented by existing (published) knowledge and experience.

2. Mixed methods are used to collect data about the elements of this initial middle range theory (intervention, outcomes, mechanisms, context).

3. During data analysis, context-mechanism-outcome configurations are identified and tested for plausibility.

4. The initial middle range theory is adapted in the light of the findings, which then becomes the starting point of new studies.

Realist synthesis

Pawson applied the principles of realist evaluation to the synthesis of evidence. Realist synthesis aims at informing policymakers through detailed ‘lessons learned’ by providing information on effectiveness but also on the mechanisms that cause the effect and the context conditions that are required to make the intervention work. This allows policymakers to better decide whether interventions can be expected to have the same results in their own setting.

This approach assesses findings of research and evaluations for evidence on the interactions between context, mechanisms and outcome. It aims at providing plausible explanations as to how interventions have produced their results, in which conditions and for which groups of the population.
Recommended reading

Advantages of theory-driven inquiry
Adherents of theory-driven inquiry argue that it has several advantages:

- It is well suited to investigate change in complex social systems: its focus on the generative causality that underlies interventions offers an approach to unravel the root causes and mechanisms of change that take into account the interaction between agency and structure and emergence of behaviours, social action and change.

- Evaluations that build on the programme theory allow the evaluator to identify the intermediate steps of the (hypothetical) causal chain. This can help in sharpening the focus of the evaluation.

- If the programme theory is made explicit together with the main actors, it can lead to a better, shared understanding of the intervention. This in turn could improve the ownership and lead to more context-adapted interventions.

- Theory-building helps to overcome the low external validity and low power to explain change of traditional case studies and process evaluations: it indicates in which conditions and how the effects were obtained. As such it improves transferability of findings to other settings.

Further reading
Realist evaluation and critical realism

Many authors refer to critical realism when discussing the philosophical roots of realist evaluation. Understanding the core elements of realism may help in applying the principles of theory-driven inquiry, and especially clarify how causality, mechanisms and context can be examined.

Critical realism
Critical realism is a particular philosophy of science developed by Ray Bhaskar, Rom Harré, Andrew Sayer and others. Adherents of realism maintain that social phenomena are real world objects and that these are not contingent on human observation: they do not only exist in people’s minds and are therefore independent of the researcher’s claims about them. In other words, social objects and the conjunctions between them exist and therefore can be described.

However, at the same time, knowledge is socially and historically constructed. Realists maintain that constructs that explain individual and social life are to be derived from observable events.

Realism bases this explanation on the perspective of ‘generative causality’. It explains change brought about by programmes by referring to the actors who act and change (or not) a situation under specific conditions and under influence of external events (including the intervention). The actors and the interventions are embedded in a stratified social reality, which is the result of an interplay between individuals and institutions, each with their own interests and objectives.

Key elements
Of main interest to us are the views of CR on stratified nature of the real world, the generative nature of causality, the interaction between social structure and agency, and retroduction, CR’s approach of learning. These principles have informed realist evaluation. For Pawson and Tilley, the exploration of the underlying generative mechanisms should reflect the embeddedness of the intervention in the social reality, the micro- and macro-level processes of change and how actors’ choices and use of their resources lead to the outcome. The CMO configuration offers a practical method to explore and analyse these generative mechanisms.

- Critical realism stratifies reality in the ‘empirical’ (the day-to-day experience), the ‘actual’ (observable patterns) and the ‘real’ (the relatively unchanging existence, which can be disclosed by science). The ‘real’ encompasses the ‘actual’, which in turn includes the ‘empirical’. What can be observed is only part of what actually is, which in turn is only the actualised part of what is real.

- Recognising the social world as a dynamic open system and the importance of social action, CR aims at identifying the generative mechanisms that underlie the social reality. Causality is generative: actors and society have potential mechanisms of causation by their very nature, which are actualised when an event or intervention combined with the right context factors trigger these generative mechanisms. These generative mechanisms are not directly observable but they are real.

- CR considers that reality comes about as a result of an interaction between agency and structure (the transformational model of social activity). Generative mechanisms are actualised through the interplay between actors and structures. Such mechanisms can be positive and facilitate change or counteracting and block change.

- Retroduction is the technique through which generative mechanisms can be identified: through iterative abstraction, an argument moves from a description of some phenomenon to a description of something that produces it or is a condition for it.
Further reading
Theory-driven inquiry in health care

While the merits of theory-driven and realist evaluation have been widely discussed in journals on evaluation, there is relatively little documented experience with theory-driven inquiry in the domain of health services and health systems. As an illustration, some interesting papers are briefly discussed below.

Theory-driven evaluation

– Birckmayer and Weiss (2000) present a review of 6 health promotion studies based on theory driven evaluation.

– Mercier et al. (2000) carried out a theory-driven evaluation of drop-in activities at a YMCA Youth Center and present a clear methodology to develop and test the programme theory.

– Grocott et al. (2002) is an example of how TDE was applied to patient care, in this case the palliative management of malignant wounds.

– Rodriguez, E. & Mead, J. (1997) describe a retrospective theory driven program evaluation to assess the complex processes and dynamics underlying the program and which influence its outcomes. They show how TDE examines the black box of processes and mechanisms and provides insights in what went wrong.

– Van Belle and colleagues present a 7 step method they designed to carry out a theory-driven evaluation of an adolescent reproductive health project in West-Africa.

Theories of change


– MacKenzie and Blamey (2005) discuss the assessment of the Scottish Health Demonstration projects in child health and heart disease prevention.

– Douglas et al. (2010) develop a Theory of Change protocol for the evaluation of smoking cessation programmes for pregnant women and young people.
Realist evaluation

- Evans and Kiloran (2000) present the evaluation of a Health Education Authority’s Integrated Purchasing Programme and specifically of demonstration projects testing partnerships to tackle health inequalities.

- Byng et al. (2005) and Byng et al. (2008) applied realist evaluation principles to evaluate a mental health programme in the NHS.

- Leone (2008) presents an overview of how a MRT for an intervention can be derived from an analysis of the programme, how this MRT is translated into research questions and how analysis is done.

- Guichard and Ridde (2009) focus on the mechanisms that underlie programmes to fight social inequalities in France.

- Ogrinc and Batalden (2009) develop a practical evaluation framework to assess teaching programmes on improvement of care on the basis of realist evaluation principles.

- Greenhalgh et al. (2009) apply realist evaluation in the assessment of the Modernisation Initiative, which aimed at whole-scale transformation in stroke, kidney and sexual health services in London.

- Pommier et al. (2010) present a research protocol for evaluation of health promotion in schools.

There are fewer publications of TD inquiry in health systems research in low and middle-income countries:


- Marchal et al. (2010) present a realist case study of human resource management in Ghana and discuss the main methodological challenges of RE.

- Blaise and colleagues (2010) explain how realist evaluation compares with other methods used in health promotion research. This is one of the few texts in French on realist evaluation.

- Maluka and colleagues (2010) carried out a realist evaluation of the introduction of the accountability for reasonableness framework at a rural district in Tanzania.

- Marchal et al. (2010) analyse how a hospital management team succeeded in reviving an ailing district hospital in Ghana using the realist evaluation approach.
Realist synthesis

- Connelly et al. (2007) present a realist synthesis of controlled trials of childhood obesity prevention programmes.

- Greenhalgh et al. (2007) enhance the analysis of findings of a Cochrane review by applying a realist analysis in order to “determine the aspects that determine success and failure in various situations.”

- Dieleman et al. (2009) present a realist synthesis of evidence on management of the health workforce in low- and middle-income countries.

- Wong et al. (2010) carried out a realist synthesis of internet-based medical education, presenting a clear methodology section.

- Kane et al. (2010) reviewed human resources management interventions aimed at improving community health workers’ performance in IMCI in Low and Middle Income countries.

Bridging realist synthesis and evaluation

- Rycroft-Malone et al. (2010) present the development of a protocol for evaluation of protocol-based care in nursing that combines elements of realist synthesis with realist evaluation.

- Robert et al. (2010) focus on user fees abolition policies in Africa and present a protocol that similarly combines realist synthesis with realist evaluation.
Bibliography


PAWSON, R. (2001) Evidence based Policy: II. The promise of ‘realist synthesis’. ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London.


