Evaluation

Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research

Bruno Marchal, Sara van Belle, Josefien van Olmen, Tom Hoerée and Guy Kegels

Additional services and information for *Evaluation* can be found at:

Email Alerts: http://evi.sagepub.com/cgi/alerts

Subscriptions: http://evi.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://evi.sagepub.com/content/18/2/192.refs.html

>> Version of Record - Apr 22, 2012

What is This?

Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research

Bruno Marchal
ITM-A, Belgium

Sara van Belle
BTC, Belgium

Josefien van Olmen
ITM-A, Belgium

Tom Hoerée
ITM-A, Belgium

Guy Kegels
ITM-A, Belgium

Abstract

This overview aims to stimulate conceptual and practical discussions to help unlock the full potential of realist evaluation in health systems research. Based on a structured literature search, this review maps how the concepts of realist evaluation are applied in health systems research and which methodological problems are encountered. We found a great diversity in the depth of application of the philosophical concepts, use of terminology and scope of application in the research process. Terms of theory-driven evaluation, theories of change and realist evaluation are often used interchangeably. Diverging views exist regarding the nature of 'mechanism' and the difference between mechanism and essential context condition. A lack of methodological guidance was highlighted by many authors. Realist evaluation is slowly gaining traction in health systems research, but more clarity is needed concerning the definitions of mechanisms and context and how the configuration of context, mechanism and intervention can be described and assessed.

Corresponding author:
Bruno Marchal, ITM-A, Nationalestraat 155, B-2000, Antwerp, Belgium.
Email: bmarchal@itg.be

Introduction

For 15 years, health systems researchers have been employing theory-driven-inquiry approaches in response to calls for better methods to deal with complexity. These calls resulted from frustration with (quasi-)experimental research and evaluation designs (Fulop et al., 2001). Quasi-experimental studies are the mainstay for effectiveness studies, using analytical techniques like randomization, linear regression and cluster analysis to isolate the effect of each variable on the outcome. Such designs are excellent to assess effectiveness of interventions, but they fail to provide valid information when applied to complex and dynamic systems – such as health care organizations (Sturmberg and Martin, 2009). By holding all other variables constant 'instead of showing how the variables combine to create outcomes' (Fiss, 2007), the results of quasi-experimental studies do not identify in which conditions and through which configuration of factors the outcome is achieved. This limits the drawing of context-sensitive conclusions, limits learning and contributes to the inadequate uptake of research evidence into practice (Kernick, 2006).

In domain of health systems research, these limits have been identified for some time (see, for instance, Barnes et al., 2003; Berwick, 2008; Eccles et al., 2003; Grimshaw et al., 2001; Victora et al., 2004; Walshe, 2007). Keynote presentations at the First Global Symposium of Health Systems Research organized by WHO, the Alliance HPSR and the Global Forum for Health Research in Montreux (16–19 November 2010) (www.hsr-symposium.org) indicate that the need for innovative research methods for complexity have reached the 'main stream' of the health systems research community.

In social sciences, alternative research approaches for complexity have been developed over the last 30 years. The emergence of theory-driven evaluation during the 1980s can be seen in this light. 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 (Chen and Rossi, 1980, 1983, 1987). They argued that for any intervention, a programme theory can be described that explains how the planners expect the intervention to reach its objective. Describing the often implicit set of assumptions that steers the choice and design of a programme or intervention is useful, because it aims to explain what is being implemented and why. The theory of change approach (Connell et al., 1995; Fulbright-Anderson et al., 1998) and realist evaluation (Pawson and Tilley, 1997) are considered as specific schools within theory-driven evaluation. Some authors prefer to use the term 'theory-driven evaluation' for all these schools (see e.g. Coryn et al., 2011). We prefer to group them as 'theory-driven inquiry' to avoid erasing the key differences, which are briefly presented below.

Theory-driven evaluation

A theory-driven evaluation focuses not only on the implementation of the intervention and its effectiveness, but also on the causal mechanisms and the contextual factors that underlie change (Chen, 1990).

Theory is defined by Chen and Rossi (1983) as the 'prosaic theories that are concerned with how human organizations work and how social problems are generated'. Within the programme theory, Chen (1989) distinguishes normative theory from causal theory.

The *normative* theory presents the theories and ideas that inform the design and implementation of the intervention, its objectives and the implementation procedure. This explains how the intervention is to be implemented and what it aims at. The normative theory can guide the assessment of the effectiveness of the intervention and the consistency of the implementation. This enables the distinction between programme-theory failure from implementation failure. Later, Chen (2005) called this the 'action model'.

The *causal* theory specifies the underlying causal mechanisms in terms of relationships between the intervention and the outcome, the influence of context and the intervening factors. This explains why the intervention is set up, how it is supposed to work and in which conditions. Chen (2005) calls this the 'change model'.

Theory-driven evaluation is increasingly undertaken in healthcare research. For example, the comprehensive review by Coryn et al. (2011) includes empirical research based on theory-driven evaluation in social sciences, education, medicine and healthcare, while other examples were published by Grocott et al. (2002), Mercier et al. (2000), Rodriguez and Mead (1997) and Sidani et al. (2004).

Theory of change

The Theory of Change (TOC) approach[1] was developed by the Roundtable on Community Change of the Aspen Institute (Connell et al., 1995; Fulbright-anderson et al., 1998; Weiss, 1995). More pragmatic in approach and oriented toward stimulating practical change, TOC was initially used to evaluate community-based programmes. These typically involved many actors, intervened at several levels, would consist of sets of activities that shift in time and had outcomes that are difficult to measure (Judge and Bauld, 2001).

TOC is essentially prospective (Dickinson, 2006) and seeks to establish the links between intervention, context and outcome (Barnes et al., 2003; Mason and Barnes, 2007; Weiss, 1995). It does so through development and testing of logic models (Douglas et al., 2010). Such logic models describe the populations that are targeted by the programme, the indicators used to monitor change, the thresholds indicators should pass to indicate significant change and the time lines (Judge and Bauld, 2001). In the USA, TOC has been integrated in programmes from the development phase onwards, while in the UK, it was most often used in evaluations of programmes that had already begun (Mason and Barnes, 2007). Examples of studies that applied TOC in health care include Barnes et al. (2003), MacKenzie and Blamey (2005), Mason and Barnes (2007) and Sullivan et al. (2002). Blamey and Mackenzie (2007) provide a comprehensive comparison of TOC and realist evaluation.

Realist evaluation

Pawson and Tilley (1997) developed the Realistic Evaluation school,[2] arguing that in order to be useful for decision makers, evaluations need to indicate 'what works, how, in which conditions and for whom', rather than to answer the question 'does it work?'. The account of the processes that explain how an intervention leads to a particular outcome is formulated as a middle-range theory (MRT). Pawson and Tilley use the definition of MRT provided by Merton (1968: 39): '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 organization and social change'. While no exact distinctions are provided by the literature, in our understanding, the MRT in realist evaluation is situated at a more abstract level than what is called

the 'operational' programme theory in theory-driven evaluation, or the logic model in TOC. Blamey and Mackenzie (2007) provide a comparison of the main elements of TOC and realist evaluation.

Its explicit philosophical foundations and its methodology set realist evaluation apart from other theory-driven approaches. Pawson and Tilley refer to scientific realism as their philosophical source of inspiration. This school of realism shares a number of elements with critical realism: it accepts that there is a reality independently of the researcher (natural realism), but that knowing this reality through science is unavoidably relative to the researcher (relativist epistemology). Realist evaluators consider causality to be generative in nature. In other words, they believe that actors have a potential for change by their very nature. Accepting as such the role of actors in change (agency), realist evaluation considers structural and institutional features to exist independently of the actors and researchers (Pawson and Tilley, 1997). In their view, both actors and programmes are rooted in a stratified social reality, which results from an interplay between individuals and institutions, each with their own interest and objectives. If all human action is embedded within such a wider range of social processes, then causal mechanisms reside in social relations and context as much as in individuals.

In practice, realist evaluations start with an MRT and end with a refined MRT (Pawson and Tilley, 1997: 84) (see Figure 1). The MRT can be formulated on the basis of existing theory, past experience and previous evaluations or research studies. The result is discussed with the stakeholders and finally results in the MRT that will be tested. The field study is then designed in relation to the MRT: the design, data-collection tools and analysis tools are developed so as to enable testing the elements of the MRT. Realist evaluation is method-neutral and both quantitative and qualitative data are routinely collected. The data-collection phase is followed by data analysis, whereby realist evaluation uses the context–mechanism–outcome (CMO) configuration as the main imaging tool. The analysis of qualitative data from interview transcripts and documents is based on coding in terms of 'description of the actual intervention', 'observed outcomes', 'context conditions' and 'underlying mechanisms'. Quantitative data are analysed with the aim of assessing the effectiveness of the intervention and to substantiate or devalidate the patterns that emerge. The resulting explanations for the observed outcomes are formulated as conjectural CMO configurations. For better comparison, these can be compiled in the form of narrative summaries, tables or diagrams. Through triangulation, the plausible patterns or demi-regularities that explain how the intervention led to the observed results are confirmed. In a final step, these are translated into the more abstract level of the initial MRT, which is modified if necessary. This then kicks off a new study. From the realist perspective, single evaluations cannot produce universally valid findings. What realist evaluation can do is to help the researcher to find out in which specific conditions the intervention works (or not) and how, and to refine the findings in a process of specification. This in turn leads to an accumulation of insights that help decision makers to assess whether interventions that proved successful in one setting may be so (or not) in another setting and how (Pawson and Tilley, 1997). Later, Pawson and colleagues applied the same principles to review and synthesis of evidence for policymakers, calling this realist synthesis (Pawson 2002, 2006; Pawson et al., 2005).

Advantages and challenges

Theory-driven inquiry has a number of advantages. Its different schools provide frameworks to systematically deconstruct an intervention into its components and to reconstruct it with the causal webs that led to the observed outcome. Exposing not only the underlying mechanisms of change, but also the influence of context on the relation between the intervention and the outcome aids the
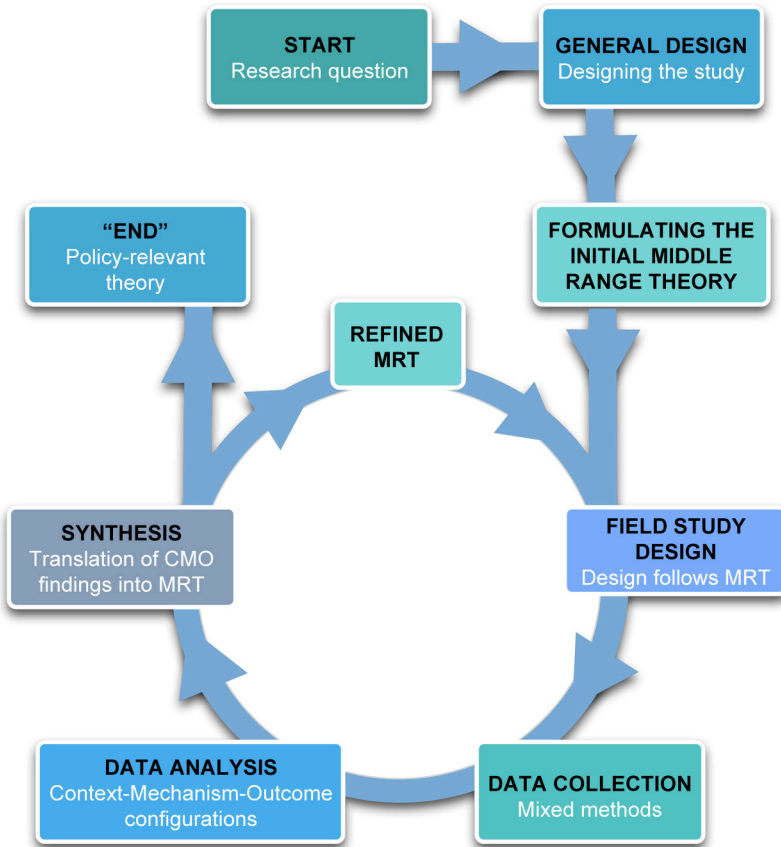
Figure 1. The realist evaluation cycle (adapted from Pawson and Tilley, 1997).

evaluation of complex issues (Berwick, 2008; Greenhalgh et al., 2009). A detailed assessment of the intermediate steps and processes between intervention and outcomes, backed up by complementary quantitative data, can improve the attribution claim (Weiss, 1997). Transferability of findings to other settings is thus enhanced (Kernick and Mannion, 2005) and the findings are more relevant for policy makers (Stame, 2004; Van Belle et al., 2010). Theory-driven inquiry also strengthens the general knowledge base, because it frames findings in existing theories (Weiss, 1995). If applied in a cyclical manner, such theory building helps to overcome the limits of traditional case studies and specifically their low external validity and low power to explain change (Calnan and Ferlie, 2003; Costner, 1989).

Theory-driven inquiry presents a number of challenges. First, in some cases, there may be little or no relevant theory that applies to the problem under consideration (Chen and Rossi, 1989). Guidance to develop the programme theory in such cases is scarce (Cole, 1999). Pawson and Tilley

(1997) propose few practical methods, if any, to do this (Julnes et al., 1998). Most initial instructive examples were developed outside the field of health (see, for instance, Kazi, 2003a, 2003b; Ying Ho, 1999), but these remain scarce. Second, the question of what constitutes a mechanism remains a challenge (Astbury and Leeuw, 2010). Third, there is the issue of context. Barnes et al. (2003), for example, warn about the risk of interpreting context as a purely external factor; they argue that in open systems, context is as much shaped by the actors as it constrains their activities. Fourth, it is not yet clear when theory-driven research is indicated. While some argue that, for instance, realist evaluation is useful for complex issues, Blamey and Mackenzie (2007) believe that it is not well-suited to evaluations of complex multi-site programmes made up of different interventions aiming at multiple outcomes. Finally, theory-driven evaluations can be quite resource- and time-intensive, because in addition to the assessment of the efficacy, the underlying theory that contains the underlying mechanisms and the influence of context also need to be assessed (Blamey and Mackenzie, 2007; Pedersen and Rieper, 2008).

Growing attention for realist evaluation

The last 15 years have witnessed a gradual increase in published papers that apply theory-driven-inquiry principles in health systems research. Weiss (1997) provided a review of the methodological development and application of theory-driven evaluation in the broad sense in 1997. The author found very few empirical applications, and these were all in the fields of health promotion and risk prevention.

Ten years later, Rogers (2007) found in a review with a wider scope that more studies had used theory-driven-inquiry principles. She concluded that programme theory had become a central part of many development programmes in the shape of logical frameworks, but that often the principles of theory-driven research were applied in a simplistic and linear way.

Most recently, Coryn et al. (2011) presented an overview of theory-driven evaluation, which they defined broadly to include realist evaluation. They retained 45 studies from social sciences, education, health and medicine, and described them with a framework built upon core principles of theory-driven evaluation. The authors conclude that there are few published studies and that among them, very few used the programme theory 'in any meaningful way for formulating or prioritizing evaluation questions nor for conceptualizing, designing, conducting, interpreting, or applying the evaluation reported'.

To the best of our knowledge, there has been no review of studies that are based on realist evaluation. This article presents a review of the application of realist evaluation methodology in health systems research. The aim of this literature review was to map how the concepts of realist evaluation are applied in health systems research and to identify the methodological problems encountered in practice. Through this overview, we aim to stimulate conceptual and practical discussions that may help to unlock the full potential of realist approach in health systems research.

Methodology

The questions we aimed at answering through this review include: (1) How are realist evaluation principles being used in health systems research in terms of research topic, scope and justification? (2) How do these studies apply the realist evaluation cycle? and (3) What methodological problems do these papers raise?

Table 1. Inclusion criteria

Inclusion criteria	Details
Studies based on realist evaluation principles	
Studies carried out in the domain of health	Domain defined as including: health care, health policy, health programmes and management of health care organizations
Paper reporting on the design and/or implementation of a realist evaluation study	

We opted for a review of papers identified through a systematic search strategy, which was designed to find peer-reviewed papers that present applications of realist evaluation in health-service and health-systems research. Since there are currently no widely accepted quality criteria for realist evaluation, we did not aim at grading the quality of application of the realist evaluation method nor the quality of the studies we found.

We started the search using the PubMed and Web of Science/Social Sciences Citation Index search engines. We used sets of key words including the approach ('realist evaluation', 'realist synthesis' and 'realist review') and the domain. Given the heterogeneous use of the word 'health systems research', we chose the following terms for the domain: 'health', 'health care', 'health services', 'program*', 'policy', 'program* evaluation'.

Our initial PubMed search results showed that the search strategy was neither specific nor sensitive. It led to a very large number of publications that were not relevant and missed important papers we already knew. A review of the current Mesh vocabulary and definitions showed that it does not include any term specifically pointing to realist evaluation. We therefore attempted to increase the sensitivity of the search in two ways. First, we used the Mesh terms 'evaluation studies as topic' and 'program evaluation'. Both include a wide set of entry terms that make the search more sensitive, although decreasing its specificity. Second, knowing from experience that many authors use terms from theory-driven evaluation, realist evaluation and theories of change interchangeably, we added the following search terms: theory driven evaluation, theory based evaluation, theor* of change. This resulted in 1395 initial hits. These were reviewed on the basis of the inclusion criteria presented in Table 1. We set the following exclusion criteria: (1) conceptual or theoretical papers; (2) advocacy papers; (3) studies that report on application in domains other than health. No time or language limits were set in order to cast the net wide and to explore the evolution of the trends in evaluation. The review was concluded in May 2011

The title and abstracts of 1395 papers identified by the PubMed/SCCI search were scanned independently by two researchers for relevance. Discrepant decisions were reviewed. 229 papers were retained. Figure 2 summarizes the review process.

At this stage of the review, we decided to try and increase the yield by running the same combinations of search terms in Google Scholar and Scirus, which resulted in 30 additional relevant abstracts. In addition, we scanned the bibliographies of relevant papers and books (reference tracking) and publication lists of key authors. We also reviewed the websites of evaluation journals and of evaluation initiatives. This yielded another 12 abstracts (see Figure 2).

Two researchers independently reviewed the resulting long list of 271 abstracts and excluded 172. Most of these papers reported upon theory-based interventions or the application of measurement tools based on theory and cannot be considered as using theory-driven inquiry as defined above. For the 99 papers of the resulting short list, the full text was retrieved. Upon reading, these
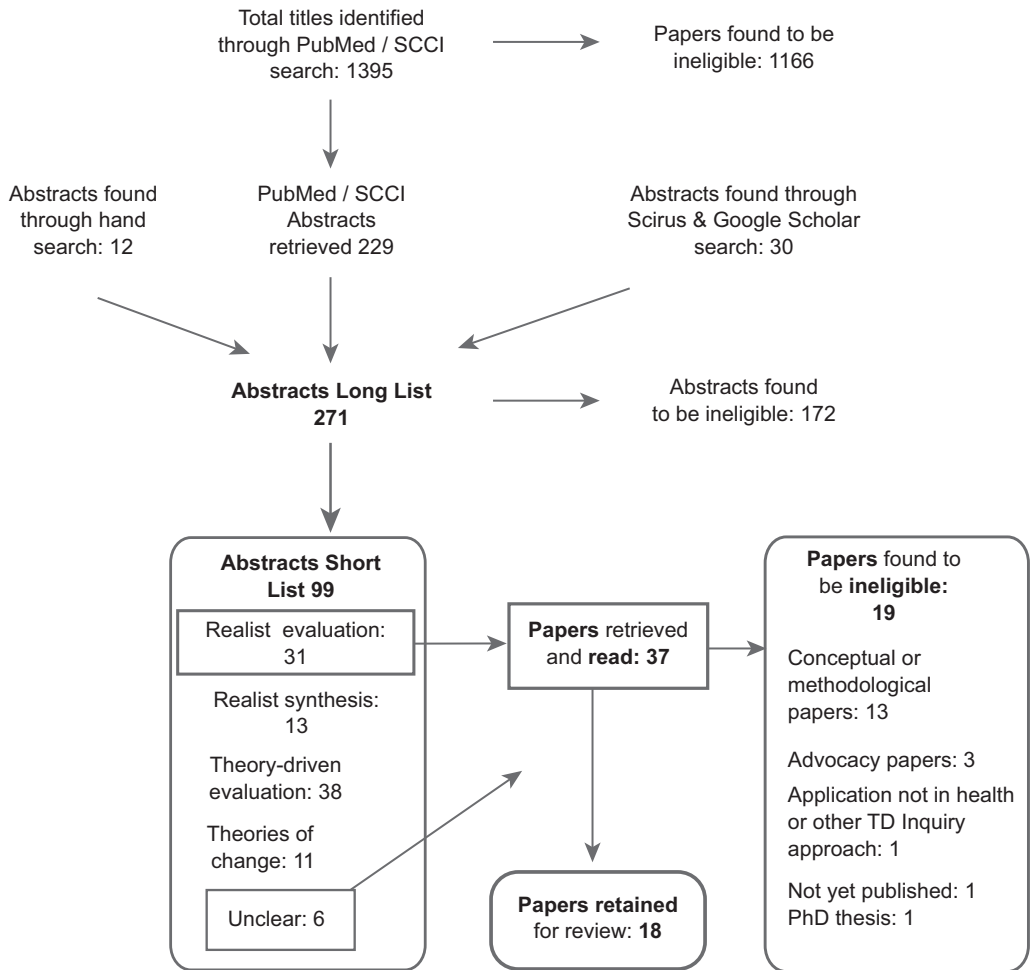
Figure 2. Overview of results of the search process by stage

papers were sorted into five categories on the basis of the description of the different approaches presented in the Introduction: realist evaluation (31 papers), realist synthesis (13), theory-driven evaluation (38), theories of change (11) and unknown (6). Because of the specific focus of this review, we analysed the papers from the categories 'realist evaluation' and 'unknown' (total of 37 papers). Upon reading, 19 papers were found to be non-eligible because they did not report on empirical applications of realist evaluation. The majority (13 papers out of 19) are advocacy papers calling for more use of realist evaluation in health research. Table 2 shows the 18 papers that were finally examined for this review. Each of these papers was reviewed independently by two research-ers to map their content in terms of the categories presented in Table 3.

Table 2. The list of papers examined for this review

Blaise P and Kegels G (2004) A realistic approach to the evaluation of the quality management movement in health care systems: a comparison between European and African contexts based on Mintzberg's organizational models. *International Journal of Health Planning and Management* 19: 337–64.

Byng R, Norman I and Redfern S (2005) Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness. *Evaluation* 11: 69–93.

Byng R, Norman I, Redfern S and Jones R (2008) Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation. *BMC Health Services Research* 8: 274.

Clark AM, Whelan HK, Barbour R and Macintyre PD (2005) A realist study of the mechanisms of cardiac rehabilitation. *Journal of Advanced Nursing* 52: 362–71.

Evans D and Killoran A (2000) Tackling health inequalities through partnership working: lessons from a realistic evaluation. *Critical Public Health* 10: 125–40.

Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C and Pawson R (2009) How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *Milbank Q* 87: 391–416.

Leone L (2008) Realistic evaluation of an illicit drug deterrence programme. *Evaluation* 14: 19–28.

Mackenzie M, Koshy P, Leslie W, Lean M and Hankey C (2009) Getting beyond outcomes: a realist approach to help understand the impact of a nutritional intervention during smoke cessation. *European Journal of Clinical Nutrition* 63: 1136–42.

Maluka S, Kamuzora P, San Sebastian M, Byskov J, Ndawi B, Olsen O and Hurtig AK (2011) Implementing accountability for reasonableness framework at district level in Tanzania: a realist evaluation. *Implementation Science* 6(1): 11.

Manzano-Santaella A (2011) A realistic evaluation of fines for hospital discharges: incorporating the history of programme evaluations in the analysis. *Evaluation* 17: 21–36.

Marchal B, Dedzo M and Kegels G (2010) A realist evaluation of the management of a well-performing regional hospital in Ghana. *BMC Health Services Research* 10: 24.

Marchal B, Dedzo M and Kegels G (2010) Turning around an ailing district hospital: a realist evaluation of strategic changes at Ho Municipal Hospital (Ghana). *BMC Public Health* 10: 787.

Ogrinc G and Batalden P (2009) Realist evaluation as a framework for the assessment of teaching about the improvement of care. *Journal of Nursing Education* 48: 661–7.

Pommier J, Guével M-R and Jourdan D (2010) Evaluation of health promotion in schools: a realistic evaluation approach using mixed methods. *BMC Public Health* 10: 43.

Ridde V and Guichard A (2011) Perception de quelques mécanismes favorables a la reduction des inégalités sociales de santé en France. *Global Health Promotion*. 18 (3): 47–60

Rycroft-Malone J, Fontenla M, Bick D and Seers K (2010) A realistic evaluation: the case of protocol-based care. *Implementation Science* 5(1): 38.

Tolson D, Mcintosh J, Loftus L and Cormie P (2007) Developing a managed clinical network in palliative care: a realistic evaluation. *International Journal of Nursing Studies* 44: 183–95.

Wand T, White K and Patching J (2010) Applying a realist(ic) framework to the evaluation of a new model of emergency department based mental health nursing practice. *Nursing Inquiry* 17: 231–9.

Findings

Our review of this body of research highlighted the following key points:

- the body of literature is very small and young;
- realistic evaluation has been applied in a variety of fields within health systems research;
- there is considerable diversity in the way in which the principles were applied; and
- the common challenge of applying the principles in practice.

Table 3. Literature-review framework

Categories	Description
Subject	Description of the domain and intervention to which realist evaluation was applied
Justification	Reasons why realist evaluation was used by the authors
Application of realist evaluation principles	Summary of how the authors applied the realist evaluation approach in practice
Methodological conclusions	Main reported methodological findings and issues as identified by the authors

The body of work and its scope

As shown in Table 2, we found only 18 papers. Strikingly, most were published between 2008 and 2010. This young body of work differs in its application. In seven studies, realist evaluation principles were applied in evaluation or research of programmes, in seven cases the application domain was health service organization and management and in four cases, clinical care. Of the 18 studies, only three applied realist evaluation in research in low- and middle-income countries (LMIC).

Most studies used realist evaluation in a standalone design, but some did use realist evaluation in addition to other designs. Byng et al. (2008) explain how realist evaluation was used in a case study design alongside a cluster RCT that assessed health care programmes for long-term mentally ill patients. In the study by Mackenzie et al. (2009), the realist evaluation was a part of a cluster randomized trial that assessed the impact of an intervention to reduce weight gain during a smoking cessation intervention.

Regarding the scope of the application of realist evaluation in the research process, we found that less than half of the papers discuss all the steps of the cycle presented by Figure 1. These include Byng et al. (2005), Evans and Killoran (2000), Leone (2008), Mackenzie et al. (2009), Marchal et al. (2010a, 2010b) and Manzano-Santaella (2011). Rycroft-Malone et al. (2010) and Pommier et al. (2010) present the development of the protocol of an realist evaluation study up to the point of proposing methods for data analysis. Other papers focus on the MRT and how it was developed (Wand et al., 2010) or on mechanisms (Ridde and Guichard, 2011).

Justification for the use of realist evaluation

Two main arguments are used to justify the use of realist evaluation in health systems research. First, it is argued that realist evaluation provides a sound framework to examine how context and mechanisms influence the outcomes of an intervention – the 'opening the black box' argument (see, for instance, Blaise and Kegels, 2004; Clark et al., 2005; Greenhalgh et al., 2009; Ogrinc and Batalden, 2009; Tolson et al., 2007).

Others state that realist evaluation is well suited to investigating complexity, either for evaluations of complex interventions (e.g. Byng et al., 2008; Maluka et al., 2011; Manzano-Santaella, 2011; Pommier et al., 2010; Wand et al., 2010) or of complex causal pathways (Ridde and Guichard, 2011; Rycroft-Malone et al., 2010). For Mackenzie et al. (2009), 'realist evaluation would offer an opportunity to develop an integrated outcome and process evaluation framework and would advance theoretical understanding of the best circumstances for increasing the impact of nutritional interventions'.

However, few of these authors commented in their discussion sections on their experience of the usefulness of realist evaluation with regard to studying complex issues. Byng et al. (2005) present a rather detailed methodological critique, pushing for more critical realism, while Pedersen and Rieper (2008) argue that realist evaluation fits complexity if the research design is adapted. Leone (2008) found that the construct of the CMO configuration was useful both while designing the evaluation and exploring the complexity of the intervention and its underlying processes. Marchal et al. (2010a) raise the issue of the attribution paradox, referring to the inherent difficulty in attributing changes in outcomes to interventions that are complex in nature.

How is the realist evaluation approach applied in practice?

We found a great diversity in the depth of the application of the philosophical concepts, the use of terminology and the scope of application in the research process.

Applying realism. The philosophical principles that underlie realist evaluation are variably interpreted and applied to different degrees. Most authors only fleetingly refer to the philosophical foundation of realist evaluation, which arguably is among its most distinctive features and provides much of its explanatory power.

If mentioned, there seem to be different interpretations of the philosophical roots. Greenhalgh et al. (2009), for instance, say that realist evaluation has a constructivist ontology and an interpretivist epistemology. This is in contrast to Connelly's (2007) claim that realist evaluation shares with critical realism a realist ontology. In Bhaskar's view, critical realism is 'reconciling ontological realism, epistemological relativism and judgmental rationality' (Bhaskar, 1998).

This view is adhered to by Byng et al. (2005), who make explicit comparisons between critical realism and realist evaluation. They note, for instance, that the feedback loops between outcomes of an intervention and the original mechanism are not much considered by Pawson and Tilley, in contrast to Bhaskar's work, in which these feedback loops are fundamental to the notion of emergence. In another related paper, Byng et al. (2008) point to the possibility of multiple mechanisms acting at the same time, which again they claim is stressed by Bhaskar and ignored by Pawson and Tilley. Also Kaneko (1999) and Wand et al. (2010) refer explicitly to critical realism.

Such different interpretations of the philosophical roots may have consequences for the way the realist evaluation study is carried out, and specifically for the manner in which mechanisms are analysed. Indeed, another notable difference has to do with the analysis of causality. Realist evaluation considers causality to be generative: actors and society have potential mechanisms of causation by their very nature. Change occurs when interventions, combined with the right contextual factors, release the generative mechanisms. The Context–Mechanism–Outcome (CMO) configuration is used as an analytical tool to analyse the data and unearth the mechanism.

Using terms interchangeably. The review shows that the terms 'theory-driven evaluation', 'theories of change' and 'realist evaluation' are often used interchangeably. Greenhalgh et al. (2009) use the term 'theory of change' to indicate what Pawson and Tilley (1997) call the MRT, and the term 'programme theory' for what the latter authors call the refined MRT. In their realist evaluation of health promotion in schools, Pommier et al., (2010) use the term 'theory of change' for the MRT they will test. Also Rycroft-Malone et al. (2010) use theory of change terminology in their paper on a realist evaluation of protocol-based care. Marchal et al. (2010a) borrow elements from theory-driven evaluation to make the description of the CMO more detailed (for instance, by adding a description of the process of the implementation of the intervention and its implementation context).

The middle range theory. The review found that a central element in the realist evaluation cycle, the MRT, is treated in various ways. Some papers do not mention the MRT (see, for instance, Clark et al., 2005; Tolson et al., 2007); others present a MRT but do not explain how it was developed (e.g. Leone, 2008; Wand et al., 2010); and a number of other papers explain clearly how the initial MRT was developed on the basis of existing theories. Examples include the paper by Blaise and Kegels (2004) who use Mintzberg's organizational configurations (Mintzberg, 1989) to categorize the structural configuration of the health systems in which they studied quality management interventions. Evans and Killoran (2000) describe an analytical framework consisting of six enabling factors, which they used to draw potential CMO configurations. Kaneko (1999) usefully describes mechanisms in smoking cessation (the medicalization mechanism, the primary group encouragement mechanism, the substitution mechanism, the role model mechanism) and context elements (the experience context, competing force context, environmental consciousness context and social pathology context). Each of these can constitute testable elements of a MRT. Mackenzie et al. (2009) present how they developed their study on a nutritional intervention during smoke cessation starting with a literature review and initial interviews with smoking-cessation advisors. They then mapped the myriad of potential CMO configurations and selected CMOs 'that could feasibly be studied in this intervention'. They tested these pathways using quantitative and qualitative data, and used the linkage of process and outcome data to refine the initial theories about what works for whom in which circumstances.

The CMO configuration. In practice, some authors apply the CMO configuration merely as a tool to describe sub-elements of the intervention and the context; this use of the CMO as a descriptive frame does not lead to the identification of the causal links between the intervention and the outcome, and the role of the context and the mechanism. Others use the CMO configuration as suggested by Pawson and Tilley: they analyse the link between intervention, context and mechanism to explain how change was brought about, in line with the idea of generative causality (see for instance Byng et al., 2005; Clark et al., 2005). Clark et al. (2005) describe the mechanisms that keep cardiac patients attending the preventive activities of cardiac rehabilitation programmes. They found that effective programmes trigger psychological mechanisms (e.g. regaining trust in the capacity of their body) and social mechanisms (such as camaraderie and building up social capital), but only if the programme operates in community-based settings where the patients feel safe (e.g. competent physical trainers).

Mechanisms. The definition of 'mechanism' is another bone of contention. Some authors stick to the definitions provided by Pawson and Tilley (1997). Pommier et al. (2010) cite from Pawson and Tilley's book: 'A mechanism is not a variable but an account of the behavior and interrelationships of the processes that are responsible for the change. A mechanism is thus a theory'. The definitions presented by Rycroft-Malone et al. (2010) and Ogrinc and Batalden (2009) seem to align with Pawson and Tilley's view. Also Clark et al. (2005) and Wand et al. (2010) applied similar definitions in their data analysis.

Yet, Evans and Killoran (2000) define mechanisms as including interventions or activities. Greenhalgh et al. (2009) define mechanism as 'the stakeholders' ideas about how change will be achieved'; the mechanisms they found include descriptions of the actual intervention. This is also the case with Maluka and colleagues (2011) and Tolson et al. (2007). The latter consider mechanisms in terms of barriers and facilitators, but seem to conflate activities and modes of implementation with mechanisms. Ridde and Guichard (2011) examined the mechanisms that may explain the effect of programmes to reduce social inequity in health. Their initial list has elements that indicate possible

Table 4. Methodological issues raised by the authors

Issues		Papers
Methodological problems that are presented	Difficulties to describe the MRT if not much is known about the issue under study	Rycroft-Malone et al. (2010)
	Problems with defining how deep one should delve into layers of social and physical reality to unearth mechanisms	Byng et al. (2005)
	Difficulties to differentiate mechanism from intervention	Marchal et al. (2010a)
	Difficulties to differentiate mechanism from context	Byng et al. (2005), Rycroft-Malone et al. (2010), Byng et al. (2008), Marchal et al. (2010a), Wand et al. (2010)
	Difficulties to describe relevant context elements	Greenhalgh et al. (2009), Marchal et al. (2010a)
	Difficulties to demonstrate attribution	Marchal et al. (2010a), Marchal et al. (2010b)
Practical issues	Time and resources required to do a good realist evaluation	Redfern et al. (2003), Wand et al. (2010), Marchal et al. (2010a)
	Lack of practical guidance	Rycroft-Malone et al. (2010)

mechanisms (e.g. degree of involvement of actors; leadership styles; empowerment), but also programme attributes, like interventions, actions, objectives, evaluation methods, etc.

Mechanism versus context. Another main source of diverging views is the difference between mechanism and essential context condition. Greenhalgh et al. (2009), for instance, call 'conditions for success' what could also be considered as 'mechanisms'. They present diagrams that spell out constraining and enabling factors (what others would call context conditions) and 'success' and 'disappointment' (outcomes). Also in the paper by Ogrinc and Batalden (2009) on the effect of teaching on the improvement of quality of care, there is some confusion between mechanism and context: is the learners' schedule a mechanism or a context element that needs to be taken into account?

Presenting the CMO configuration and the MRT. A range of methods is used to present the CMO configurations and the MRT. Some authors use narratives to describe the individual elements and interactions of the CMO configuration (for instance Byng et al., 2008; Clark et al., 2005; Marchal et al., 2010a). Others use tables (Greenhalgh et al., 2009; Leone, 2008; Tolson et al., 2007). Ogrinc and Batalden (2009) developed a realist hypothesis grid that contains elements of plausible mechanisms, contexts and outcomes and that is used to generate potential CMO configurations during the preparation phase of a realist evaluation study. Similarly, Byng et al. (2008) used predictor-outcome matrices that incorporated the CMO configurations and clarified the analytical induction processes. Others also use diagrams (Byng et al., 2005; Maluka et al., 2011).

The methodological issues as raised by the authors. Table 4 presents the methodological issues raised by the authors of the papers we analysed. It is striking that despite the diverging views on applying

the principles of realist evaluation as presented above, relatively few authors raise questions about methodological problems. Some authors mention the difficulty of describing the initial MRT if not much is known about the issue under study. Identifying mechanisms and describing and assessing the weight of context elements were also highlighted as challenges.

Discussion

Scope and limitations of this review

As explained in the Methods section above, this review does not claim to be the ultimate systematic review of realist evaluation in health systems research, given the difficulties of retrieving papers using a traditional search strategy. However, we believe our structured search found most of the relevant papers and many more than other researchers have reported. Greenhalgh et al. (2009) reported they found 5 realist evaluation papers in health. A comparison of our search results with the results of the recent review by Coryn et al. (2011) found that the authors of the latter review did not identify any of the papers retained in our review, although they broadly defined theory-driven evaluation to include realist evaluations. The results of our review were presented to researchers who are currently using realist evaluation in their work during the workshop on theory-driven inquiry for health systems research at the Institute of Tropical Medicine in Antwerp in November 2010 (see www.itg.be/tdi). This led to the identification of only one additional empirical study, which was being submitted at that stage, so our search could not yet have identified it.

Our search identified a large number of papers reporting on studies that assessed the effectiveness of theory-based interventions or the application of measurement tools based on theory. However, such studies cannot be called realist evaluation studies, nor theory-driven inquiry studies in a broader sense: none of them aim at unearthing the mechanisms and influence of the context, and nor do they attempt to refine a MRT.

About half of the papers on the realist evaluation shortlist did not meet our inclusion criteria because they did not report on the empirical application of realist evaluation principles. A number of them are, however, interesting as they deal with the philosophical underpinnings of realist evaluation, and more specifically with critical realism (see for instance Connelly, 2007; Dickinson, 2006; Julnes et al., 1998; McEvoy and Richards, 2003). These papers present much-needed clear introductions to the principles of critical realism. Others show that the call for theory-driven inquiry or realist evaluation in health research started about 10 years ago (Connelly, 2000; Judge and Bauld, 2001; Mingers, 2000). Gradually, this call became stronger (Mills et al., 2008; Walshe, 2007) and conceptually better framed (Connelly, 2000, 2004; Dickinson, 2006).

Application of realist evaluation in health research

We found a small body of work that covers a wide range of health research domains, from clinical practice and management to evaluation of health programmes and policies. It is also a strikingly young literature.

This suggests a slow uptake of realist evaluation principles in health systems research, given that Pawson and Tilley published their seminal book in 1997. This may be for a number of reasons. First, the analysis of the methodology of the papers we reviewed confirms most of the methodological challenges mentioned in the introduction. As we said earlier, these are recognized by some authors (see Table 4). Lack of methodological guidance remains an issue, highlighted by, for instance, Byng et al. (2008) and Rycroft-Malone et al. (2010). Second, not only do realist

evaluations demand both methodological and substantive expertise, they are also time-consuming. This is supported by Wand et al. (2010) and Redfern et al. (2003); for example, the authors of the latter study report that it proved too time consuming to fully exhaust all plausible CMO configurations. Third, researchers may be reluctant to carry out realist evaluations because they believe that realist evaluation cannot cope with complex interventions. This is refuted by Pedersen and Rieper (2008) who conclude that realist evaluation can be applied if some modifications to the design are made, as proposed by Pawson and Tilley (1997). For instance, the policy or intervention should be unpacked into its components, for each of which a typical CMO can be drafted and tested.

There may also be other factors leading to a slow uptake. In 1999, Cole (1999) highlighted difficulties around realist evaluation gaining academic credibility. This is not mentioned in any of the papers we reviewed, but peer pressure may play a role in deterring academic researchers from undertaking realist inquiry. The perceived credibility of the approach may also impact on publication of articles, although some journals have been calling for more attention to improving external validity through theory-based research (Kernick, 2006; Smith, 2000; Steckler and McLeroy, 2008).

Getting to the bottom of realism interpretations of the philosophical roots

We found that the principles of critical realism, on which realist evaluation is built, are applied to different degrees. It could be argued that this is not really an issue, since, compared with outcome evaluations, even a superficial application of realist evaluation has the advantage of exploring the processes and context instead of only the link between intervention and outcome (i.e. effectiveness). However, such application of realist evaluation would remain a mere process evaluation if it does not include critical realism's search for the deeper explanations of change. Realist evaluators state that it is not the intervention but the mechanism that needs to be transferred. Therefore, the patterns that underlie social change (i.e. the explanatory theory) need to be discovered.

The methodological issues presented in the next section, however, demonstrate that it may be rather more difficult to do this in practice, especially in the case of research or evaluations of complicated, multi-component interventions taking place in different contexts.

Methodological problems

The middle range theory. As our findings indicate, many authors find the practical application of the concept of 'middle range theory' difficult. Similar problems were reported in the field of theory-driven evaluation (Van Belle, 2010) and interesting methods can be found in this literature. For instance, Trochim proposes pattern matching, whereby the stakeholders are interviewed and the pattern of the theory proposed by them (or by extension derived from prior theory) is compared with the observations made by the evaluators (Trochim, 1989). Leeuw (2003) proposes to use argumentational analysis to draw useful propositions from data sources such as stakeholder interviews, project documents and published documentation. Argumentational analysis is a 'model for analysing chains of arguments and it helps to reconstruct and 'fill in' argumentations' (Leeuw, 2003). Applied to realist evaluation, this technique can be used to search the data for statements that point to expected mechanisms and context conditions. These resulting statements are listed and the mechanisms, outcomes and context conditions they refer to are linked into configurations. These are in turn analysed and compared with existing research findings and theories. In the soft systems approach, developed by Checkland (1999), the evaluators and the people involved in the intervention together describe the problem situation in the form of rich pictures. The resulting 'activity' model can in fact be viewed as a middle range theory.

In our own work, we found it useful to add a conceptually intermediate level of theory between the MRT and the CMO configuration. In the approach of Pawson and Tilley (1997), there is nothing between the MRT and the CMO configuration. While the latter is a very useful tool to categorize data and link the outcomes to mechanisms and context elements, the MRT as these authors define it is sometimes too abstract to serve as a tangible starting point for empirical work. This happens specifically when there is little published or known about the intervention one is examining. Pawson and Tilley (1997) provide a clue for an intermediate level. They describe how the interpretations of the actors involved in the intervention can be used to construct the MRT. This seems to correspond to the concept of 'programme theory', which in the theory-driven evaluation school is the set of assumptions held by the designers and implementers of a programme (Chen, 2005). This programme theory is tangible and can be described on the basis of interviews with programme designers and implementers. Using it as a starting point, links can be made both 'downward' to the categories that will be used to develop CMOs during the analysis and 'upward' to the drafting of the MRT.

Mechanisms. The definition of what constitutes a mechanism is essential in realist evaluation, as it affects the depth of analysis; and we found that the various authors define 'mechanism' in many different ways. These findings confirm the discussion presented by Astbury and Leeuw (2010). We found, for instance, that sub-categories of interventions or intervention modalities are often presented as 'mechanisms'. This is in contrast with a realist search for mechanisms in social-science research that is based on the interplay between structure and agency and the social embeddedness of change.

In our view, one of the main added values of a realist approach is that it looks for mechanisms at individual, group, organizational and societal levels. A hypothetical realist evaluation of a merger of hospitals within a city may need to look for mechanisms:

- at the level of health workers, where resistance to change may be due to fear and uncertainty, changes in workload, etc.
- At the team level, where many of these psychological reactions may play out as well, besides the effects such a merger may have on staff transfers and team composition, closing down of specific services, etc.
- At the organizational level, where other logics may influence the managers in implementing the change: besides psychological reactions, there will be institutional motives, policy pressure and financial incentives or constraints that may steer them toward policy adherence or resistance.

In other words, realist evaluation searches for psychological or social explanations of behaviour and seeks how social structure interacts with individual or group agency. This systemic approach provides a way to unravel complex issues.

Context. Defining 'context' and separating 'mechanism' from 'context' remains a difficult issue. Improving our understanding of the influence of the context on the outcomes of an intervention or on the problem at hand is, however, one of the key elements that set realist evaluation apart from effectiveness evaluations. We think it is useful to consider context elements as actors or other factors that are external to the intervention, present or occurring even if the intervention does not lead to an outcome, and which may have an influence on the outcome.

Our review shows that in practice, the context poses two major problems. First, both in the phase of describing the initial MRT as in the process of data collection, realist researchers are likely to be confronted with the problem of identifying those context elements that really matter. From a critical realism perspective, context is made up by the norms, regulations and procedures, as well as the barriers and facilitators that reside within the relationships between the involved actors and between the actors and the broader social structure (Connelly, 2000). The challenge is then to identify those factors that may affect the intervention and its outcome. Second, once they have been identified, these elements need to be monitored in time and their linkages with intervention and outcome need to be assessed.

Conclusions

This review shows that realist evaluation principles are gaining traction in health systems research, be it at a slow pace. A number of reasons may explain this slow uptake, despite the generally claimed usefulness of the approach for research and evaluation of complex issues and interventions.

At this (early) stage of applying realist evaluation principles in health systems research, diverging views and interpretations of concepts and methods should not surprise us. After all, the ontological underpinnings and their epistemological consequences can be quite challenging and the books by Bhaskar and colleagues are not for the faint-hearted.

We found, however, that there is now a series of papers that clearly present the basic principles of critical realism and apply these to research of health services and systems. We also found a broad scope of application, from research in clinical care to policy analysis, and a number of papers that systematically present the methods used in applying realist evaluation.

While we would agree with Byng et al. (2005) that realist evaluation presents a useful approach but that its principles should not slavishly be adhered to, we also believe that to fully realize the potential of this approach, more clarity is needed concerning the definitions of mechanisms and context and how the configuration of context, mechanism and intervention can be described and assessed. Only more and well-documented empirical and conceptual work will allow us to better understand how this can be done.

Funding

Notes

1. The terms 'theories of change' and' theory of change' are often used interchangeably.
2. Pawson and Tilley's 1997 book used the term 'realistic evaluation'. Because the term 'realist' is now widely used by other authors (see, for instance, Kazi, 2003), they switched to the latter (Pawson and Tilley, 2006).

References

Astbury B and Leeuw F (2010) Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31: 363–81.

Bamberger P (2008) Beyond contextualisation: using context theories to narrow the micro-macro gap in management research. *Academy Of Management Journal* 51: 839–46.

Barnes M, Matka E and Sullivan H (2003) Evidence, understanding and complexity: evaluation in non-linear systems. *Evaluation* 9: 265–84.

Berwick DM (2008) The science of improvement. *JAMA* 299: 1182–4.

Bhaskar R (1998) Philosophy and scientific realism. In: Archer M, Bhaskar R, Collier A, Lawson T and Norrie A (eds) *Critical Realism. Essential Readings.* London & New York: Routledge.

Blaise P and Kegels G (2004) A realistic approach to the evaluation of the quality management movement in health care systems: a comparison between European and African contexts based on Mintzberg's organizational models. *International Journal of Health Planning and Management* 19: 337–64.

Blamey A and Mackenzie M (2007) Theories of change and realistic evaluation: peas in a pod or apples and oranges? *Evaluation* 13(4): 439–55.

Byng R, Norman I and Redfern S (2005) Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness. *Evaluation* 11: 69–93.

Byng R, Norman I, Redfern S and Jones R (2008) Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation. *BMC Health Services Research* 8: 274.

Calnan M and Ferlie E (2003) Analysing process in healthcare: the methodological and theoretical challenges. *Policy and Politics* 31: 185–93.

Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P and Kinmonth AL (2007) Designing and evaluating complex interventions to improve health care. *British Medical Journal* 334: 455–9.

Checkland P (1999) *Systems Thinking, Systems Practice*. Chichester: Wiley.

Chen H-T (1989) The conceptual framework of the theory-driven perspective. *Evaluation and Program Planning* 12: 391–6.

Chen H-T (1990) *Theory-Driven Evaluations*. Newbury Park, CA: SAGE.

Chen H-T (2005) *Practical Program Evaluation*. Thousand Oaks, CA: SAGE.

Chen H-T and Rossi P (1980) Multi-goal theory-driven approach: a model linking basic and applied science. *Social Forces* 59(1): 106–22.

Chen H-T and Rossi P (1983) Evaluating with Sense – The Theory Driven Approach. *Evaluation Review* 7(3): 283–302.

Chen H-T and Rossi P (1987) The theory-driven approach to validity. *Evaluation and Program Planning* 10: 95–103.

Chen H-T and Rossi P (1989) Issues in the theory-driven perspective. *Evaluation and Program Planning* 12: 299–306.

Clark AM, Whelan HK, Barbour R and Macintyre PD (2005) A realist study of the mechanisms of cardiac rehabilitation. *Journal of Advanced Nursing* 52: 362–71.

Cole G (1999) Advancing the development and application of theory-based evaluation in the practice of public health. *American Journal of Evaluation* 20: 453–70.

Connell JP, Kubisch A, Schorr L and Weiss C (1995) *New Approaches To Evaluating Community Initiatives. Concepts, Methods, and Contexts*. Washington, DC: The Aspen Institute.

Connelly J (2000) A realistic theory of health sector management: the case for critical realism. *Journal of Management in Medicine* 14: 262–71.

Connelly J (2004) Realism in evidence based medicine: interpreting the randomised controlled trial. *Journal of Health Organization and Management* 18: 70–5.

Connelly J (2007) Evaluating complex public health interventions: theory, methods and scope of realist enquiry. *Journal of Evaluation in Clinical Practice* 13(6): 935–41.

Coryn C, Noakes L, Westine C and Schröter D (2011) A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation* 32: 199–226.

Costner H (1989) The validity of conclusions in evaluation research: a further development of Chen and Rossi's theory-driven approach. *Evaluation and Program Planning* 12: 345–53.

Dickinson H (2006) The evaluation of health and social care partnerships: an analysis of approaches and synthesis for the future. *Health and Social Care in the Community* 14: 375–83.

Douglas FC, Gray DA and Van Teijlingen ER (2010) Using a realist approach to evaluate smoking cessation interventions targeting pregnant women and young people. *BMC Health Services Research* 10: 49.

Eccles M, Grimshaw J, Campbell M and Ramsay C (2003) Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality and Safety in Health Care* 12: 47–52.

Evans D and Killoran A (2000) Tackling health inequalities through partnership working: lessons from a realistic evaluation. *Critical Public Health* 10: 125–40.

Fiss P (2007) A set-theoretic approach to organizational configurations. *Academy of Management Review* 32: 1180–98.

Fulbright-Anderson A, Kubisch A and Connell J (1998) *New Approaches To Evaluating Community Initiatives. Theory, Measurement, and Analysis*. Washington, DC: Aspen Institute.

Fulop N, Allen P, Clarke A and Black N (2001) *Studying the Organisation and Delivery of Health Services: Research Methods*. London: Routledge.

Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C and Pawson R (2009) How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *Milbank Q* 87: 391–416.

Grimshaw J, Wilson B, Campbell M, Eccles M and Ramsay C (2001) Epidemiological methods. In: Fulop N, Allen P, Clarke A and Black N (eds) *Studying the Organisation and Delivery of Health Services: Research Methods.* London: Routledge.

Grocott P, Cowley S and Richardson A (2002) Solving methodological challenges using a theory-driven evaluation in the study of complex patient care. *Evaluation* 8: 306–21.

Judge K and Bauld L (2001) Strong theory, flexible methods: evaluating complex community-based initiatives. *Critical Public Health* 11: 19–38.

Julnes G, Mark M and Henry G (1998) Promoting realism in evaluation. realistic evaluation and the broader context. *Evaluation* 4: 483–504.

Kaneko M (1999) A methodological inquiry into the evaluation of smoking cessation programmes. *Health Education Research* 14(3): 433–41.

Kazi M (2003a) Realist evaluation for practice. *British Journal of Social Work* 33: 803–18.

Kazi M (2003b) *Realist Evaluation in Practice. Health and Social Work*. London: SAGE.

Kernick D (2006) Wanted–new methodologies for health service research. is complexity theory the answer? *Family Practice* 23: 385–90.

Kernick D and Mannion R (2005) Developing an evidence base for intermediate care delivered by GPs with a special interest. *British Journal of Medical Practice* 55(521): 908–10.

Leeuw F (2003) Reconstructing program theories: methods available and problems to be solved. *American Journal of Evaluation* 24: 5–20.

Leone L (2008) Realistic evaluation of an illicit drug deterrence programme. *Evaluation* 14: 19–28.

Mackenzie M and Blamey A (2005) The practice and theory: lessons from the application of a theories of change approach. *Evaluation* 11(2): 151–68.

Mackenzie M, Koshy P, Leslie W, Lean M and Hankey C (2009) Getting beyond outcomes: a realist approach to help understand the impact of a nutritional intervention during smoke cessation. *European Journal of Clinical Nutrition* 63: 1136–42.

Maluka S, Kamuzora P, San Sebastian M, Byskov J, Ndawi B, Olsen O and Hurtig AK (2011) Implementing accountability for reasonableness framework at district level in Tanzania: a realist evaluation. *Implementation Science* 6(1): 11.

Manzano-Santaella A (2011) A realistic evaluation of fines for hospital discharges: incorporating the history of programme evaluations in the analysis. *Evaluation* 17: 21–36.

Marchal B, Dedzo M and Kegels G (2010a) A realist evaluation of the management of a well-performing regional hospital in Ghana. *BMC Health Services Research* 10: 24.

Marchal B, Dedzo M and Kegels G (2010b) Turning around an ailing district hospital: a realist evaluation of strategic changes at Ho Municipal Hospital (Ghana). *BMC Public Health* 10: 787.

Mason P and Barnes M (2007) Constructing theories of change: methods and sources. *Evaluation* 13: 151–70.

McEvoy P and Richards D (2003) Critical realism: a way forward for evaluation research in nursing? *Journal of Advanced Nursing* 43: 411–20.

Mercier C, Piat M, Peladeau N and Dagenais C (2000) An application of theory-driven evaluation to a drop-in youth center. *Evaluation Review* 1: 73–91.

Merton RK (1968) *Social Theory and Social Structure*. New York: The Free Press.

Mills A, Gilson L, Hanson K, Palmer N and Largarde M (2008) What do we mean by rigorous health-systems research? *Lancet* 372: 1527–9.

Mingers J (2000) The contribution of critical realism as an underpinning philosophy for Or/Ms and systems. *Journal of the Operational Research Society* 51: 1256–70.

Mintzberg H (1989) *Mintzberg on Management. Inside Our Strange World of Organizations*. New York: The Free Press.

Ogrinc G and Batalden P (2009) Realist evaluation as a framework for the assessment of teaching about the improvement of care. *Journal of Nursing Education* 48: 661–7.

Pawson R (2002) Evidence-based policy: the promise of 'realist synthesis'. *Evaluation* 8: 340–58.

Pawson R (2006) *Evidence-Based Policy. A Realist Perspective*. London: SAGE.

Pawson R and Tilley N (1997) *Realistic Evaluation*. London: SAGE.

Pawson R, Greenhalgh T, Harvey G and Walshe K (2005) Realist review–a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy* 10(Suppl. 1): 21–34.

Pedersen L and Rieper O (2008) Is realist evaluation a realistic approach for complex reforms? *Evaluation* 14: 271–93.

Pommier J, Guével M-R and Jourdan D (2010) Evaluation of health promotion in schools: a realistic evaluation approach using mixed methods. *BMC Public Health* 10: 43.

Redfern S, Christian S and Norman I (2003) Evaluating change in health care practice: lessons from three studies. *Journal of Evaluation in Clinical Practice* 9: 239–49.

Ridde V and Guichard A (Forthcoming) Perception de quelques mécanismes favorables a la reduction des inégalités sociales de santé en France. *Global Health Promotion*.

Rodriguez E and Mead J (1997) Evaluating a community oriented primary care program: lessons learned through a theory-oriented approach. *Evaluation and Program Planning* 20: 217–24.

Rogers P (2007) Theory-based evaluation: reflections ten years on. *New Directions for Evaluation* 114: 63–84.

Rycroft-Malone J, Fontenla M, Bick D and Seers K (2010) A realistic evaluation: the case of protocol-based care. *Implementation Science* 5(1): 38.

Sidani S, Doran DM and Mitchell PH (2004) A theory-driven approach to evaluating quality of nursing care. *Journal of Nursing Scholarship* 36(1): 60–5.

Smith R (2000) Quality improvement reports: a new kind of article. *British Medical Journal* 321: 1428.

Stame N (2004) Theory-based evaluation and types of complexity. *Evaluation* 10: 58–76.

Steckler A and Mcleroy KR (2008) The importance of external validity. *American Journal of Public Health* 98: 9–10.

Sturmberg J and Martin C (2009) Complexity and health – yesterday's traditions, tomorrow's future. *Journal of Evaluation in Clinical Practice* 15: 543–8.

Sullivan H, Barnes M and Matka E (2002) Building collaborative capacity through 'theories of change': early lessons from the evaluation of health action zones in England. *Evaluation* 8: 205–26.

Tolson D, Mcintosh J, Loftus L and Cormie P (2007) Developing a managed clinical network in palliative care: a realistic evaluation. *International Journal of Nursing Studies* 44: 183–95.

Trochim W (1989) Pattern matching and program theory. *Evaluation and Program Planning* 12: 355–66.

Van Belle S, Marchal B, Dubourg D and Kegels G (2010) How to develop a theory-driven evaluation design – lessons learned from an adolescent sexual and reproductive health programme in West-Africa. *BMC Public Health* 10: 741.

Victora CG, Habicht JP and Bryce J (2004) Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health* 94: 400–5.

Walshe K (2007) Understanding what works–and why–in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care* 19: 57–9.

Wand T, White K and Patching J (2010) Applying a realist(ic) framework to the evaluation of a new model of emergency department based mental health nursing practice. *Nursing Inquiry* 17: 231–9.

Weiss C (1995) Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives. In: Connell J, Kubisch A, Schorr L and Weiss C (eds) *New Approaches to*

Evaluating Comprehensive Community Initiatives: Concepts, Methods and Contexts. Washington, DC: The Aspen Institute.

Weiss C (1997) Theory-based evaluation: past, present, and future. *New Directions for Program Evaluation* 76: 41–55.

Ying Ho S (1999) Evaluating urban regeneration programmes in Britain. exploring the potential of the realist approach. *Evaluation* 5: 422–38.

Bruno Marchal, MD, MPH is a research fellow at the Institute of Tropical Medicine, Antwerp (ITM-A) in Belgium. He wrote a PhD study on the role of (human resource) management in performance of hospitals in Ghana using realist evaluation. His research interests lie in management of local health systems and research of complex issues. Please address correspondence to: ITM-A, Nationalestraat 155, B-2000, Antwerp, Belgium. [email: bmarchal@itg.be]

Sara van Belle is a political scientist and anthropologist who worked at ITM-A as research fellow before joining the Belgian Technical Cooperation (BTC) as senior expert. She is undertaking a PhD study of accountability of non-state actors in the field of reproductive health in Africa and has a keen interest in theory-driven inquiry. Please address correspondence to: BTC, Hoogstraat 147, B-1000 Brussels, Belgium. [email: sara.vanbelle@btcctb.org]

Josefien van Olmen (MD, MPH) is currently doing research at the ITM-A in the field of organization of health services for life-long diseases in RD Congo, the Philippines and Cambodia, focusing on diabetes. She has an interest in health systems research, health systems strengthening, evaluation methods and realist synthesis. Please address correspondence to: ITM-A, Nationalestraat 155, B-2000, Antwerp, Belgium. [email: jvolmen@itg.be]

Tom Hoerée (MD, MPH) is course coordinator of the MPH programme Health systems management and policy at the ITM-A. He has been long involved in realist evaluation and evaluation methods for complexity. He is currently involved in the evaluation of training of district medical officers in Karnataka, India. Other areas of interest lie in health systems research, management and policy. Please address correspondence to: ITM-A, Nationalestraat 155, B-2000, Antwerp, Belgium. [email: thoeree@itg.be]

Guy Kegels (MD, PhD) is professor in public health at the ITM-A, with a long experience in health systems research in Africa and Asia. His most recent research interests cover theory-driven inquiry, complexity, decision-making and chronic diseases. Please address correspondence to: ITM-A, Nationalestraat 155, B-2000, Antwerp, Belgium. [email: gkegels@itg.be]