Realist evaluation of health promotion interventions:

a scoping review

Malengreaux S.[1] (segolene.malengreaux@uclouvain.be), Doumont D.[1], Scheen B.[1], Van Durme T.[2], Aujoulat I.[1]

Abstract

This scoping review aims to give a narrative account of existing realist evaluation practices in health promotion. Realist evaluations of health promotion interventions published between 2010 and 2021 were identified from searching five academic databases: Embase, Pubmed, PsycINFO, ScienceDirect and Scopus. A data-charting form was created based on the characteristics of realist evaluation and four core features of an approach appropriate for evaluating health promotion interventions. Seventeen (17) articles met the inclusion criteria. These were classified into two types of studies: those aiming to build an initial program theory, and those aiming to test an initial program theory. Our results revealed a great variety of realist evaluation practices and uncovered a growing interest in realist evaluation over the years. Our searches identified a lack of participative practice and capacity-building intention. Our examination of the data collection and analysis methods points to some common practices in using multi-methods. Perspectives on realist evaluation practices and on assessing the effectiveness of health promotion have been identified.

Lay summary

This scoping review aims to critically examine current practices of realist evaluation in the field of health promotion with respect to four core features of an approach appropriate for evaluating health promotion initiatives, namely: the need to accommodate the complex nature of health promotion interventions; drawing on a variety of disciplines and a broad range of information-gathering procedures; involving stakeholders in the evaluation; and building capacity for

addressing health promotion concerns. Seventeen (17) articles met the inclusion criteria. These were classified into two types of studies: those aiming to build an initial program theory, and those aiming to test an initial program theory. Our results suggest that the use of realist evaluation approach in the field of health promotion is guided mainly by the need to accommodate the complex nature of health promotion interventions. Our searches identified a lack of participative practice and capacity-building intention in current practice. Our examination of the data collection and analysis methods points to some common practices in using multi-methods. Perspectives on more meaningful practices of realist evaluation and on more relevant evaluation practices of the effectiveness of health promotion have been identified.

Introduction

In tackling the diversity of health determinants, health promotion (HP) strategies require complex interventions (Skivington *et al.,* 2021). The interactivity between an intervention and the context of its implementation generates a series of complex characteristics, such as non-linearity, the emergence of patterns, ongoing adaptations, uncertainty, dynamism and the co-evolvement of agents (Bisset *et al.*, 2013; Jolley, 2014). As suggested by Cambon et al., the crucial role of context in HP calls for using the term 'interventional system' rather than 'intervention', since the "interrelated human and non-human contextual agents within spatial and temporal boundaries [generate] mechanistic configurations which are prerequisites for change in health" (Cambon *et al.*, 2019). HP interventions should rank among complex interventions, requiring evaluation approaches embedded in complexity science. Occurring within intertwined systems, HP interventions do not respond in a linear way according to what the initiators of change intend (Greenhalgh and Papoutsi, 2018). By recognising these properties and the role of the context of their implementation when producing changes, the separation

between process and outcome evaluation becomes blurred. Therefore, the focus of evaluation broadens towards a "contextualized understanding of how an intervention induces change" (Skivington *et al.*, 2021). While recent publications have embraced this shift, methodological debates remain (Moore *et al.*, 2019).

Evaluating HP interventions also calls for a process of evaluation that is aligned with important principles for action within the field of HP (Potvin *et al.*, 2009). The World Health Organisation European Working Group on Health Promotion Evaluation (Rootman *et al.*, 2001) concluded that approaches appropriate for the evaluation of HP interventions should be based on the following four core features: (i) the evaluation design should be appropriate to the complex nature of HP intervention; (ii) the evaluation approach should implement a "broad range of information-gathering procedures" and rely on "a variety of disciplines"; (iii) the evaluation approach should be participative, meaning that it should involve legitimate stakeholders in "appropriate ways"; (iv) the evaluation process should represent an opportunity for the capacity-building of individuals, communities, organisations and governments (Rootman *et al.*, 2001).

The realist evaluation approach

Given the complex nature of HP interventions, understanding their effectiveness implies moving from the traditional question ("Is this intervention working?") to a more comprehensive question: "How, why, for whom and in what circumstances is this intervention working?" Realist evaluation (RE) as a methodology is particularly well-designed for addressing these types of questions. Developed by Pawson and Tilley, RE looks at the effectiveness question in a more open and nuanced way, adopting a view of causality that is less deterministic and simplistic because it aims to explain change and causal links by revealing the mechanisms of change and the influence of context (Marchal *et al.*, 2012). This has always been a puzzle for

HP practitioners, researchers and policy makers alike. RE therefore seems to offer some valuable perspectives on HP interventions.

One key feature at the core of RE is the generative approach to causality. Simply put, causality is not explained by the interaction between the intervention and its outcome but rather by the interplay and integration of three explanatory components: *Context*, *Mechanism* and *Outcome* (CMO) (Lacouture *et al.*, 2015). The concept of *Mechanism* refers to the interaction between the choices (reasoning) of individuals and the capacities (resources) provided by a given intervention, helping to understand what it is about the program that makes it work – or not. *Context* is a wider concept than location and infrastructure, as it may encompass pre-existing social conditions (rules, norms, values, etc.), institutional characteristics, interpersonal relationships, characteristics and capacities of individuals (Pawson, 2013). Absence or presence of these contextual characteristics will activate *Mechanisms* (or fail to activate them) in a causal relationship, which in turn generates *Outcomes*. In RE, the central task is therefore to model the different ways in which these "ingredients" of causality come together in "configurations" (i.e. CMO configurations) (Pawson and Tilley, 1997).

A realist evaluation starts and ends with a theory, like a cycle, but with movements back and forth involving multiple revision stages. The evaluation starts by framing "propositions about how mechanisms are fired in contexts to produce outcomes" (Pawson and Tilley, 1997, p.85) using initial theoretical CMO configurations in a heuristic way. These constitute the initial program theory which is then tested systematically through several iterations of data collection and analysis. Finally, based on the findings of the analysis, the initial program theory is refined or consolidated, and sometimes rejected. The use of RE may appear particularly relevant for quality improvement in HP, as the definition of a program theory and its refinement or consolidation may offer interesting perspectives to support decision-making and guide action in specific contexts (Blaise, 2004).

Aim and research questions

The aim of this paper is to provide an overview of current evaluation practices that are informed by RE principles, and to critically examine the contribution of RE for the evaluation of HP interventions with respect to the four important afore mentioned features defined by Rootman *et al.* (2001) for the evaluation of HP interventions. A scoping review method was selected as it addresses broad topics beyond intervention effectiveness and is adapted to present an account of existing literature (Arksey and O'Malley, 2005).

Methods

Search strategy

Using PRISMA-ScR guidelines (Tricco *et al.*, 2018), we conducted a literature review to identify cases of RE applied to HP intervention. Our search was performed across five databases (Embase, Pubmed, PsycINFO, ScienceDirect and Scopus), using the following keywords in the articles and abstracts: "HEALTH PROMOTION" and "REALIST EVALUATION". The literature search was conducted between December 2020 and December 2021, and encompasses publications between 1st January 2010 and 31st December 2021 (see Supplementary file 1 for more details about the search strategy). We used an iterative process to identify papers to be included, building on the insights of the first papers identified.

Papers were included if they presented (i) the results of an evaluation labelled as "RE", and (ii) an evaluation of projects or programs labelled as "health promotion interventions" – meaning that the intervention met more than one of a series of recognised health promotion strategies and principles, which we summarise hereafter in Table 1. We did not restrict our search to specific populations or themes. Reviews and protocols were excluded.

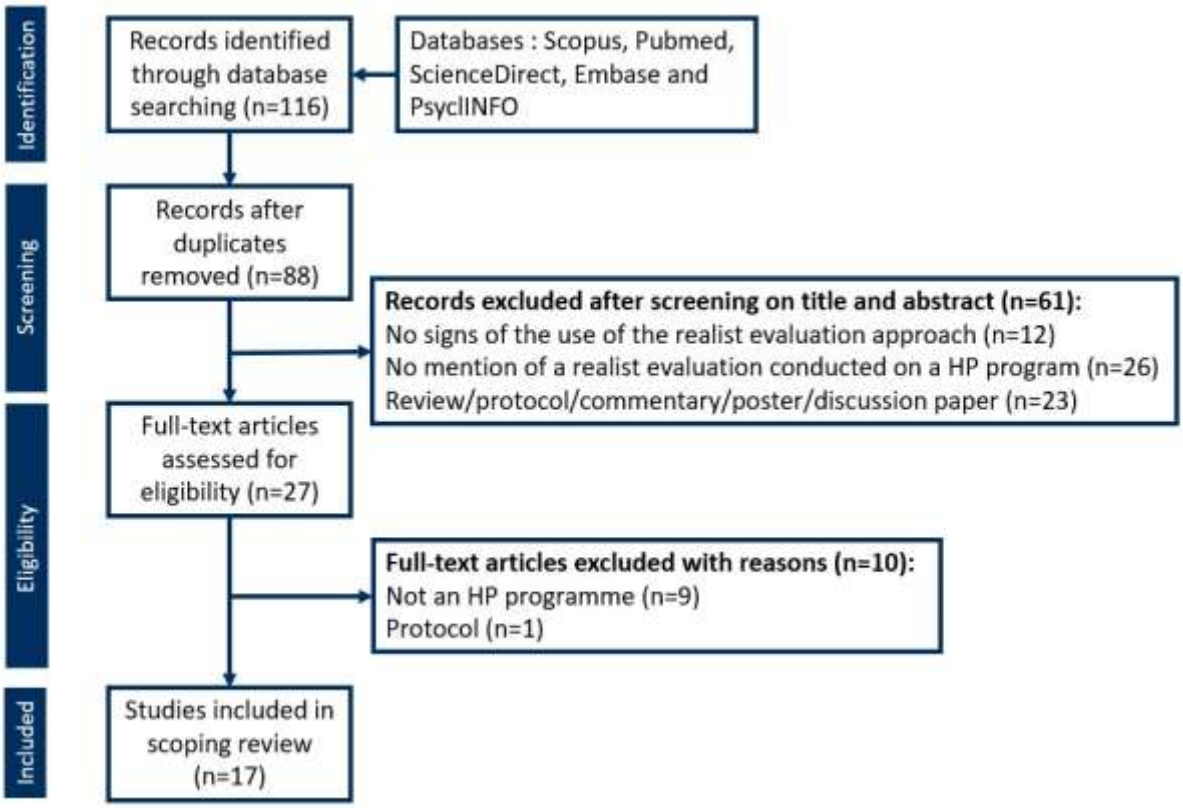
Table 1: Key principles and strategies for defining an intervention as a HP intervention

Strategies / Key principles	Build healthy public policy	Create supportive environments	Strengthen community action	Develop personal skills	Reorient health services
Empowering					
Equitable					
Holistic					
Intersectoral					
Participatory					
Sustainable					
Multi-strategy					

Source: Adapted from Rootman et al., 2001

One investigator (first author) conducted the database search and two investigators (first and third authors) screened the article titles and abstracts in order to reach a consensus on a list of articles to be reviewed. The full texts of those articles were then independently screened by two investigators (first and second or fifth author) to determine their eligibility.

Our initial search retrieved 88 papers (after removal of duplicates). After screening titles and abstracts, 27 papers were selected for full-text assessment. Seventeen papers met the inclusion criteria and were included in the final selection. The PRISMA flow diagram (Figure 1) illustrates the selection process.



Data analysis

Three investigators (first, second and fifth authors) developed a data-charting form on Excel that helped extract information related to the study characteristics (such as the evaluation questions, objectives and the description of the program), the RE characteristics (such as the rationale for using RE, information linked to the initial and revised program theories and CMO configurations), and the core features of an evaluation approach appropriate for HP interventions (i.e. participation, capacity-building, relying on a variety of disciplines and information-gathering techniques and acknowledging complexity). The same investigators independently read one of the included studies and completed the data-charting form as a trial to test its applicability. The form was adapted in order to reach a common understanding of each category (see Supplementary file 2 for the list of all topics for which data were sought), and was then used as a framework for the analysis of the selected articles, and periodically revisited by the research team. One investigator (first author) charted the data from the selected

articles. In addition to analysing the content of these articles, the authors' affiliation and contributions were systematically screened in order to document the range of academic and non-academic disciplines involved in the study.

Results

The 17 included papers described the evaluation processes and outcomes of 17 different interventions that we categorised as HP interventions based on recognised HP strategies and principles (see Table 1). Therefore, in this article, the term "studies" refers to the content of the evaluation process and its results, while the terms "(evaluated) intervention" refer to the object of evaluation.

Characteristics of the evaluated interventions

Table 2 presents a global description of the interventions evaluated in the included studies, as well as their classification in relation to HP strategies and key principles outlined in Table 1.

Table 2: Characteristics of the evaluated programs

Authors and year	Evaluated programs description and scale of the evaluation	Group type (A or B)*	HP strategies					HP principles						
			Build healthy public policy	Create supportive environments	Strengthen community action	Develop personal skills	Reorient health services	Participatory	Intersectoral	Empowering	Holistic	Equitable	Sustainable	Multi-strategy
Darlington *et al.,* 2017	Support school practice to prevent childhood obesity. The program was implemented in multiple schools in one region of France.	A		■		■			■			■		■
Jonsson and Goicolea, 2020	Two different but similar interventions that seek to facilitate the social integration of young people in education and employment related spheres. These interventions were implemented in different parts of a region in Sweden.	A				■			■	■				
O'Rourke *et al.,* 2019	Community-based volunteer doula support program which provides free antenatal, birth and postnatal doula support for women experiencing socioeconomic disadvantage. This program was implemented in one city of Australia.	A				■				■		■		
Taylor *et al.,* 2010	Consumer-driven community mental health centre providing safe and supportive environments, social connections, and activities for people with a lived experience of mental illness. This unique centre is situated in Australia.	A		■	■	■		■		■	■			■
Willis *et al.,* 2018	Program facilitating involvement in physical activity for children and youths with disabilities in local environments implemented in a rehabilitation centre in Norway.	A			■	■		■	■	■	■			■
Bergeron *et al.,* 2020	Intersectoral oral health promotion program implemented by health centres in multiple schools in Peru.	B		■					■			■		■
Darlington *et al.,* 2018	Training in health promotion strategies for school professionals and local agents implemented in multiple schools in France.	B				■		■	■					
Deschesnes *et al.,* 2014	Professional development in the 'healthy school' approach implemented in multiple schools in one province of Canada.	B				■		■	■					
Pals *et al.,* 2016	Toolkit for professionals involved in therapeutic education with people with chronic disease implemented in multiple health centres across Denmark.	B				■		■		■				
Tennant *et al.,* 2020	Provide long-term multidisciplinary care coordination, capacity building and integrated care for families with complex health and social needs in one region of Australia.	B		■		■		■	■	■	■	■		■
Owusu-Addo *et al.,* 2020	Provide cash and free access to health insurance to extremely poor and labour-constrained households in one region of Ghana.	B		■						■		■		
Van der Veken *et al.,* 2020	Program of sports activities generating social cohesion, health and inclusion (especially for vulnerable populations) through a local football club in Belgium.	B		■	■	■				■		■		

Mathias *et al.,* 2019	Peer-led, community-based, participatory group intervention on social inclusion and mental health among young people affected by psycho-social disability in one region in India.	B													
Uzochukwu *et al.,* 2020	Advocacy activities for impressing on stakeholders the need to sustain the maternal and child health services as a policy priority that were organized in a region in Nigeria.	B													
Bysted *et al.,* 2020	In Denmark, health professionals work from local offices in each neighborhood. They engage in active recruitment by going door-to-door, informing residents about the intervention and inviting them to participate in the free local health services.	B													
Howarth *et al.,* 2021	UK collaborative pilot project of social prescribing to a garden (Royal Horticultural Society Garden Bridgewater). Up to 75 people were to be referred by a health care professional to the garden on 3-month placements where they undertook therapeutic gardening activities.	B													
Martin-Fernandez *et al.,* 2021	Project TC-REG aims to improve use of scientific knowledge by implementing tailor-made knowledge transfer strategies in the field of health promotion and disease prevention in four regions in France.	B													

**Group A: describes how the program might work (initial program theory); Group B: how the program actually worked (refined program theory)*

Characteristics of the RE design

With regards to the RE cycle, 6 studies (hereafter referred to as belonging to Group A: Darlington *et al.*, 2017; Howarth *et al.*, 2021; Jonsson and Goicolea, 2020; O'Rourke *et al.*, 2020; Taylor *et al.*, 2010; Willis *et al.*, 2018) were at the stage of formulating hypotheses of how the program *might work*, for whom and in what circumstances (i.e. the initial program theory), while 11 studies (hereafter referred to as belonging to Group B: Bergeron *et al.*, 2020; Bysted *et al.*, 2020; Martin-Fernandez *et al.*, 2021; Darlington *et al.*, 2018; Deschesnes *et al.*, 2014; Mathias *et al.*, 2019; Owusu-Addo *et al.*, 2020; Pals *et al.*, 2016; Tennant *et al.*, 2020; Uzochukwu *et al.*, 2020; Van der Veken *et al.*, 2020) were at the stage of assessing how the program *worked*, for whom and in what circumstances (i.e. the revision of the initial program theory). First, we briefly describe how the initial and revised program theories were defined in the articles included.

Initial Program Theory (IPT)

For studies from Group A, the IPT was presented in the results section, while for studies from Group B, the IPT was presented in the introductory section, except for three studies from Group B (Bergeron *et al.*, 2020; Owusu-Addo *et al.*, 2020; Uzochukwu *et al.*, 2020) which did not present the IPT in their articles. The CMO framework, or its variants *Context-Intervention-Mechanism-Outcome* (Taylor *et al.*, 2010) and *Intervention-Context-Actor-Mechanism-Outcome* (Jonsson and Goicolea, 2020), were used to formulate the IPT with the help of graphs, tables or text, except in two studies (Darlington *et al.*, 2017; Howarth *et al.*, 2021) which did not use the CMO framework. All but one of the studies defined the IPT by listing potential *Contexts* and *Mechanisms* that could explain one or more *Outcomes*. In contrast, Jonsson and Goicolea (Jonsson and Goicolea, 2020) designed their IPT as an overall causal relationship.

Revised Program Theory (RPT)

Because they did not aim at revising the program theory, studies from Group A were not considered in this section. In the studies from Group B, the revision of the IPT took three forms: (i) precision; (ii) rejection; (iii) addition of some features regarding *Contexts*, *Mechanisms* or *Outcomes*. The revisions were not necessarily illustrated by a new figure or table, but rather in words. We expected to be able to easily identify where revisions had been made, but this was not the case, nor were the arguments for revision fully explained. None of the included studies compared the IPT and RPT, and instead the results were presented as if they were new, without referring to the IPT. Only in one study (Owusu-Addo *et al.*, 2020) did the authors explain what brought them to make the adjustments for each revised CMO.

RE practices in the light of Rootman's (2001) four core features for the evaluation of HP interventions

Feature 1: Evaluation accommodates the complex nature of HP interventions

For most of the authors of the included studies, the relevance of RE resided largely in its suitability to the complex nature of HP interventions. One of the main contributions of RE to accommodating the complex nature of HP is based in the CMO configurations framework. The notion of '*Mechanism'* was, as expressed by several authors (Bergeron *et al.*, 2020; Deschesnes *et al.*, 2014; Owusu-Addo *et al.*, 2020; Pals *et al.*, 2016; Van der Veken *et al.*, 2020; Willis *et al.*, 2018), particularly helpful in understanding program-context interactions. Because CMO configurations are commonly identified as a tool of RE in order to clarify and capture the complexity of interventions, we examined how the authors of the selected articles had interpreted the explanatory components of the realist causation pattern (*Context*, *Mechanism* and *Outcome*) both as stand-alone elements and as parts of CMO configurations.

Context

The elements reported by the authors under the term '*Context'* described the physical and social environment (setting) in which the intervention took place. These elements could be related to individuals, interpersonal relationships, institutional characteristics or infrastructure, as defined by Pawson (2013). Such elements were pre-existing or external to the program, or internal, created by the program. In one article (Martin-Fernandez *et al.,* 2021), a distinction was made explicitly between the contextual elements linked to the intervention (named 'Ci') and those external to the intervention (named 'Ce').

Mechanisms

As with the elements of *Context*, the reported *Mechanisms* were situated at different levels: the individual level, the collective level, the institutional level and the level of the resources created by the intervention. The majority of *Mechanisms* referred to the individual level. The reported definitions of *Mechanism* are variable; some were quite precise while others were rather vague. In two studies, *Mechanisms* were defined according to their pivotal role in the relationship between *Context* and *Outcomes* (Darlington *et al.*, 2017; Darlington *et al.*, 2018). In seven studies, *Mechanisms* were defined more succinctly, by referring to them as forces that generate *Outcomes* (Deschesnes *et al.*, 2014; Mathias *et al.,* 2019; Pals *et al.,* 2016; Taylor *et al.*, 2010; Tennant *et al.*, 2020; Van der Veken *et al.*, 2020; Willis *et al.*, 2018). In five studies, *Mechanisms* were divided into resources (which the intervention provided) and responses (how people responded to these resources) (Bergeron *et al.*, 2020; Bysted *et al.*, 2020; Martin-Fernandez *et al.*, 2021; O'Rourke *et al.,* 2020; Owusu-Addo *et al.,* 2020) with reference to Pawson and Tilley definition of mechanisms (Pawson and Tilley, 1997). In one study, *Mechanisms* were defined as "how the programme was operationalised" (Howarth *et al.*, 2021). The authors of another study suggested looking at *Mechanisms* as processes rather than on/off

switches (Jonsson and Goicolea, 2020). In one last study (Uzochukwu *et al.*, 2020) there was no definition of *Mechanism*.

Outcomes

All of the included studies considered one or more *Outcomes*. In all but one study (Bergeron *et al.*, 2020), the *Outcomes* were quite specific to a program level and mostly related to personal resources or behaviours. Only in Bergeron et al.'s study (2019) was the *Outcome* expressed in terms of quite general health improvement (more specifically, dental health improvement). In a few studies (Bysted *et al.*, 2020; Owusu-Addo *et al.*, 2020; Van der Veken *et al.*, 2020; Willis *et al.*, 2018), *Outcomes* included intended and non-intended *Outcomes*. In one study (Howarth *et al.*, 2021), *Outcomes* were considered both for participants of the intervention and also for those who delivered the intervention. None of the reported *Outcomes* were related to physical environments.

CMO configurations

Although all the included studies set the RE approach with reference to the work of Pawson and Tilley, the ways in which authors presented CMO configurations differed from one another. In this section, we considered the CMO configurations comprising in the initial program theory of the studies from Group A and the CMO configurations comprising in the revised program theory of the studies from Group B.

From the five studies from Group A, two studies (Willis *et al.*, 2018; Taylor *et al.*, 2010) presented each of the C, M and O components separately in a descriptive and general way, while two studies (O'Rourke *et al.*, 2020; Jonsson and Goicolea, 2020) presented configured CMO components: one study had a single configuration (Jonsson and Goicolea, 2020) while the other had multiple configurations (O'Rourke *et al.*, 2020). In the remaining Group A study

(Darlington *et al.*, 2017), the distinction between C and M components was not clearly stated and consequently the context alone seems to explain the outcomes.

From the 12 studies from Group B, one study (Howarth *et al.*, 2021) had not presented its results by distinguishing the C, M and O components. The remaining 11 studies provided configured CMOs that were presented either as a unique configuration (Deschesnes *et al.*, 2014; Uzochukwu *et al.*, 2020; Van der Veken *et al.*, 2020; Mathias *et al.*, 2019; Pals *et al.*, 2016) or as multiple configurations (Owusu-Addo *et al.*, 2020; Bysted *et al.*, 2020; Bergeron *et al.*, 2020; Martin-Fernandez *et al.*, 2021; Darlington *et al.*, 2018; Tennant *et al.*, 2020). In the latter case, a few authors (Bergeron *et al.*, 2020; Darlington *et al.*, 2018; Tennant *et al.*, 2020; Bysted *et al.*, 2020) organised the CMO configurations into categories.

Feature 2: Evaluation draws on a variety of disciplines and a broad range of information-gathering procedures

Table 3 summarises the characteristics of the included studies in terms of study design, data collection methods and data analysis methods. A majority of the included studies were qualitative, and in all studies interviews were used to collect data in addition to one or more other data collection methods. Up to four different data collection methods were used. Data analysis methods were more diverse within the included data. Surprisingly, in the case of four studies from Group B (Bergeron *et al.*, 2019; Howarth *et al.*, 2021; Pals *et al.*, 2016; Tennant *et al.*, 2020), the authors did not explicitly use the IPT as a framework to guide the analysis.

By looking at the affiliations of the articles' authors, we note that six articles (Deschesnes *et al.*, 2014; Howarth *et al.*, 2021; Taylor *et al.*, 2010; Tennant *et al.*, 2020; Mathias *et al.*, 2019; Willis *et al.*, 2018) involved non-academic authors, and three of those involved more than two disciplines (Tennant *et al.*, 2020; Deschesnes *et al.*, 2014; Willis *et al.*, 2018). Another two articles (Owusu-Addo *et al.*, 2020; Uzochukwu *et al.*, 2020) involved more than two disciplines

without involving non-academic authors. The remaining 12 studies were co-authored by researchers only and involved less than two different disciplines.

In 12 studies, the data was coded and analysed by more than one person (Bergeron *et al.*, 2020; Howarth *et al.*, 2021; Martin-Fernandez *et al.*, 2021; Mathias *et al.*, 2019; O'Rourke *et al.*, 2020; Tennant *et al.*, 2020; Van der Veken *et al.*, 2020; Willis *et al.*, 2018; Pals *et al.*, 2016; Jonsson and Goicolea, 2020; Uzochukwu *et al.*, 2020; Darlington *et al.*, 2018). In two of these studies (Uzochukwu *et al.*, 2020; Willis *et al.*, 2018), the participants in the analysis seem to be affiliated with different disciplines and in one study (Mathias *et al.*, 2019) the analysis involved non-academics. Conversely, in three studies one person analysed the data (Darlington *et al.*, 2017; Owusu-Addo *et al.*, 2020; Bysted *et al.*, 2020) and in the two remaining studies (Deschesnes *et al.*, 2014; Taylor *et al.*, 2010) information about how many individuals had been involved in the analysis was not clearly stated.

Table 3: Study characteristics

Authors and year	Study methods		Data collection methods							Data analysis methods		
	Qualitative study	Mixed methods	Interviews	Focus Groups	Document analysis	Observations	Literature review	Informal discussions	Questionnaires	Inductive coding (no use of the IPT)	Framework analysis using the IPT	Alternating stages of inductive and deductive approaches
Group A												
Darlington *et al.,* 2017	■		■		■					■		
Jonsson and Goicolea, 2020	■		■		■						■	
O'Rourke *et al.,* 2019	■		■				■	■		■		
Taylor *et al.,* 2010	■		■	■	■						■	
Willis *et al.,* 2018	■		■	■		■					■	
Group B												
Bergeron *et al.,* 2020		■	■	■					■	■		
Darlington *et al.,* 2018		■	■		■				■		■	
Deschesnes *et al.,* 2014	■		■	■	■							■
Pals *et al.,* 2016		■	■			■			■	■		
Tennant *et al.,* 2020	■		■					■		■		
Owusu-Addo *et al.,* 2020	■		■	■		■						■
Van der Veken *et al.,* 2020	■		■	■	■	■					■	
Mathias *et al.,* 2019		■	■	■					■		■	
Uzochukwu *et al.,* 2020	■		■		■		■					■
Bysted *et al.,* 2020	■		■			■					■	
Howarth *et al.,* 2021	■		■	■						■		
Martin-Fernandez *et al.,* 2021	■		■	■		■						■

Feature 3: Evaluation involves stakeholders in appropriate ways

The way stakeholders contributed to the RE was identified in eight articles (Bergeron *et al.*, 2020; Darlington *et al.*, 2018; Martin-Fernandez *et al.*, 2021; O'Rourke *et al.*, 2020; Taylor *et al.*, 2010; Tennant *et al.*, 2020; Van der Veken *et al.*, 2020; Willis *et al.*, 2018). In these articles, stakeholders contributed in one or more stages of the evaluation process, but how their contribution was organised was not always detailed in the studies. In Taylor *et al*. (Taylor *et al.*, 2010), the members (ie. volunteers, managers, staff) of the service being evaluated were involved at each stage of the evaluation (identification of the research questions, identification of the information needed to answer them and how this information should be collected). This is also the only study where a sustainable 'feedback component' allowing researchers and stakeholders who wished to be involved in the evaluation process to reflect together during that process was reported. In three papers (Darlington *et al.*, 2018; Martin-Fernandez *et al.*, 2021; Tennant *et al.*, 2020), stakeholders were involved in the formulation of the IPT, and later in the revision of it. In Willis *et al.* (Willis *et al.*, 2018), one member of the organisation contributed in almost every step of the evaluation process (except for analysis). In O'Rourke *et al.* (O'Rourke *et al.*, 2020), a reference group composed of members of the organisation subject to the experiment was created to help build the IPT. In Bergeron *et al.* (Bergeron *et al.*, 2020), stakeholders validated the CMO configurations, and in that of Van der Veken *et al.* (Van der Veken *et al.*, 2020) they discussed the initial hypotheses in a focus group.

Feature 4: Evaluation builds the capacity of individuals, communities, organisations and governments in addressing health promotion concerns

As concluded in the previous section, in none of the included studies did the authors explicitly acknowledge any intention to distribute 'power' and thus reinforce the capacity of stakeholders to conduct evaluations and make use of evaluation results. However, some of the participative processes detailed above, such as in the case of Taylor *et al.* (2010) or in the articles co-authored

with non-academics, may have allowed the stakeholders involved to develop competencies, strengthen reflective practices or increase the use of evaluation results for instance. This suggests that in some studies the evaluation process may have contributed to reinforcing the capacity of stakeholders to carry out the evaluation and undertake HP interventions. However, this was mostly unintentional and has not been identified or explicitly acknowledged in the papers.

Discussion

Health promotion is in need of evaluation that reflects its characteristics, values and principles. In this paper, we sought to provide an overview of the current use of RE in the field of HP. We analysed the practices of RE reported in 17 articles, with respect to Rootman et al.'s (Rootman *et al.*, 2001) four core features of approaches appropriate for the evaluation of HP interventions. Of the included studies, we identified two types: those aiming to build an IPT; and those aiming to test an IPT, leading to the formulation of a RPT. The way RE captures the complex nature of HP interventions is rooted in its CMO framework. Our results revealed a great variety of practices, both from 'configured CMO' to 'unconfigured or disconnected CMO', and from global and descriptive CMO configurations to specific and explanatory ones. Our examination of the data collection and analysis methods points to some common practices in the use of multi-methods – particularly qualitative methods. Authors' affiliation and contributions point to interdisciplinarity and transdisciplinarity approaches (Tress *et al*. 2004). Yet, these approaches did not guide the analysis of data, though it generally involved multiple persons. Although scarce, or at least not set out by authors, there were various types of participative practices. Such participative practices aimed to reinforce data validity rather than stakeholder capacity. Capacity-building through evaluation might have unintentionally taken place in the studies included.

Before discussing our results, we feel the need to clarify that by exploring authors' practices of RE. We did not seek to judge whether or not the included studies are 'really' realist or not – as Pawson wrote "[…] this is a tough question and arguably the wrong question to ask" (Pawson, 2013, p13) – but rather to learn from their experience of applying RE in the field of HP, and reflect on its contribution to HP evaluation practices. The narrative account set out through our analysis of the current scientific literature aimed to identify perspectives on more relevant evaluation practices in the field of HP. In addition, our findings might also shed light on more meaningful practices of RE.

Perspectives on more relevant evaluation practices of HP effectiveness

RE puts emphasis on the diversity of people, of their behaviors, understandings, experiences – and on the importance of processes. We believe it is quite logical to work with an evaluative approach that recognises, as HP practitioners do, the complexity of reality and the importance of processes. Our findings reveal two ways in which this emphasis is perceptible when RE is applied to the evaluation of HP interventions: (i) in its way of building an IPT before collecting data, and (ii) in its understanding of causality within the CMO configurations framework.

By aiming to clarify the implicit set of assumptions in a program, RE contributes to the evaluability assessment of such a program which, in turn, has the potential to create a meaningful outcome and impact evaluation (Leviton *et al.*, 2010). Some would argue that this initial step is not specific to RE, and they would be right; but not every evaluative approach includes the assumption of evaluability in its methodology. In the case of RE, program theory is at the centre of evaluation: it helps identify the data to be collected, build the data collection tools, analyse the data and use the results by revising the program theory. RE does not just attempt to clarify the program theory; rather, it is at the core of the evaluative process. We believe HP could benefit from putting theory at the centre of the evaluation process, as this

allows the generation of meaningful results that may be used to guide decisions on action, and therefore program improvement. This perspective on the evaluation of HP was recently supported by Cambon and Alla (Cambon and Alla, 2021), who suggested developing a context-dependent theory (called Intervention System Theory) to guide evaluation design.

The emphasis on people and processes is also reflected in the use of CMO configurations in order to understand causation. The articulation between the three components of realist causation has great potential for increasing the 'readability' of HP interventions and for emphasising the non-linearity of causality in complex interventions. The role of *Context* has always been difficult to define in causality – and RE offers a promising way of defining the explanatory role of *Context*. However, the notion of *Mechanism* could be counter-productive if our understanding of causal mechanism and our use of program theory do not represent the role of the multiple determinants of health. We explore this idea further in the next section.

Perspectives for more meaningful practices of RE

In order to clarify the theory on which interventions are based, RE methods suggest that stakeholders explain their assumptions about how they think the program works, why, for whom and in what circumstances. To that aim, our results show that the practice of RE in the 17 included studies focusses mainly on multi-methods and, to a lesser extent, on participative processes. The participative practices reported in our selection of articles were more about collecting different points of view in order to produce a solid methodology than about building a common understanding of the program by cross-referencing visions and experiences. The nuance between these two types of participative approaches lies mainly in the desire to produce either knowledge or social transformation. In order to better meet HP principles for evaluation, we believe RE practices in the field of HP would benefit from being applied with transparent

participative intentions, and from sharing the participative components within scientific articles.

Our findings also show that most of the CMO configurations were related to personal resources or behavioural outcomes, and only a very few to social, organisational or policy dimensions. A possible, but incomplete, explanation for this trend is that many of the interventions evaluated in the included studies aimed at developing personal skills (cf. Table 2), a strategy that targets personal resources and behaviours by nature. This explanation is incomplete because, in HP, making sure that the evaluation approach is able to recognise the interplay of determinants of health is fundamental. However, with regards to our results, the use of RE failed to emphasise important determinants of health, other than the personal resources and behavioural determinant. This result is consistent with a chapter written by Gill Westhorp who pleads for and provides "ways of conceptualising mechanisms that can be applied at other levels of systems" (Westhorp, 2018 p.56). We believe that future practices of RE in the field of HP should look to extending the range of CMO configurations to other levels of systems beyond the individual.

By looking for evidence of participation and capacity-building, our findings emphasise a lack of consideration of the social processes of evaluation in current RE practices. Indeed, even in the only article that does demonstrate the benefits of practicing RE in a participative and empowerment-oriented approach, these benefits, do not relate to the social processes of the approach. Currently, the scientific literature focusses more on the technical aspects of realist evaluation, and fails to provide insight to the 'social practice' of realist evaluation. This result echoes the work of Abma and Widdershoven (Abma and Widdershoven, 2008) who, more than ten years ago, already called for more attention to be paid to the social relationships between stakeholders in the evaluation setting. We believe researchers and/or evaluators involved in RE should: acknowledge the importance of creating a real dialogue between stakeholders; think of

the conditions for creating this dialogue; practice RE in a more participative and empowering way; and report more systematically how, why and in what circumstances the RE setting empowered stakeholders (or did not) in addressing contemporary issues.

Strenghts and limitations

This study is limited by the use of the unique term ‘health promotion’ in relation to RE in the search strategy. Indeed, our experience suggests that many interventions may promote health in the way we have defined it, without explicitly referring to this public health approach.

Our analysis is based on the information available in the included studies. However, because as scientific publications tend to focus on results rather than on processes, our analysis is limited to what was explicitly set out by the authors in their articles, which might not be a full account of what and how things were done. A way to overcome this limitation could have been to check the full reports, if only referenced and accessible.

We decided not to assess the studies against the RAMASES standards (Wong *et al.*, 2016) to determine whether the studies were REs. This methodological choice could be considered a limitation because some of the included studies would possibly have been rejected. However, this choice is consistent with our research aim which is to produce an overview of how a RE is applied to evaluate HP interventions. From this perspective we were also interested in the adaptations of the realist approach in order to align the methods with HP principles.

Among the strengths of our study are: interdisciplinary research; triangulation of sources, and of researchers’ perspectives; use of thematic framework analysis; and traceability of the selection and coding process.

Conclusion

We intended to make recommendations based on research experiences that would inform how to evaluate HP interventions using the RE approach, whilst fulfilling Rootman *et al*'s (Rootman *et al.*, 2001) four core features of an approach appropriate for the evaluation of HP interventions. The narrative account of realist evaluation practices in the scientific literature helped us to reach this objective. The use of the four principles for HP evaluation proved helpful for examining the current practices of RE through a HP lens while identifying perspectives that may guide future practices in both HP and RE.

Keywords

Program evaluation, empowerment, capacity building, review, collaboration

Abbreviations

CMO: Context Mechanism Outcome; HP: Health Promotion; IPT: Initial Program Theory; RPT: Revised Program Theory; RE: Realist Evaluation.

Ethical approval

Not applicable.

Author contributions

S.M. and I.A. were involved in the design of the scoping review. B.S., D.D., I.A. and S.M. were involved in data collection and in building the data charting form. S.M. analysed data and wrote the manuscript. All authors critically revised the manuscript and agreed on final approval for the present version to be published.

Funding

This study was supported by the Brussels' Strategic Plan for Health Promotion, directed by the French Community Commission (COCOF), a Regional representative of French-speaking authorities in the Brussels-Capital Region.

Conflict of interest

There is no conflict of interest.

Author details

[1]University centre for Health Promotion RESO, Institute of Health and Society, Université catholique de Louvain, Brussels, Belgium

Clos chapelle-aux-champs 30, B1.30.14

1200 Woluwé-St-Lambert

[2]Institute of Health and Society, Université catholique de Louvain, Brussels, Belgium

Clos chapelle-aux-champs 30, B1.30.13

1200 Woluwé-St-Lambert

Caption list

Table 1: Key principles and strategies for defining an intervention as a HP intervention

Table 2: Characteristics of the evaluated programs

Table 3: Study characteristics

Figure 1: The PRISMA flow diagram of the selection process

List of supplementary materials

Supplementary file 1: Search strategy for each database

Supplementary file 2: List of all variables for which data were sought

References

Abma, T. A. and Widdershoven, G. A. M. (2008) Evaluation and/as Social Relation. *Evaluation*, 14, 209-225.

Arksey, H. and O'Malley, L. (2005) Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8, 19-32.

Bergeron, D. A., Talbot, L. R. and Gaboury, I. (2020) Intersectoral oral health promotion interventions for schoolchildren living in remote rural Andean communities: a realist evaluation. *Global Health Promotion*, 27, 103-112.

Bisset, S., Potvin, L. and Daniel, M. (2013) The adaptive nature of implementation practice: Case study of a school-based nutrition education intervention. *Evaluation and Program Planning*, 39, 10-18.

Blaise, P. (2004) Culture qualité et organisation bureaucratique, le défi du changement dans les systèmes publics de santé. Une évaluation réaliste de projets qualité en Afrique., *Ecole de Santé Publique - Université Libre de Bruxelles*.

Bysted, S., Overgaard, C., Simpson, S. A., Curtis, T. and Bøggild, H. (2020) Recruiting residents from disadvantaged neighbourhoods for community-based health promotion and disease prevention services in Denmark—How, why and under what circumstances does an active door-to-door recruitment approach work? *Health and Social Care in the Community*, 1-12.

Cambon, L., Terral, P. and Alla, F. (2019) From intervention to interventional system: towards greater theorization in population health intervention research. *BMC Public Health*, 19, 339.

Cambon, L. and Alla, F. (2021) Understanding the complexity of population health interventions: assessing intervention system theory (ISyT). *Health Research Policy and Systems*, 19, 95.

Darlington, E. J., Simar, C. and Jourdan, D. (2017) Implementation of a health promotion programme: a ten-year retrospective study. *Health Education*, 117, 252-279.

Darlington, E. J., Violon, N. and Jourdan, D. (2018) Implementation of health promotion programmes in schools: An approach to understand the influence of contextual factors on the process? B*MC Public Health*, 18.

Deschesnes, M., Drouin N., Tessier, C. and Couturier, Y. (2014) Schools' capacity to absorb a Healthy School approach into their operations: Insights from a realist evaluation. *Health Education*, 114(3), 208-224.

Greenhalgh, T. and Papoutsi, C. (2018) Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med.*, 16, 95.

Howarth, M., Cormac, L. and da Silva, A. (2021) Creating a transformative space for change: A qualitative evaluation of the RHS Wellbeing Programme for people with long term conditions. *Health & Place*, 71, 102654.

Jolley, G. (2014) Evaluating complex community-based health promotion: Addressing the challenges. *Evaluation and Program Planning*, 45, 71-81.

Jonsson, F. and Goicolea, I. (2020) "We believe in you, like really believe in you": Initiating a realist study of (re)engagement initiatives for youth not in employment, education or training with experiences from northern Sweden. *Evaluation and Program Planning*, 83, 101851.

Lacouture, A., Breton, E., Guichard, A. and Ridde, V. (2015) The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implement Sci*, 10, 153.

Leviton, L. C., Kettel Khan, L., Rog, D., Dawkins, N. and Cotton, D. (2010) Evaluability Assessment to Improve Public Health Policies, Programs, and Practices. *Annual Review of Public Health*, 31, 213-233.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. and Kegels, G. (2012) Is realist evaluation keeping its promise ? A review of published empirical studies in the field of health systems research. *Evaluation*, 18, 192-212.

Martin-Fernandez, J., Aromatario, O., Prigent, O., Porcherie, M., Ridde, V. and Cambon, L. (2021) Evaluation of a knowledge translation strategy to improve policymaking and practices in health promotion and disease prevention setting in French regions: TC-REG, a realist study. *BMJ Open*, 11.

Mathias, K., Singh, P., Butcher, N., Grills, N., Srinivasan, V., and Kermode, M. (2019) Promoting social inclusion for young people affected by psycho-social disability in India - a realist evaluation of a pilot intervention. *Glob Public Health*, 14, 1718-1732.

Moore, G.F., Evans, R.E., Hawkins, J., Littlecott, H., Melendez-Torres G.J., Bonell, C., *et al.* (2019) From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation*, 25(1), 23-45.

O'Rourke, K. M., Yelland, J., Newton, M. and Shafiei, T. (2020) An Australian doula program for socially disadvantaged women: Developing realist evaluation theories. *Women and Birth*, 33, e438-e446.

Owusu-Addo, E., Renzaho, A. M. N. and Smith, B. J. (2020) Developing a middle-range theory to explain how cash transfers work to tackle the social determinants of health: A realist case study. *World Development*, 130, 104920.

Pals, R. A., Olesen, K. and Willaing, I. (2016) What does theory-driven evaluation add to the analysis of self-reported outcomes of diabetes education? A comparative realist evaluation of a participatory patient education approach. *Patient Educ Couns*, 99, 995-1001.

Pawson, R. (2013) The science of evaluation : a realist manifesto. *SAGE Publications*, London, United Kingdom.

Pawson, R. and Tilley, N. (1997) Realistic Evaluation. *SAGE Publications*, London, United Kingdom.

Potvin, L., McQueen, D.V. and Hall, M. (2009) Introduction. Aligning Evaluation Research and Health Promotion Values: Practices from the Americas. In Potvin, L., McQueen, D.V., Hall, M., de Salazar, L., Anderson, L.M. and Hartz Z.M.A. (eds) Health Promotion Evaluation Practices in the Americas: Values and Research. Springer, New York, pp. 1-9.

Rootman, I., Goodstadt, M., Potvin, L. and Springett, J. (2001) A framework for health promotion evaluation In Rootman, I., Goodstadt, M., Hyndman, H., McQueen, D.V., Potvin, L., Springett, J. and Ziglio, E. Evaluation in health promotion. Principles and perspectives. World Health Organisation Regional Publications, European Series, 92, pp. 3-38.

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M. *et al.* (2021) A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance *BMJ*; 374, 2061.

Taylor, J., Jones, R. M., O'Reilly, P., Oldfield, W. and Blackburn, A. (2010) The Station Community Mental Health Centre Inc: nurturing and empowering. *Rural Remote Health*, 10, 1411.

Tennant, E., Miller, E., Costantino, K., De Souza, D., Coupland, H., Fotheringham, P. *et al.* (2020) A critical realist evaluation of an integrated care project for vulnerable families in Sydney, Australia. *BMC Health Services Research*, 20, 995.

Tress, G., Tress, B. and Fry, G. (2004) Clarifying integrative research concepts in landscape ecology. *Landscape Ecology*, 20, 479-493.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., *et al.* (2018) PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169, 467-473.

Uzochukwu, B., Onyedinma, C., Okeke, C., Onwujekwe, O., Manzano, A., Ebenso, B., *et al.* (2020) What makes advocacy work? Stakeholders' voices and insights from prioritisation of maternal and child health programme in Nigeria. *BMC Health Services Research,* 20, 884.

Van der Veken, K., Lauwerier, E. and Willems, S. (2020) "To mean something to someone": sport-for-development as a lever for social inclusion. *International Journal for Equity in Health*, 19, 11.

Westhorp, G. (2018) Understanding mechanisms in realist evaluation and research In Emmel, N., Greenhalgh, J., Manzano, A., Monaghan, M. and Dalkin, S. Doing realist research, SAGE Publications, pp. 44-57.

Willis, C. E., Reid, S., Elliott, C., Rosenberg, M., Nyquist, A., Jahnsen, R., *et al.* (2018) A realist evaluation of a physical activity participation intervention for children and youth with disabilities: What works, for whom, in what circumstances, and how? *BMC Pediatrics*, 18.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T. (2016) RAMESES II reporting standards for realist evaluations. *BMC Med* 14, 96.