

Methods in realist evaluation: a mapping review

Dimitri Renmans^{a,b}, Virginia Castellano Pleguezuelo^c

<https://doi.org/10.1016/j.evalprogplan.2022.102209>

a Ecole de Santé Publique, Université Libre de Bruxelles, Route du Lennik 808, 1070 Brussels, Belgium

b Institute of Development Policy (IOB), University of Antwerp, Lange Sint-Annastraat 7, 2000 Antwerp, Belgium

c Department of Public Health, Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium

Abstract

Realist evaluation is becoming increasingly popular as an evaluation methodology. Its main objective is to uncover the mechanisms that lead to observed outcomes following an intervention and the contextual conditions that enabled this. The focus is on explaining why, for whom and in what circumstances an intervention works. It is a theory-driven approach and is explicitly method neutral, meaning that both quantitative and qualitative data collection methods can be used to unearth the underlying mechanisms that cause the intervention outcomes. In this review, we aim to map the methods used in realist evaluation studies, to draw lessons from the findings and to reflect on ways forward. We found that qualitative methods and interviews specifically are most commonly used in realist evaluations; that theory is often absent behind the methods and sampling techniques used; and that more innovative methods remain underexplored. We conclude the review by proposing four ways forward: (1) developing realist surveys, (2) exploring the relevance of innovative methods, (3) increasing the attention paid to sampling procedures and (4) strengthening the theory-driven nature of method. We believe that these four action points can strengthen the practice of realist evaluation and its outcomes.

Keywords: Realist evaluation, data collection methods, interview, observation, surveys, theory-driven evaluation

Abbreviations: RE = realist evaluation; CMOC = context-mechanism-outcome configuration; II = individual interviews; GI = group interviews

Introduction

Several authors have observed a shift in the evaluation landscape towards an increased focus on and interest in mechanisms and the underlying processes of interventions that lead to (un)desirable outcomes (Brousselle & Buregeya, 2018; Lemire, Peck, & Porowski, 2020). Many of these approaches (qualitative comparative analysis, contribution analysis, realist evaluation) are at least to some extent underpinned by scientific/critical realism (Brousselle & Buregeya, 2018; Rutten, 2021), a philosophy

of science that maintains a generative view on causation (Bhaskar, 2008 [1975]; Pawson, 2013; Sayer, 1992). Simply put, this approach considers that interventions lead to specific outcomes in specific contexts because of the underlying generative mechanisms that are triggered. One of the approaches in this paradigm shift is realist evaluation (RE), developed by Pawson and Tilley (1997).

RE is strongly grounded in the realist idea of generative mechanisms and their context-contingent nature (Pawson, 2013). While there is much debate on the definition of a mechanism (Lacouture, Breton, Guichard, & Ridde, 2015; Lemire, Kwako, et al., 2020), the following definition by George and Bennet offers a good starting point, especially if the word 'processes' is changed into the word 'powers': 'unobservable physical, social, or psychological processes through which agents with causal capacities operate, but only in specific contexts or conditions, to transfer energy, information, or matter to other entities,' thereby changing the latter entities' 'characteristics, capacities, or propensities in ways that persist until subsequent causal mechanisms act upon it' (George and Bennet, 2005, cited in Bennett (2013, p. 466).

RE's main objective is to uncover the mechanisms that lead to observed outcomes following an intervention and the contextual conditions that enabled this. The focus is on explaining why, for whom and in what circumstances an intervention works. Similar to most theory-based evaluations, RE starts by eliciting an initial programme theory (or even multiple programme theories). Specifically, for RE, this programme theory is built around a specific heuristic called the Context-Mechanism-Outcome Configuration (CMOC). The latter depicts the central form of causal reasoning within RE (Pawson, 2013).

In a second step, RE sets out to collect empirical data to refine these CMOCs. Data collection is method neutral (Pawson & Tilley, 1997; Sayer, 1992) and can thus be qualitative or quantitative depending on the evidence needed to refine specific parts of the CMOCs. The data is subsequently analysed using the CMOC as a guide. In the final step, the initial programme theory is refined and updated. This process is iterative and entails going back and forth between theory (the first phase) and the data collection and analysis phase whenever new mechanisms are uncovered.

As this review and other scoping reviews show (Lemire, Kwako, et al., 2020; Nielsen, Lemire, & Tangsig, 2022), RE is becoming increasingly popular among evaluators. However, its application does not come without challenges. Several reviews have highlighted difficulties with several elements, including identifying mechanisms, differentiating between mechanisms and contextual factors, and uncovering the most relevant contextual factors (Lemire, Kwako, et al., 2020; Marchal, van Belle, van Olmen, Hoerée, & Kegels, 2012; Nielsen, et al., 2022). While this is partly due to conceptual issues, and conceptual clarification may help us in this respect, it is clear that the methods we use to unearth

mechanisms and relevant contextual conditions are equally important. Therefore, the objective of this review is to map out the methods used in the RE literature, to draw lessons from the findings and to reflect on ways forward.

Methodology

We performed a mapping of the literature to see which methods were being used within published peer-reviewed realist evaluations. A mapping review aims ‘to map out and categorize existing literature on a particular topic’ (Grant & Booth, 2009, p. 97). In this case, the topic was the use of methods within realist evaluations. The analysis aimed to obtain a broad view of the methods used and offer insight into how to improve the use of methods within realist evaluations.

Search process

We searched five different databases (Scopus, Web of Science, Global Health [Ovid], EMBASE [Ovid], and MEDLINE [Ovid]) using the search terms presented in Table 1. The search was performed on 1 February 2021 for the EMBASE, MEDLINE and Global Health databases and on 6 February 2021 for the Scopus and Web of Science databases.

Table 1: Search terms used in the five databases

Databases	Search strategy
Medline (Ovid)	(realist* or realism) ADJ4 (evaluat* or inquir* or research* or lens or approach* or perspective* or review* or synthes* or RAMESES)
Embase (Ovid)	(realist* or realism) ADJ4 (evaluat* or inquir* or research* or lens or approach* or perspective* or review* or synthes* or RAMESES)
Global Health (Ovid)	(realist* or realism) ADJ4 (evaluat* or inquir* or research* or lens or approach* or perspective* or review* or synthes* or RAMESES)
Web Of Science	(realist* or realism) NEAR/4 (evaluat* or inquir* or research* or lens or approach* or perspective* or review* or synthes* or RAMESES)
Scopus	TITLE-ABS-KEY ((realist* OR realism) PRE/4 (evaluat* OR inquir* OR research* OR lens OR approach* OR perspective* or review* or synthes* or RAMESES)

The inclusion criteria were, first, peer-reviewed journal article published between 1997 and 2021, with a full text, self-identified as RE and at least referencing some seminal works on RE (e.g. Pawson and Tilley (1997); Wong, et al. (2016)). Conference abstracts, papers only presenting an initial programme theory, commentaries, book reviews, realist reviews, non-English papers, conceptual papers, papers on realist research, protocols and grey literature were not included. Second, because we were

specifically interested in the methods used to unearth mechanisms and the related contextual factors and outcome, the other selection criterion was the explicit reporting of the context, mechanisms and outcomes in the results or discussion sections.

Analysis

For each RE paper selected, the data collection method used was coded inductively. We then grouped these codes into four groups of methods as can be seen in Figure 1. These groups largely reflect common typologies that differentiate between interviews, observations and surveys (see for example Esterberg, 2002; Green & Thorogood, 2009). However, we included documentary review, or 'documentary observation' as we call it – which is often seen as a separate category – in the group of observation (see e.g. Esterberg (2002). This is justified by the idea that observation entails the collection of direct information without interpretation by stakeholders. This is an assumption that, to a large extent, also holds for management documents, meeting minutes, monitoring data and health care records, all labelled 'documentary observation' in this review. We also added the category of 'innovative methods' to capture less common methods not included in the other categories. Whether they are truly innovative can always be debated.

The coded data was then entered into a data matrix and mapped descriptively. During the mapping, we carried out new data extraction whenever this seemed relevant. For example, after having identified which REs used interviews, we decided to look at who the respondents were. The final coding tree can be found in Figure 1. The studies and associated codes can be found in the supplementary materials.

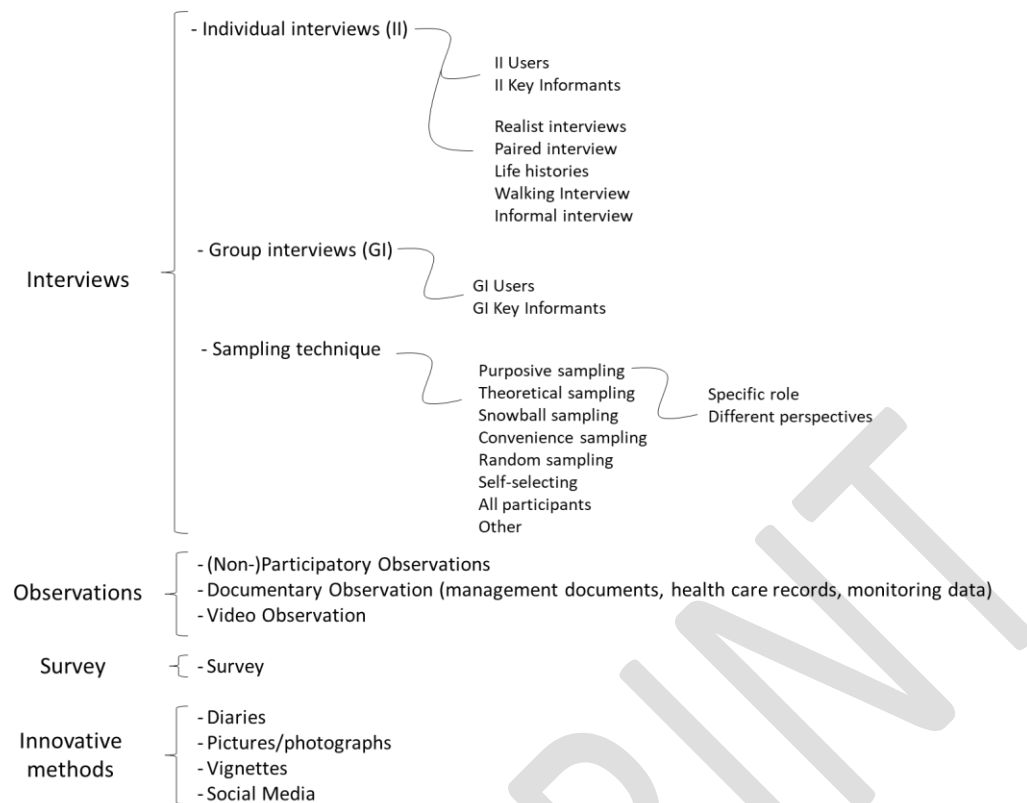


Figure 1: Coding tree of the mapping review

Limitations

This mapping review has some important limitations that need to be considered. First, the choice of databases, EMBASE and MEDLINE in particular, may have induced a bias towards studies from the health care sector. However, there is no strong reason to claim that this may have shifted the results in a certain direction. Second, limiting the selection to papers that explicitly mentioned the context, mechanism and outcomes may be seen as restrictive. However, we believe that the CMO configuration (and its adaptations) is the signature feature of RE, which means that without it the RE nature of the study may not always be as obvious. Finally, not all REs are published in the peer-reviewed literature. In fact, many remain unpublished due to time constraints of the evaluators, or contractual constraints put in place by the commissioner of the RE. Therefore, many interesting and valuable RE have not been incorporated into this mapping review. Again, we have no reason to believe that this would drastically change the overall conclusion of our study.

Findings

After the deletion of duplicates, we identified 6,446 publications across the different databases (see Figure 2). Of these, 283 publications were found to be REs, of which 167 reported explicitly about the relevant contexts, mechanisms and outcomes. Of these, 102 papers made CMO configurations that

explicitly linked the context, mechanism and outcome to each other and only included one mechanism, which can be seen as an ‘advised’ practice (Pawson & Manzano-Santaella, 2012). However, this does not mean that they were necessarily of good quality. Indeed, as observed by Lemire, Kwako, et al. (2020) many of the mechanisms were not such in the realist sense (The RAMESES II project, 2017) but programme components or activities. In our reporting of the findings, we did not distinguish between those who adhered to the ‘advised’ practices and those who did not, because the results were similar for the two groups and we did not want to give the impression of an established causal relationship in either direction between the methods used and the use of the CMO configuration. After applying these inclusion-exclusion criteria, 166 studies were descriptively analysed.

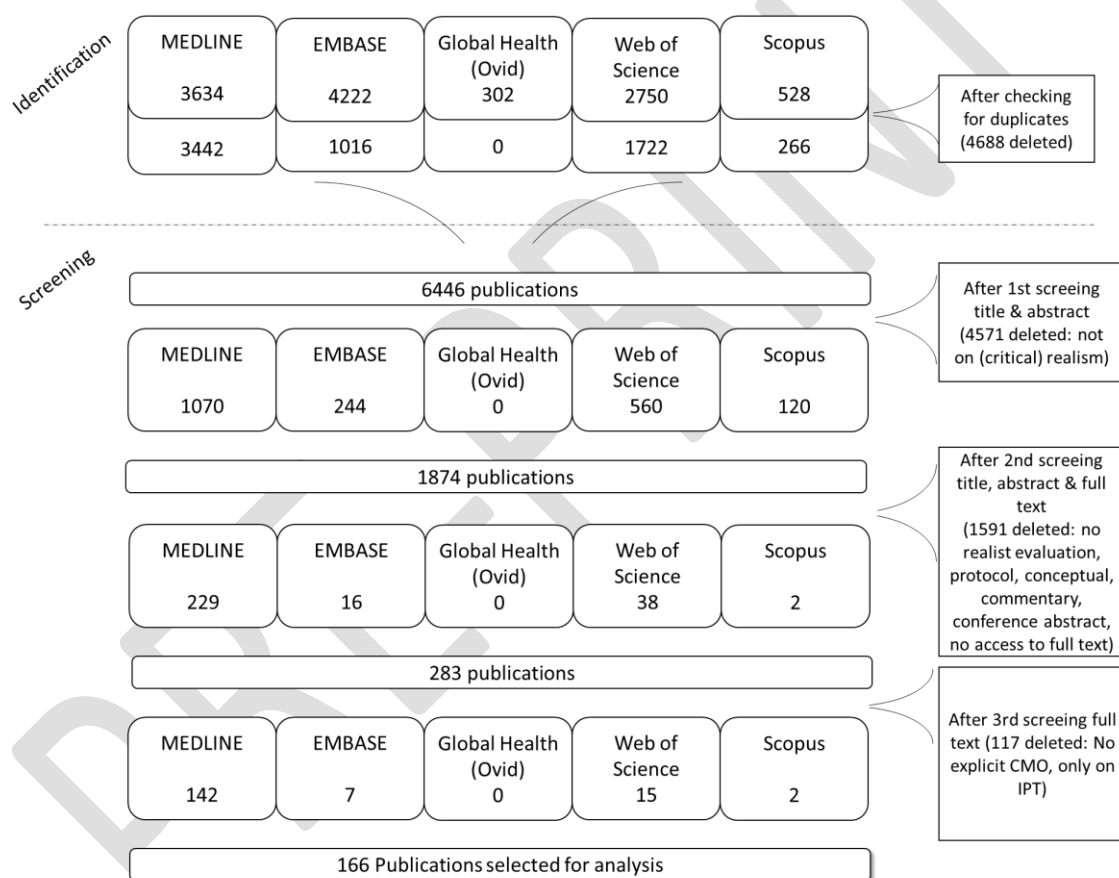


Figure 2: Search outcomes and selection process of the mapping review

The first interesting finding was that RE has become increasingly popular since the beginning of this century, with a remarkably sharp rise during recent years (see Figure 3). This increased popularity comes with an increased need for methodological guidance and understanding of how REs are performed. Hence, the importance of the current mapping review of the use of methods within RE.

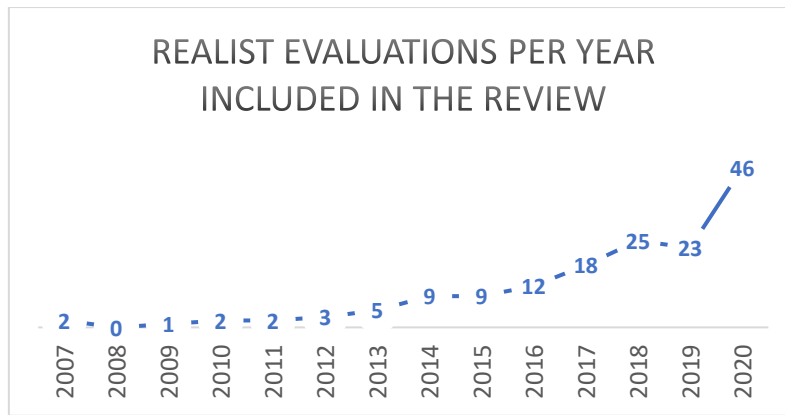


Figure 3: RE studies included per year

Table 2 shows that almost all the REs included some kind of interviewing, either individual interviews or group interviews. Observations through videos, (non-)participatory observation and documentary observation (monitoring data, policy and management documents, health care records) occurred in more than half of the REs. Surveys were used much less frequently, with only one quarter of the studies applying them. More innovative data collection methods, such as the use of social media, pictures, diaries or vignettes, were used in only 8% of the studies.

Table 2: General data collection methods used in selected REs

Method	Number (Total = 166)	%
Interview	161	97%
Observation	92	55%
Survey	43	26%
Innovative methods	13	8%

RE is not only method neutral but also advocates the use of different methods to triangulate data and strengthen claims about the underlying programme theory. Table 3 shows that one third of the studies relied only on interviews, while the large majority (65%) of the REs combined different methods. In our sample, one quarter of the REs combined interviews with the survey method, hence quantitative with qualitative methods. Overall, the data showed that, in general, REs adhere to the best practice of combining different data collection methods, although a significant number of REs (34%) relied on single-method approaches, which may limit, from a realist perspective, the depth of the causal claims they make.

Table 3: Combination of methods used in selected REs

Only interviews:				53 (32%)
Interviews combined with ... :				108 (65%)
<i>AND</i>	Survey	Observation	Innovative	
Survey	13 (8%)	25 (15%)	0 (0%)	
Observation		58 (35%)	6 (4%)	
Innovative			5 (3%)	
Surv. & Obs.			1 (1%)	
No interviews:				5 (3%)
	<i>Survey</i>	<i>Survey + Inn.</i>	<i>Obs.+ Inn.</i>	
	3 (2%)	1 (1%)	1 (1%)	

* All percentages are in relation to the total sample of 166 studies and rounded.

Interviews

As almost all of the REs applied some form of interviewing, it is useful to take a closer look at this element. Of the studies using interviews, 96% used individual interviews, with group interviews less common but still appearing in one third of the studies (Table 4). A particular in-depth interviewing technique related to RE is known as ‘realist interviewing’ (Manzano, 2016; Mukumbang, Marchal, Van Belle, & van Wyk, 2020; Pawson, 1996). During realist interviews, the evaluator and the respondent enter into a teacher-learner cycle in which the evaluator first explains their concepts and programme theory, after which the respondent takes on the role of teacher and explains to what extent the evaluator’s theory is correct or needs to be adapted (Pawson, 1996). What makes this approach distinct from other interviewing approaches is its openness with respect to the initial ideas, assumptions and hypotheses of the researcher and the focus on the construction of a programme theory (Mukumbang, et al., 2020). Despite its closeness to RE, a remarkably low number of evaluations (less than one in five) used the realist interviewing approach. The most commonly used techniques were the more classic semi-structured or in-depth interview, with other techniques such as life histories (Richardson, Phillips, Colom, & Nichols, 2019), walking interviews and paired interviews (Whitelaw, Gibson, Wild, Hall, & Molloy, 2021) being less frequently used. One third of the studies using interviews employed group interviews (focus group discussions or consensus panels), while a small number used informal interviews.

Table 4: Interviewing methods used in selected REs

Method <i>(Interviews)</i>	Number (n = 161)	%
Individual interviews (II)	155	96%
<i>Realist Interviews</i>	28	18%
<i>Life history, informal, paired or walking int.</i>	9	6%
<i>Semi-structured or in-depth interview</i>	118	76%
Group interviews (Focus group discussions, consensus panels)	56	35%
Informal Interviews	7	4%

Because who you interview is as important as how you interview, we also analysed which respondents were included in the interviews and found that almost half of the studies used both key informants (policymakers, implementers, service deliverers, etc.) and users and/or beneficiaries as interview respondents (Table 5). This concurs with the idea of having multiple perspectives included in an RE. About one third of the studies only relied on key informants, that is, those who were actively involved in the implementation of the intervention, while one in five only gathered interview data from users and/or beneficiaries.

Table 5: Overview of the interview participants

Interview participants	Number (n = 161)	%
Only key informants	51	32%
Only users and/or beneficiaries	33	20%
Both	77	48%

‘Purposive sampling’ was predominantly used as the sampling technique to select the respondents. Other common sampling techniques such as convenience sampling, snowballing or self-selection were much less frequently used. Of those studies that used purposive sampling, seven out of ten did so in order to select those respondents that had a specific role in the intervention, enabling them to obtain relevant information about how the intervention worked. Half of the studies using purposive sampling attempted to create a sample with a variety of perspectives. Often, the two rationales were used at the same time. Again, notwithstanding REs theory-driven nature, only 5% of the studies using purposive sampling gave theoretical reasons for this (also called realist or theoretical sampling). The purposive selection was sometimes done by the researchers themselves; however, often key stakeholders, such as managers or health workers, participated in the selection process (e.g. Hernandez, Hurtig, Dahlblom, & San Sebastian, 2014; Wye, et al., 2014). In one out of ten studies, the sampling technique used was unclear.

Table 6: Sampling techniques used

Sampling technique (Interviews)	Number (n = 161)	%
Purposive	111	69%
Different perspectives	58	52%
Specific role	80	72%
Theoretical	6	5%
Random	6	4%
Convenience	13	8%
Snowball	10	6%
Self-selecting	9	6%
All participants	26	16%
Unclear	18	11%

Observation

More than half of the studies used some form of observation (Table 7). Of these, three out of five used participatory or non-participatory observations and half of these used the review of documents, including annual reports, progress reports, newspaper articles, emails, meeting minutes, etc. Whereas these documents are often used as a way to analyse the context (e.g. Efstathiou et al., 2020; Reddy, Carey, & Wakerman, 2016), routinely collected quantitative data were more often used to collect data on outcomes (Dalkin, Lhussier, Jones, Phillipson, & Cunningham, 2018; Vugts, Zedlitz, Joosen, & Vrijhoef, 2020).

Table 7: Different forms of observation used in the studies covered by the mapping review

Method (Observations)	Number (Total = 92)	% (n = 92)
(Non-) Participatory Observation	57	61%
Documentary observation	42	46%
Monitoring data	29	32%
Health Care Records Audit	2	2%
Video Observation	1	1%

Surveys

One quarter of the evaluations used the survey method. These included standardized, established or validated tests used to analyse the context and make categorizations (e.g. Spitzer-Shohat, et al., 2018), to follow up on outcomes (e.g. Desveaux, et al., 2018) or to describe beliefs that may signal mechanisms (e.g. Martin & Tannenbaum, 2017). Other studies used less standardized self-created surveys to elicit views on the different elements of the CMO configuration (Oroviogicoechea & Watson, 2009). Once data had been collected, studies either used a thorough statistical analytical approach (e.g. Ford, et al., 2018; Oroviogicoechea & Watson, 2009) or used the data in a more descriptive way (Bertotti, Frostick, Hutt, Sohanpal, & Carnes, 2018).

Innovative data collection methods

One out of ten studies used more innovative and less frequently employed methods to collect data (Table 8). These were new, advanced and/or creative methods that were often employed alongside more commonly used methods such as interviews, surveys and observations.

Table 8: Less commonly used and innovative data collection methods

Method	Description	Studies
Pictures/photographs	Rose's (2016) visual methodology, in which observers put together a photo essay which is 'a series of photographs that creates emotions about the research topic and provides information about how culture and social relationships play out in the field of study' (Søndergaard, Frederiksen, Sørensen, & Lorentzen, 2019, p. 503)	(Søndergaard, et al., 2019; Søndergaard, Lorentzen, Sørensen, & Frederiksen, 2017)
Social Media	Using Facebook, Twitter, comment sections on news websites, etc., as sources of data on people's opinions, concerns, etc.	(Cleal, Willaing, Hoybye, & Thomsen, 2020)
Diaries	Field notes or daily reflections written down or recorded by participants or implementers of an intervention	(Bhanbhro, et al., 2016; Haruta & Yamamoto, 2020; Hobbs & Tully, 2020; Lefroy, et al., 2017; Limbani, Thorogood, Gomez-Olive, Kabudula, & Goudge, 2019; Ogrinc, et al., 2014; Parker, Mawson, Mountain, Nasr, & Zheng, 2014; van Gool, Bierbooms, Janssen, & Bongers, 2020)
Vignettes	'Short stories about a hypothetical person, presented to participants during qualitative research (e.g. within an interview or group discussion) or quantitative research, to glean information about their own set of beliefs' (Gourlay, et al., 2014, p. 2)	(Ohly, Crossland, Dykes, Lowe, & Moran, 2018; Sriranjana, Abrams, Wong, & Park, 2020)

Søndergaard, et al. (2019) used photographs which accompanied their ethnographic observations and were inspired by Rose's (2016) visual methodology in research. Citing Rose (2016), Søndergaard, et al. (2019) argued that '[w]hat is seen and how it is seen are culturally constructed; therefore, they contribute to research by appending information that elaborates on the variety and complexity of the data' (p. 502). In accord with this methodology, photos are taken by the observers and brought together in a photo essay which is 'a series of photographs that creates emotions about the research

topic and provides information about how culture and social relationships play out in the field of study' (Søndergaard, et al., 2019, p. 503).

Another use of photos can be found in Mukumbang and van Wyk (2020), who employed the method of photovoice. In this approach, rather than the observers, the respondents take photos. Although self-identified as a critical realist study (because of which it was not included in the review), the study also used the CMO configuration, hence its relevance to realist evaluators and its brief mention here.

One study in the review used social media to collect data (Cleal, et al., 2020). However, this was mainly due to the fact that the intervention took place on social media. There was no real discussion about the rationale behind or the implications of using social media in this specific way. However, such a discussion can be found in a study by Jamison, Sutton, Mant, and De Simoni (2018), which explored how data gathered through social media differs from data gathered from interviews.

Several studies used diaries as a source of data. Diaries are field notes or daily reflections written down or recorded by participants in or implementers of an intervention. Diaries can give more direct insight into the thinking of participants, as the initiative of what and when to write comes from the respondent, avoiding recall and interviewer bias to some extent (Lefroy, et al., 2017). However, our review shows that these diaries were rarely central and their precise role in the data analysis remained undiscussed in the studies under review. The study by Hobbs and Tully (2020) was an exception, but only because the diaries were part of the intervention.

Finally, two studies used vignettes during in-depth interviews. Sriranjana, et al. (2020) presented patient cases to their respondents in order to 'stimulate participant discussion about identification and management of postnatal depression in a range of situations' (p. 2-3), with the aim of identifying the respondents reasoning and looking for mechanisms. Interestingly, Ohly, et al. (2018) used vignettes to present initial programme theories to the respondents during a realist interview. They claim that it 'helped to elicit in-depth explanatory data at the level of context and mechanisms' (p. 3).

We also found several studies that combined other methodologies with RE. These were not included in our analysis because they are not pure data collection methods. However, examples include: social network analysis (Spitzer-Shohat, et al., 2018), qualitative comparative analysis (Goicolea, et al., 2015), ethnography (among others Dainty, et al., 2018; Orchard, et al., 2019; Redgate, Potrac, Boocock, & Dalkin, 2020; Rycroft-Malone, et al., 2013; Willis, et al., 2018), systems thinking (Kwamie, van Dijk, & Agyepong, 2014; Renmans, Holvoet, & Criel, 2020), Q sort methodology (Benmore, Henderson, Mountfield, & Wink, 2018; Harris, Henderson, & Wink, 2019) and randomized controlled trial (among others Vugts, et al., 2020; Warren, Melendez-Torres, Viner, & Bonell, 2020).

Discussion

In this section, we discuss some of the main findings arising from this mapping review and link them with discussions found in the theoretical, philosophical and methodological literature. We conclude with some ways forward for future REs and research.

Qualitative methods predominate. Rather unsurprisingly and in line with an earlier, more limited review by Manzano (2016), we found that qualitative methods are most commonly used. This predominance of qualitative research methods reflects the general situation in the social sciences, as well as the fact that the mechanisms sought by RE are often hidden and difficult to measure and thus not easily captured with more quantitative approaches that lack interpretative depth (Astbury & Leeuw, 2010; Bonell, Warren, & Melendez-Torres, 2022; Sayer, 1992). Therefore, the utility of quantitative methods is still debated in realism (Sayer, 2000), especially when they are used in a randomized controlled trial design (Marchal, et al., 2013; Van Belle, et al., 2016). The main objection is that the underlying focus on correlations and regularities in most statistical approaches is not in line with the realist generative view on causation (Marchal, et al., 2013). Moreover, the nature of the generative mechanisms that are central to any realist approach makes quantification very difficult, or even impossible, because of their complexity and unobservability (Bonell, et al., 2022).

However, several studies do use quantitative approaches such as structural equation modelling (for example Ford, et al., 2018; Oroviogoicochea & Watson, 2009), descriptive statistics or logistic regression (Ravn, 2019), or embed RE within a randomized controlled trial (for example Husted, Esbensen, Hommel, Thorsteinsson, & Zoffmann, 2014; Warren, et al., 2020). Dyer and Williams (2021) have even claimed that for realism 'to be relevant' (p. 110) quantification is necessary, as 'the way we theorise the world inevitably requires measurement to test such theories' (p. 112). However, ultimately, as they rightly claimed, 'all quantitative research (...) requires interpretation' (p.120), which necessitates some level of qualitative data collection. Similarly, Ravn (2019) emphasized that his quantitative approach 'should be regarded as supplementary to qualitative evidence and analysis' (p. 173).

Interview methods are the most common. Among the qualitative methods, interviews, especially individual interviews, are most commonly used. Different theoretical arguments justify this. One argument relates to the epistemological position of RE, which posits that 'all enquiry and observation are shaped and filtered through the human brain' (Westhorp, 2014, p. 4). Hence, our perceptions and observations of the world may be biased by our interests, our socio-economic position, our political stances, cognitive illusions, etc. Consequently, we do not have direct access to the mechanisms that we attempt to uncover (Sayer, 1992). This means that as evaluators we cannot merely rely on our own

observations and interpretations. A second argument is related to the predominant conceptualization of a mechanism in RE as an interaction between resources provided by the intervention and the reasoning of relevant change actors (Dalkin, Greenhalgh, Jones, Cunningham, & Lhussier, 2015; Pawson & Tilley, 1997). These change actors make the intervention work or not; hence, insight into their reasoning and perception of the intervention is crucial. An important caveat is that this might lead to an explanation that is overly reliant on the individual, while overlooking other causal mechanisms (see Westhorp (2018), for an overview of different kinds of mechanisms).

Mainly purposive sampling, very little theoretical sampling. Both of the above arguments for the use of interviews influence, at least in theory, the selection of the sample, although in different ways. The first argument focuses on the inclusion of as many different perspectives as possible (including the beneficiaries or users), while the second argument highlights the need to interview the actors who are most closely involved in the implementation of the intervention and service delivery and thus those who have the greatest causal power (policymakers, implementers, service deliverers, etc.). As observed, a large majority of the studies use purposive sampling and justify it using one or both of these arguments. However, an alternative kind of purposive sampling is theoretical or even realist sampling (Emmel, 2013), in which respondents are chosen based on how well they are placed to provide insight into certain claims of the programme theory under investigation. Manzano (2016) also suggests the selection of different stakeholders during the different phases of a realist investigation: policymakers during the initial phase of theory gleaning, and frontline workers during the refinement and consolidation phase, while users and beneficiaries should be used to analyse the outcomes rather than mechanisms. Interestingly, despite the theory-driven nature of RE, this sampling method is rarely used.

Half of the REs rely on only one group of respondents. As mentioned above, different respondents can contribute different pieces of data and information; however, half of the REs are based on data solely from the users/beneficiaries or the key informants (policymakers, implementers, service deliverers, etc.). Although specific evaluation circumstances may justify this focus on one group of respondents, the influence of interests and social position, discussed above, may give a biased and incomplete understanding of the intervention.

Low use of realist interviewing. In line with the findings of Manzano (2016), only one in five REs that use interviews as a source of data, apply a realist interviewing technique. Mukumbang, et al. (2020) have suggested that this might be related to 'the influence of other paradigms and research approaches, such as the grounded theory or the investigator's inclination to understand the process, which pre-supposes unstructured interviews' (p. 487). Other reasons might be a lack of awareness

about the interviewing technique or a fear of power and language issues. Indeed, Gilmore (2019) highlighted the ethical and methodological issues that arise when a foreign evaluator performs a realist interview in the context of low and middle-income countries. For example, concepts may be difficult to translate into local languages and the power imbalance may lead to respondents merely agreeing with the interviewer. These issues are not restricted to low and middle-income countries but may occur in every evaluation performed by an outsider. Bonell, et al. (2022) and Thomsen (2022) proposed that this may be overcome by first giving the respondents room to express themselves in their own words, after which the researcher introduces the concepts that match the story of the respondents, thereby turning the sequence of a realist interview around.

Observations are common. Observations are not as common as interviews, although they are still used in the majority of the studies. This may reflect, to a certain extent, the relative kinship of RE with ethnography (Van Belle, van de Pas, & Marchal, 2017). Indeed, some studies explicitly mention ethnography as a relevant framework for their data collection (Dainty, et al., 2018; Redgate, et al., 2020). The analysis of the context in particular, can greatly benefit from the use of observation methods. However, the analysis of mechanisms can also be furthered using observations. According to Bonell, et al. (2022), observations ‘can allow researchers to witness the actions and interactions that constitute or spin off from the enactment of interventions, which may offer insights into how outcomes might be generated’ (p. 8). However, they warn us that observations may be limited to the mechanism sections that occur directly after the intervention actions, and that more long-term mechanisms behind outcomes are more difficult to observe. Moreover, as mechanisms are unobservable and often lie in the minds of causal agents (Pawson & Tilley, 1997), combining observation with other data collection methods, most notably interviews, is indispensable (Handley, Bunn, Lynch, & Goodman, 2020). Indeed, while Handley, et al. (2020) found ‘observation a useful method for helping to surface mechanisms located at an individual level’ (p. 390), it was only when the observants communicated their way of thinking to the observers that these mechanisms were actually revealed. Hence, it is not the observation as such that provides access to the underlying mechanisms but the informal interview that occurs during the observation at the precise moment that the mechanism is playing out. Clearly, the role of observation in RE and its search for mechanisms and an updated programme theory may benefit from further clarifications and methodological development.

Surveys are less common and lack a clear function. Surveys are used much less than the qualitative approaches, and their function differs between studies. Surveys are either used to measure outcomes (e.g. Grace & Horstmannshof, 2019; Robertson, Pointing, Stevenson, & Clough, 2013), describe the context (e.g. Spitzer-Shohat, et al., 2018), analyse mechanisms (e.g. James, Romo-Murphy, & Oczon-

Quirante, 2019) or all three at the same time (e.g. Mutschler, Rouse, McShane, & Habal-Brosek, 2018). Notwithstanding the previously mentioned objections to the use of quantitative methods, surveying remains an interesting method as it can gather a lot of information at the individual level, making it worthwhile to take it into account in RE.

Using multiple methods is common. Combining multiple methods is seen as good practice in RE (Greenhalgh, et al., 2015) and, in the majority of cases, interviews were accompanied by another method to strengthen the causal claims. However, only one quarter of these studies combined their qualitative approach with a quantitatively oriented survey. Interestingly, apart from a discussion about whether realism and statistical approaches can be properly combined at all, another discussion seems to focus on the role and timing of quantitative and qualitative methods. Dyer and Williams (2021) emphasized the importance of quantification and the necessity of interpreting the statistics after quantitative inquiry. In contrast, Bonell, et al. (2022) argued that ‘there is a role for correlational quantitative research in checking whether broader patterns of regularities appear to align with the theories developed or refined through qualitative research’ (p. 12). Hence, the qualitative part should come before the quantitative inquiry.

Few innovative methods. Finally, interviewing different stakeholders to gain different perspectives may already achieve a great deal. Moreover, triangulating different data sources will strengthen causal claims about mechanisms (Manzano, 2016). However, the unobservable character of mechanisms should encourage creativity among evaluators when it comes to data collection methods to unearth these mechanisms. The review revealed that innovative methods that may have more potential to do the task of unearthing mechanisms are not used very often, which creates room for significant advances.

Below, we present some lessons and ways forward for the improvement of the use of data collection methods in RE.

1) [Develop realist surveys](#)

Surveys are a common method used in science and evaluation in general, but are underdeveloped in RE. While there are some objections to an overreliance on quantitative methods, surveys may largely be of benefit to RE when used in a realist way, as they make it possible to gather a wealth of information at the individual level in a relatively efficient way. Indeed, individual characteristics, perceptions, reasoning and behaviour changes (all part of a CMO configuration) may be relatively easily linked to each other. It may therefore be worthwhile to develop a realist adaptation of the survey. One crucial element to take into account in this regard is that RE is a case-based approach, while most quantitative methods are used in a variable-based approach (Dyer & Williams, 2021). The

analysis of the survey, therefore, needs to be done at the level of the respondents (which constitutes a case in the RE) and not at the level of the variables. Such a realist survey might be less relevant when looking at mechanisms at levels above the individual (e.g. an organization or society).

2) Explore the possibilities of more innovative methods

The hidden character of mechanisms and the complex causal structures that imply them makes them difficult to unearth. Resorting to more innovative methods that are less common may be an interesting way forward that should be further explored. Some examples of promising methods include Q methodology (Brown, 1993), photovoice (Mukumbang & van Wyk, 2020), soft systems methodology (Dalkin, Lhussier, Williams, Burton, & Rycroft-Malone, 2018) and causal loop diagramming (Renmans, et al., 2020). Importantly, when using such methods it is crucial to be aware of the underlying ontological and epistemological assumptions and, if necessary, to make the adaptations needed to bring the methods in line with the underlying philosophy of RE.

3) Sampling deserves more attention

Most attention in methodological writings on RE is given to the conceptualization of mechanisms, the role of context, the use of theory and the development of CMO configurations. Little has been written about the importance and relevance of sampling within RE, with a few exceptions (e.g. Emmel, 2013). Nonetheless, the sampling of respondents is a key step in every research cycle and especially in RE. Indeed, scientific realists believe that although there is one reality, every actor will have a different interpretation of that reality, which makes it crucial to sample the right respondents for the interviews. In some instances, a broad range of perspectives might be necessary, while in other instances the perspectives of very specific stakeholders may be more useful. Similarly, sampling can be based on theoretical needs (see Emmel, 2013) or, in situations where power imbalances are structural, widespread and underpin the evaluand, a more power-sensitive sampling procedure might be useful or even necessary to overcome these imbalances, gain a more in-depth understanding and overcome the evaluator's own cultural biases (Renmans, et al., 2022). Either way, the sampling procedure needs to be well thought through and deserves attention during the research process and in the methodological section of publications.

4) There is room to strengthen the theory-driven nature of the methods

RE is more than just theory-based and, arguably, can be labelled theory-driven, as theory strongly influences each step, including the analysis. However, this mapping review has shown that theory is not always present in the methods used. Realist interviews were not as common and realist sampling was rarely applied. Moreover, realist and theory-driven surveys and observations need to be further developed. We therefore argue that there is still much room for improvement in the theory-focus of

our data collection methods, which will undoubtedly have a positive effect on the theory-driven analysis and the quality of any refined programme theory. At the same time, such theory-driven data collection methods may amplify ethical concerns when applied in a situation where there is a great cultural, social, economic, hierarchical distance between the evaluator and the respondents and/or stakeholders (see Gilmore, 2019; Renmans, et al., 2022). Indeed, in such situations, questions about power differences, ownership and knowledge of the concepts, language barriers and translation issues concerning culturally embedded concepts need to be addressed.

Conclusion

We conducted a mapping review of the data collection methods used by REs published in the peer-reviewed literature. We found that REs primarily make use of qualitative methods, and interviews in particular. Sampling is mostly done in a purposeful way, with the objective to either involve different perspectives or to interview the actors most closely involved in the implementation of an intervention. Theoretical sampling is much less common than realist interviewing, which emphasizes the need to strengthen the theory-driven nature of our data collection methods. Similarly, the survey method is under-theorized, so it may need to be further adapted to the RE paradigm. This will grant it a stronger position in the RE methods toolkit. Finally, the hidden nature of mechanisms may make it useful to look for more fine-grained and creative data methods, which are still underexplored.

Funding

DR was funded by a postdoctoral research fellowship for fundamental research from the Research Foundation – Flanders (FWO), file number: 12V2720N.

References

- Astbury, B. & Leeuw, F.L. 2010. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *Am J Eval*, 31, 363-381.
- Benmore, G., Henderson, S., Mountfield, J., & Wink, B. 2018. The Stopit! programme to reduce bullying and undermining behaviour in hospitals. *Journal of Health Organization & Management*, 32, 428-443.
- Bennett, A. 2013. The mother of all isms: Causal mechanisms and structured pluralism in International Relations theory. *European Journal of International Relations*, 19, 459-481.
- Bertotti, M., Frostick, C., Hutt, P., Sohanpal, R., & Carnes, D. 2018. A realist evaluation of social prescribing: an exploration into the context and mechanisms underpinning a pathway linking primary care with the voluntary sector. *Primary Health Care Research & Development*, 19, 232-245.
- Bhanbhro, S., Gee, M., Cook, S., Marston, L., Lean, M., & Killaspy, H. 2016. Recovery-based staff training intervention within mental health rehabilitation units: a two-stage analysis using realistic evaluation principles and framework approach. *BMC Psychiatry*, 16, 292.
- Bhaskar, R. 2008 [1975]. *A realist theory of science*. Verso books, London.
- Bonell, C., Warren, E., & Melendez-Torres, G. 2022. Methodological reflections on using qualitative research to explore the causal mechanisms of complex health interventions. *Evaluation*, 0.
- Brousselle, A. & Buregeya, J.M. 2018. Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 24, 153-168.
- Brown, S.R. 1993. A primer on Q methodology. *Operant subjectivity*, 16, 91-138.
- Cleal, B., Willaing, I., Hoybye, M.T., & Thomsen, H.H. 2020. Facebook as a Medium for the Support and Enhancement of Ambulatory Care for People With Diabetes: Qualitative Realist Evaluation of a Real-World Trial. *JMIR Diabetes*, 5, e18146.
- Dainty, K.N., Golden, B.R., Hannam, R., Webster, F., Browne, G., Mittmann, N., Stern, A., & Zwarenstein, M. 2018. A realist evaluation of value-based care delivery in home care: The influence of actors, autonomy and accountability. *Soc Sci Med*, 206, 100-109.
- Dalkin, S.M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. 2015. What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science : IS*, 10, 49.
- Dalkin, S.M., Lhussier, M., Jones, D., Phillipson, P., & Cunningham, W. 2018. Open communication strategies between a triad of 'experts' facilitates death in usual place of residence: A realist evaluation. *Palliative Medicine*, 32, 980-989.
- Dalkin, S.M., Lhussier, M., Williams, L., Burton, C.R., & Rycroft-Malone, J. 2018. Exploring the use of Soft Systems Methodology with realist approaches: A

- novel way to map programme complexity and develop and refine programme theory. *Evaluation*, 24, 84-97.
- Desveaux, L., Shaw, J., Saragosa, M., Soobiah, C., Marani, H., Hensel, J., Agarwal, P., Onabajo, N., Bhatia, R.S., & Jeffs, L. 2018. A Mobile App to Improve Self-Management of Individuals With Type 2 Diabetes: Qualitative Realist Evaluation. *Journal of Medical Internet Research*, 20, e81.
- Dyer, W. & Williams, M. 2021. Quantification and realist methodologies. *International Journal of Social Research Methodology*, 24, 109-121.
- Efstathiou, N., Lock, A., Ahmed, S., Parkes, L., Davies, T., & Law, S. 2020. A realist evaluation of a "single point of contact" end-of-life care service. *Journal of Health Organization & Management. ahead of print*, 25, 25.
- Emmel, N., 2013. *Sampling and Choosing Cases in Qualitative Research: A Realist Approach*, Publishing, London.
- Esterberg, K.G. 2002. *Qualitative Methods in Social Research*. McGraw-Hill, Boston.
- Ford, J.A., Jones, A., Wong, G., Clark, A., Porter, T., & Steel, N. 2018. Access to primary care for socio-economically disadvantaged older people in rural areas: exploring realist theory using structural equation modelling in a linked dataset. *BMC Medical Research Methodology*, 18, 57.
- Gilmore, B. 2019. Realist evaluations in low- and middle-income countries: reflections and recommendations from the experiences of a foreign researcher. *BMJ Global Health*, 4, e001638.
- Goicolea, I., Vives-Cases, C., Hurtig, A.K., Marchal, B., Briones-Vozmediano, E., Otero-Garcia, L., Garcia-Quinto, M., & San Sebastian, M. 2015. Mechanisms that Trigger a Good Health-Care Response to Intimate Partner Violence in Spain. Combining Realist Evaluation and Qualitative Comparative Analysis Approaches. *PLoS ONE [Electronic Resource]*, 10, e0135167.
- Gourlay, A., Mshana, G., Birdthistle, I., Bulugu, G., Zaba, B., & Urassa, M. 2014. Using vignettes in qualitative research to explore barriers and facilitating factors to the uptake of prevention of mother-to-child transmission services in rural Tanzania: a critical analysis. *BMC Medical Research Methodology*, 14, 21.
- Grace, S. & Horstmanshof, L. 2019. A realist evaluation of a regional Dementia Health Literacy Project. *Health Expectations*, 22, 426-434.
- Grant, M.J. & Booth, A. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26, 91-108.
- Green, J. & Thorogood, N. 2009. *Qualitative Methods for Health Research*. SAGE Publications.
- Greenhalgh, T., Wong, G., Jagosh, J., Greenhalgh, J., Manzano, A., Westhorp, G., & Pawson, R. 2015. Protocol-the RAMESES II study: Developing guidance and reporting standards for realist evaluation. *BMJ open*, 5.
- Handley, M., Bunn, F., Lynch, J., & Goodman, C. 2020. Using non-participant observation to uncover mechanisms: Insights from a realist evaluation. *Evaluation*, 26, 380-393.

- Harris, K., Henderson, S., & Wink, B. 2019. Mobilising Q methodology within a realist evaluation: Lessons from an empirical study. *Evaluation, 25*, 430-448.
- Haruta, J. & Yamamoto, Y. 2020. Realist approach to evaluating an interprofessional education program for medical students in clinical practice at a community hospital. *Medical Teacher, 42*, 101-110.
- Hernandez, A., Hurtig, A.-K., Dahlblom, K., & San Sebastian, M. 2014. More than a checklist: a realist evaluation of supervision of mid-level health workers in rural Guatemala. *BMC Health Serv Res, 14*, 112.
- Hobbs, G. & Tully, M.P. 2020. Realist evaluation of public engagement and involvement in data-intensive health research. *Research Involvement & Engagement, 6*, 37.
- Husted, G.R., Esbensen, B.A., Hommel, E., Thorsteinsson, B., & Zoffmann, V. 2014. Adolescents developing life skills for managing type 1 diabetes: a qualitative, realistic evaluation of a guided self-determination-youth intervention. *Journal of Advanced Nursing, 70*, 2634-2650.
- James, R.W., Romo-Murphy, E., & Oczon-Quirante, M.M. 2019. A Realist Evaluation of a Community-Centered Radio Initiative for Health and Development in Mindanao, Philippines. *Asia-Pacific Journal of Public Health, 31*, 559-571.
- Jamison, J., Sutton, S., Mant, J., & De Simoni, A. 2018. Online stroke forum as source of data for qualitative research: insights from a comparison with patients' interviews. *BMJ open, 8*, e020133.
- Kwamie, A., van Dijk, H., & Agyepong, I.A. 2014. Advancing the application of systems thinking in health: realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana. *Health Research Policy & Systems, 12*, 29.
- Lacouture, A., Breton, E., Guichard, A., & 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.
- Lefroy, J., Yardley, S., Kinston, R., Gay, S., McBain, S., & McKinley, R. 2017. Qualitative research using realist evaluation to explain preparedness for doctors' memorable 'firsts'. *Medical Education, 51*, 1037-1048.
- Lemire, S., Kwako, A., Nielsen, S.B., Christie, C., Donaldson, S., & Leeuw, F. 2020. What is this thing called a mechanism? Findings from a review of realist evaluations. *New Directions for Evaluation, 2020*, 73-86.
- Lemire, S., Peck, L.R., & Porowski, A. 2020. The Growth of the Evaluation Tree in the Policy Analysis Forest: Recent Developments in Evaluation. *Policy Studies Journal, 48*, S47-S70.
- Limbani, F., Thorogood, M., Gomez-Olive, F.X., Kabudula, C., & Goudge, J. 2019. Task shifting to improve the provision of integrated chronic care: realist evaluation of a lay health worker intervention in rural South Africa. *BMJ Global Health, 4*, e001084.
- Manzano, A. 2016. The craft of interviewing in realist evaluation. *Evaluation, 22*, 342-360.

- Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & 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.
- Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G., & Pawson, R. 2013. Realist RCTs of complex interventions – An oxymoron. *Soc Sci Med*, 94, 124-128.
- Martin, P. & Tannenbaum, C. 2017. A realist evaluation of patients' decisions to deprescribe in the EMPOWER trial. *BMJ open*, 7, e015959.
- Mukumbang, F.C., Marchal, B.E., Van Belle, S., & van Wyk, B. 2020. Using the realist interview approach to maintain theoretical awareness in realist studies. *Qualitative Research*, 20, 485-515.
- Mukumbang, F.C. & van Wyk, B. 2020. Leveraging the Photovoice Methodology for Critical Realist Theorizing. *International Journal of Qualitative Methods*, 19.
- Mutschler, C., Rouse, J., McShane, K., & Habal-Brosek, C. 2018. Developing a realist theory of psychosocial rehabilitation: the Clubhouse model. *BMC Health Serv Res*, 18, 442.
- Nielsen, S.B., Lemire, S., & Tangsig, S. 2022. Unpacking context in realist evaluations: Findings from a comprehensive review. *Evaluation*, 28, 91-112.
- Ogrinc, G., Ercolano, E., Cohen, E.S., Harwood, B., Baum, K., van Aalst, R., Jones, A.C., & Davies, L. 2014. Educational system factors that engage resident physicians in an integrated quality improvement curriculum at a VA hospital: a realist evaluation. *Academic Medicine*, 89, 1380-1385.
- Ohly, H., Crossland, N., Dykes, F., Lowe, N., & Moran, V.H. 2018. A realist qualitative study to explore how low-income pregnant women use Healthy Start food vouchers. *Maternal & Child Nutrition*, 15, e12632.
- Orchard, J., Li, J., Gallagher, R., Freedman, B., Lowres, N., & Neubeck, L. 2019. Uptake of a primary care atrial fibrillation screening program (AF-SMART): a realist evaluation of implementation in metropolitan and rural general practice. *BMC Family Practice*, 20, 170.
- Oroviogicochea, C. & Watson, R. 2009. A quantitative analysis of the impact of a computerised information system on nurses' clinical practice using a realistic evaluation framework. *International Journal of Medical Informatics*, 78, 839-849.
- Parker, J., Mawson, S., Mountain, G., Nasr, N., & Zheng, H. 2014. Stroke patients' utilisation of extrinsic feedback from computer-based technology in the home: a multiple case study realistic evaluation. *BMC Medical Informatics & Decision Making*, 14, 46.
- Pawson, R. 1996. Theorizing the Interview. *The British Journal of Sociology*, 47, 295-314.
- Pawson, R. 2013. *The science of evaluation: A realist manifesto*. Sage publications, London.
- Pawson, R. & Manzano-Santaella, A. 2012. A realist diagnostic workshop. *Evaluation*, 18, 176-191.

- Pawson, R. & Tilley, N. 1997. *Realistic evaluation*. Sage, London.
- Ravn, R. 2019. Testing mechanisms in large-N realistic evaluations. *Evaluation*, 25, 171-188.
- Reddy, S., Carey, T.A., & Wakerman, J. 2016. A Realist Case Study of a Regional Hospital's Response to Improve Emergency Department Access in the Context of Australian Health Care Reforms. *Health Services Research & Managerial Epidemiology*, 3, 2333392816631101.
- Redgate, S., Potrac, P., Boocock, E., & Dalkin, S. 2020. Realist evaluation of the Football Association's Post Graduate Diploma (PG Dip) in Coach Development. *Sport Education and Society*.
- Renmans, D., Holvoet, N., & Criel, B. 2020. No Mechanism Without Context: Strengthening the Analysis of Context in Realist Evaluations Using Causal Loop Diagramming. *New Directions for Evaluation*, 101-114.
- Renmans, D., Sarkar, N., Van Belle, S., Affun-Adegbulu, C., Marchal, B., & Mukumbang, F.C. 2022. Realist evaluation in times of decolonising global health. *Int J Health Plann Manage*, n/a.
- Richardson, E.Z.L., Phillips, M., Colom, A., & Nichols, J. 2019. Honoring Lived Experience: Life Histories as a Realist Evaluation Method. *Am J Eval*, 40, 509-532.
- Robertson, J., Pointing, B.S., Stevenson, L., & Clough, A.R. 2013. "We made the rule, we have to stick to it": towards effective management of environmental tobacco smoke in remote Australian Aboriginal communities. *International Journal of Environmental Research & Public Health [Electronic Resource]*, 10, 4944-4966.
- Rose, G. 2016. *Visual Methodologies: An Introduction to Researching with Visual Materials*. SAGE Publications, London.
- Rutten, R. 2021. Uncertainty, possibility and causal power in QCA. *Sociological Methods & Research*.
- Rycroft-Malone, J., Wilkinson, J., Burton, C.R., Harvey, G., McCormack, B., Graham, I., & Staniszewska, S. 2013. Collaborative action around implementation in Collaborations for Leadership in Applied Health Research and Care: towards a programme theory. *Journal of Health Services & Research Policy*, 18, 13-26.
- Sayer, A. 1992. *Method in Social Science: A Realist Approach*. Routledge, London, New York.
- Sayer, A. 2000. *Realism and Social Science*. SAGE Publications, London.
- Søndergaard, S.F., Frederiksen, K., Sørensen, E.E., & Lorentzen, V. 2019. A Realistic Evaluation of Danish Perioperative Nurses' Documentation Practices. *AORN Journal*, 110, 500-509.
- Søndergaard, S.F., Lorentzen, V., Sorensen, E.E., & Frederiksen, K. 2017. Danish Perioperative Nurses' Documentation: A Complex, Multifaceted Practice Connected With Unit Culture and Nursing Leadership. *AORN Journal*, 106, 31-41.

- Spitzer-Shohat, S., Shadmi, E., Goldfracht, M., Key, C., Hoshen, M., & Balicer, R.D. 2018. Evaluating an organization-wide disparity reduction program: Understanding what works for whom and why. *PLoS ONE [Electronic Resource]*, *13*, e0193179.
- Sriranjan, A.N., Abrams, R., Wong, G., & Park, S. 2020. Towards an understanding of GPs' viewpoint on diagnosing postnatal depression in general practice: a small-scale realist evaluation. *Primary Health Care Research & Development*, *21*, e42.
- The RAMESES II project, 2017. What is a mechanism? What is a programme mechanism?, Publishing, Oxford, UK.
- Thomsen, L.L., 2022. Reflections on realist interviewing: testing a program theory with parents in psychologically and socially vulnerable positions, in: Jagosh, J. (Ed.), CARES 2022 Realist Methodology Training Conference, Publishing, Online conference.
- Van Belle, S., van de Pas, R., & Marchal, B. 2017. Towards an agenda for implementation science in global health: there is nothing more practical than good (social science) theories. *BMJ Global Health*, *2*, e000181.
- Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A., & Marchal, B. 2016. Can “realist” randomised controlled trials be genuinely realist? *Trials*, *17*, 313.
- van Gool, F., Bierbooms, J., Janssen, R., & Bongers, I. 2020. Realistic evaluation of the Flexmonitor, a team program to maintain and reflect on flexibility. *Journal of Health Organization & Management. ahead of print*, *09*, 09.
- Vugts, M.A., Zedlitz, A.M., Joosen, M.C., & Vrijhoef, H.J. 2020. Serious Gaming During Multidisciplinary Rehabilitation for Patients With Chronic Pain or Fatigue Symptoms: Mixed Methods Design of a Realist Process Evaluation. *Journal of Medical Internet Research*, *22*, e14766.
- Warren, E., Melendez-Torres, G.J., Viner, R., & Bonell, C. 2020. Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention. *Trials [Electronic Resource]*, *21*, 774.
- Westhorp, G., 2014. Realist impact evaluation: An introduction, Publishing, London.
- Westhorp, G., 2018. Understanding mechanisms in realist evaluation and research, in: Emmel, N., Greenhalgh, J., Manzano, A., Monaghan, M. & Dalkin, S. (Eds.), *Doing realist research*, Publishing, London, pp. 41-57.
- Whitelaw, S., Gibson, I., Wild, A., Hall, H., & Molloy, H. 2021. The Dumfries Arts Award Project: towards building a programme theory of innovation transfer across two social organisations. *Soc. Enterp. J.*
- Willis, C.E., Reid, S., Elliott, C., Rosenberg, M., Nyquist, A., Jahnsen, R., & Girdler, S. 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*, 113.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J., & Greenhalgh, T. 2016. RAMESES II reporting standards for realist evaluations. *BMC Medicine*, 14, 96.

Wye, L., Lasseter, G., Percival, J., Duncan, L., Simmonds, B., & Purdy, S. 2014. What works in 'real life' to facilitate home deaths and fewer hospital admissions for those at end of life?: results from a realist evaluation of new palliative care services in two English counties. *BMC Palliative Care*, 13, 37.

PRE-PRINT