

# Applying and reporting relevance, richness and rigour in realist evidence appraisals: Advancing key concepts in realist reviews

Sara Dada<sup>1,2</sup>  | Sonia Dalkin<sup>3,4</sup>  | Brynne Gilmore<sup>1,2</sup>  | Rebecca Hunter<sup>5</sup>  | Ferdinand C. Mukumbang<sup>6</sup> 

<sup>1</sup>UCD Centre for Interdisciplinary Research Education and Innovation in Health Systems (UCD IRIS Centre), School of Nursing Midwifery and Health Systems, University College Dublin, Dublin, Ireland

<sup>2</sup>School of Nursing Midwifery and Health Systems, University College Dublin, Dublin, Ireland

<sup>3</sup>Faculty of Health and Life Sciences, Northumbria University, Newcastle Upon Tyne, UK

<sup>4</sup>Fuse (The Centre for Translational Research in Public Health), Northumbria University, Newcastle Upon Tyne, UK

<sup>5</sup>Department of Nursing and Midwifery, University of Highlands and Islands, Inverness, UK

<sup>6</sup>Department of Global Health, University of Washington, Seattle, Washington, USA

## Correspondence

Sara Dada, UCD Centre for Interdisciplinary Research Education and Innovation in Health Systems (UCD IRIS Centre), School of Nursing Midwifery and Health Systems, University College Dublin, Dublin, Ireland.

Email: [sara.dada@ucdconnect.ie](mailto:sara.dada@ucdconnect.ie)

## Abstract

The realist review/synthesis has become an increasingly prominent methodological approach to evidence synthesis that can inform policy and practice. While there are publication standards and guidelines for the conduct of realist reviews, published reviews often provide minimal detail regarding how they have conducted some methodological steps. This includes selecting and appraising evidence sources, which are often considered for their ‘relevance, richness and rigour.’ In contrast to other review approaches, for example, narrative reviews and meta-analyses, the inclusion criteria and appraisal of evidence within realist reviews depend less on the study’s methodological quality and more on its contribution to our understanding of generative causation, uncovered through the process of retroductive theorising. This research brief aims to discuss the current challenges and practices for appraising documents’ relevance, richness and rigour and to provide pragmatic suggestions for how realist reviewers can put this into practice.

## KEYWORDS

evidence appraisal, quality appraisal, realist synthesis, review methodology

## Highlights

### What is Already Known

- Existing guidance for realist reviews suggests considering sources’ relevance, richness and rigour.
- Realist reviews do not follow the traditional hierarchy of evidence; thus, the appraisal of evidence is different to more traditional systematic reviews.
- However, many published realist reviews provide minimal detail on the selection and appraisal of evidence sources.

### What is New

- This research brief suggests processes for considering the relevance, richness and rigour of evidence sources.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

- This article clarifies and reiterates previous guidelines for assessing the rigour of included documents by considering the trustworthiness of a data source and the coherence of theories.

#### Impact for Readers

- This article serves as a resource by presenting examples of reported evidence appraisals that can provide guidance to both novice and experienced researchers.
- The recommendations for transparency described in this article may also be applied to quality appraisals across different review types.

## 1 | THE REALIST REVIEW

Realist reviews (or realist syntheses) have gained prominence as a systematic review approach for capturing the complexity of policies and programmes implemented in complex systems.<sup>1</sup> Realist reviews are explanatory and strive to unpack ‘how, why, for whom, and in what contexts’ policies and programmes work or do not work.<sup>1</sup> Consequently, realist reviews address some of the shortcomings of traditional review methods in their ability to provide explanatory recommendations where traditional reviews may be too specific, inflexible and focus on measuring and reporting programme effectiveness.<sup>2</sup> Realist reviews, therefore, resist the notion of generalisability and give more value to exploratory theories about how policies and programmes are shaped by context.<sup>3</sup> This is done by theorising on the underlying mechanisms that may explain why and how change occurs.<sup>1,4–6</sup> These programme theories are developed, refined and tested through data provided by the review’s included sources.

Realist reviews, like other realist-informed approaches, are ontologically underpinned by tenets of critical realism and scientific realism and are guided by the notions of ontological depth and mechanism-based theorising. This is in line with Pawson and Tilley’s<sup>7</sup> description of generative causation. They explain that programmes alone do not bring about the observed outcome, but generate outcomes through individuals’ responses to the resources, ideas and practices that those programmes introduce (programme ‘mechanisms’), which are shaped by wider contexts.<sup>7</sup> A realist review starts and ends with theory, aiming to elucidate patterns of generative causation through retroductive theorising. This requires the application of abduction, induction, and deduction while sieving through relevant literature to test and refine programme theories.<sup>8,9</sup> Retroductive theories obtained from realist reviews are ontologically considered as an approximation of the truth regarding the phenomenon under consideration. Therefore, in a realist review, it is important to focus on not only the identified literature, but also the theory to which it contributes. This is achieved through the non-linear retroductive theorising process—in other words, the to’ing and fro’ing between the programme

theory developed and tested in the review, and the evidence included in the review.

The realist review is iterative, making it less linear or rigid than a traditional systematic review.<sup>10,11</sup> The steps of a realist review, demonstrated in Figure 1, are guided by the four approaches to inference-making in realist-informed inquiries: induction, abduction, deduction and retroduction.<sup>12</sup> Inductive thinking involves projecting from what we know to what we do not know and is also applied to identify patterns in the evidence to draw conclusions.<sup>12</sup> Deductive thinking is applied to distil constructs from the sources of evidence towards constructing the explanatory theories.<sup>12</sup> Abduction is used for interpreting and re-contextualising observed outcomes and events, identified mechanisms and context conditions to formulate the “best explanation”.<sup>12</sup>

## 2 | SELECTING AND APPRAISING EVIDENCE IN REALIST REVIEWS

To gain explanatory depth and breadth, researchers are encouraged to use data from a wide range of sources, including academic (e.g. peer-reviewed articles, conference posters) and non-academic sources (e.g. social media posts, news articles).<sup>15</sup> A realist review focuses on developing explanatory theory. Therefore, criteria for evidence appraisal and inclusion differ from other systematic evidence syntheses. Traditional reviews typically apply a strict inclusion/exclusion criteria to studies and may assess their rigour based on quality assessment checklists and methodological hierarchies. In contrast, the inclusion criteria for a realist review is based on the data’s ability to contribute to theory building and testing. To this end, the proposed criteria for appraising the evidence to be included in a realist review are relevance and rigour,<sup>16</sup> and richness.<sup>17</sup>

### 2.1 | Relevance and richness

In a realist review, relevance is not only applied to a particular *topic* but to the theory being tested.<sup>18</sup> The Realist

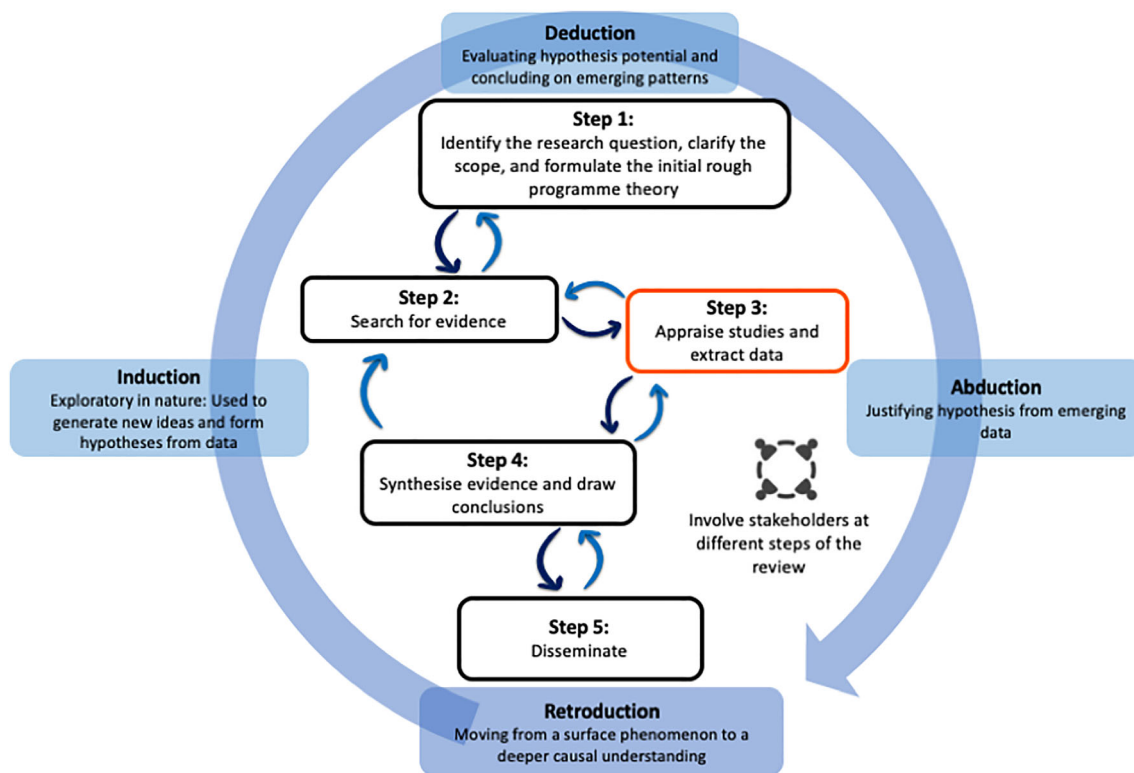


FIGURE 1 Proposed iterative process for searching articles in the synthesis. Adapted from Copper et al.<sup>13</sup> and Mukumbang et al.<sup>14</sup> [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

And METa-narrative Evidence Syntheses: Evolving Standards (RAMESES) quality standards for realist reviews define relevance as ‘whether [the data] can contribute to theory building and/or testing’.<sup>19</sup> However, this is constantly negotiated within realist reviews as the relevance of data may change as theories evolve. Pieces of evidence or sources that were previously excluded as ‘not relevant’ may prove influential in subsequent iterative rounds. Due to the fluctuating nature of relevance and the iterative nature of realist reviews, it is recommended that reviewers retain search results and excluded sources in case they need to be revisited and report this process transparently.<sup>20</sup> However, this can result in an overwhelmingly large amount of data which may or may not have sufficient detail to support theory development and testing. Adding the category of richness can support researchers to identify resources that have more explanatory evidence than others, making the review more manageable and the findings more relevant.

More recently Booth et al.<sup>17</sup> suggested evidence should also be appraised for richness. Booth et al.<sup>17</sup> noted that sources can have ‘conceptual richness’ and/or ‘contextual thickness.’ This conceptual richness is the degree of theoretical and conceptual development that explains how an intervention is expected to work. Contextual thickness entails sufficient detail that enables the reader

(i) to establish what is occurring in the intervention as well as in the wider context and (ii) to infer whether findings can be transferred to other people, places, situations or environments. Appraising for richness ensures only sources which can meaningfully contribute to the research question are included at a particular time.

## 2.2 | Rigour

The RAMESES quality standards for realist reviews defined rigour as ‘whether the method used to generate that particular piece of data is credible and “trustworthy”’.<sup>19</sup> In practice, multiple terms have been used to describe what rigour means in a realist review, causing confusion. Within prominent realist review methods papers and guidelines, terms such as credible, plausible, believable, trustworthy and coherent have been used individually or in combination to describe rigour.<sup>16,18,19,21</sup>

This difficulty of defining and applying rigour is compounded when using data from different research paradigms and non-academic sources.

Pawson et al.<sup>18</sup> suggest sample size, data collection techniques, analysis methods, and research claims should be considered when assessing rigour. Quality guidelines and checklists have been created to help researchers

make these judgements for traditional reviews. However, Pawson<sup>22</sup> also claims these traditional assessments of quality should be balanced, as “‘nuggets’ of wisdom can be found in methodologically weak studies, and realist review encourages the use of non-academic sources’.

### 3 | SNAPSHOT OF EVIDENCE APPRAISALS REPORTED IN REALIST REVIEWS

A search for manuscripts published in 2021 was conducted on PubMed using the keywords “realist reviews/syntheses” and “health systems.” This focus on health systems was used to narrow the scope and provide a snapshot of one discipline. The Methods sections of the resulting 73 published reviews were read and screened for terms ‘relevance,’ ‘richness’ and ‘rigour’ to determine how reviewers applied these appraisals to the articles screened in their reviews (Appendix S1). While 67 publications (over 91% of reviews) explicitly stated relevance criteria within the manuscript text, a supplement, or referred to a protocol, richness and rigour were often not included in the same level of detail. Fifty-seven publications (78% of reviews) acknowledged conducting some sort of rigour assessment, with only 28 (38% of all papers) describing the criteria or tools used. Richness was the least often explained and was not always explicitly named with this term; instead, it was commonly alluded to as the ‘level of relevance’. While 33 reviews acknowledged some form of ranking by richness (45% of reviews), only 13 (18% of all publications) explained how this was determined.

#### 3.1 | Relevance

Screening based on relevance for realist reviews mirrors the screening processes used in other types of reviews. This screening process is often presented as a list of inclusion and exclusion criteria given in the manuscript’s main text or in a supplement. Relevance was applied in two ways: the article (i) is relevant to the content/topic of interest, and (ii) provides evidence relevant to theory development, refinement or testing. Reviewers conducted relevance assessments in one or both of these ways. For example, Epstein et al.<sup>23</sup> explicitly state that ‘studies were expected to have adequate relevance to build the program theory’ while Morton et al.<sup>24</sup> present initial criteria for the types of interventions that should be included or excluded in the review, followed by additional guidelines on further assessing relevance. They explain:

An article should comply with the inclusion/exclusion criteria in the first instance, except

where agreed by the team for inclusion for a specific reason e.g. containing data that is broadly transferable and of use to the programme theory. A low rating might mean the article only contains a few relevant lines, with the bulk of the text focused on other, non-relevant matters. A medium rating might mean an article has a lot of detail on one relevant issue (e.g. engaging people and keeping them engaged) which is pertinent to sustainability, but otherwise little on other important factors. A high rating will mean an article has a direct focus on maintaining an intervention sustainable long term, with a good level of detail.<sup>24</sup>

#### 3.2 | Richness

Descriptions of how richness assessments were applied were often interwoven with relevance. Ulrich et al.<sup>25</sup> describe how relevant papers were classified on ‘the thick/thin continuum, which refers to a paper’s density of evidence regarding CMOs that are relevant to the study’s scope’ where ‘thick’ articles offered more detail on CMOs relevant to the review topic and ‘thin’ articles provided less. This ranking of relevance in terms of contribution to theory testing was a common operationalisation of richness assessments. For example, Pearce et al.<sup>26</sup> explain how the lead author classified articles ‘into categories of high and low relevance, based on her judgments on the relevance of the data within these articles for programme theory development’ while Morton et al.<sup>24</sup> use ‘a “traffic light” assessment system of low, medium and high relevance’. A review published by Calderón-Larrañaga et al.<sup>27</sup> was one of two publications to explicitly name and provide richness criteria, which they define in their supplementary material as:

Conceptually rich: studies with well-grounded and clearly described theories and concepts.

Conceptually thick: studies with a rich description of a programme was provided, but without explicit reference to the theory underpinning it.

Conceptually thin: studies with weak programme descriptions where discerning the theory would have been problematic.<sup>27</sup>

The other article, Waldron et al.,<sup>28</sup> describes how they considered richness:

We assessed it by scoring the articles in relation to the richness relative to the research

questions. To score highly an article should provide sufficient details in relation to how the approach used was expected to work; documenting the process and explaining contextual factors that influenced implementation and/or outcomes. We rated the richness as follows:

0 = nothing of interest, not focused on design, implementation or use,

1 = limited data of interest, likely to appear in other articles,

2 = limited data of interest, but quick to extract it and could add weight to findings,

3 = some good quality data,

4 = much valuable data.

The richness assessment at full text reading allowed us to identify the articles with the most potential for providing rich data.<sup>28</sup>

### 3.3 | Rigour

While rigour was reportedly considered in 78% of the realist reviews ( $n = 57$ ), half of these did not provide any explanation of how it was assessed. When authors did describe the assessment process applied, they included scales created by the research team or their own judgement ( $n = 10$ ) or validated quality assessment tools such as: MMAT ( $n = 9$ ), CASP ( $n = 7$ ), Cochrane ( $n = 4$ ), NIH quality assessment ( $n = 1$ ), ROBINS-I ( $n = 1$ ), CCAT ( $n = 1$ ), AACODS ( $n = 1$ ), JBI ( $n = 1$ ) and Newcastle Ottawa Scale ( $n = 1$ ). The tools listed here, except AACODS, are applied to peer-reviewed literature. While AACODS can be applied to grey literature such as reports and unpublished dissertations, it is still aimed at academic literature.

Rigour was interpreted and applied in various ways, from considering the methodological conduct of the included article to the more granular level of the evidence included in a review. Price et al.<sup>29</sup> assessed rigour ‘at the level of the included data (where needed) and programme theory’. In another example, Grünwald et al.<sup>30</sup> assessed the overall document for quality using formalised checklists and appraised individual CMOCs by ‘assessing the set of documents that contributed data to each CMOC in relation to. ...the quality of their contribution to the CMOC (as each included document may have contributed a different type of data)’. In other cases, reviews acknowledged that RAMESES does not recommend using a formal checklist. Kinsey et al.<sup>31</sup> suggest:

Each fragment of evidence is considered within the synthesis as a whole, rather than in isolation – relevant data that has lower

trustworthiness may be supported by other data with higher trustworthiness (for example, a theory put forward in a discussion section may be supported by empirical data in another paper). The quality of the review as a whole is based on this data trustworthiness, but also on the coherence of the theories developed.<sup>31</sup>

As an additional example, Morton et al.<sup>24</sup> describe guiding questions that could be considered in assessing rigour:

This is an assessment of the likely validity and reliability only of the relevant data contained in an article, not an assessment of the rigour of a study or intervention programme as a whole. Useful questions might include: Is this data likely to be biased? Is it dealt with critically? Is it from a real-world example or theoretical speculation? Was the data gathered in some depth over time or in a quick “snapshot”? Is it safe to generalise from this data?<sup>24</sup>

## 4 | DISCUSSION

RAMESES guidelines emphasise the importance of transparently detailing the overall synthesis and appraisal processes to clarify the logic guiding a review.<sup>16,21</sup> In order to balance this transparent reporting without creating an undue burden on readers and reviewers, an updated reporting checklist for realist reviews based on the RAMESES publication standards<sup>16</sup> should encompass minimum reporting requirements for the evidence appraisal. This could include what criteria was applied in the appraisal process and how it was conducted.

The snapshot of realist reviews published in 2021 demonstrates a need for clearer and more transparent descriptions of how relevance, richness, and rigour are applied and how these appraisals impact the next steps of the review. Relevance was the most cited and transparently applied quality assessment criterion. Explicitly considering the richness of sources is a newer addition to the realist review methodology and has expectedly been carried out in varying degrees. Rigour was the least described, yet consistently applied across the included studies. While we observed numerous terms used to describe rigour across guidelines (including credible, plausible, believable, trustworthy and coherent), Wong<sup>32</sup> outlines two aspects of rigour: (1) trustworthiness and (2) coherence and plausibility.

Given the objective of this paper is to clarify and not confuse, we suggest ‘rigour’ should refer to *the*

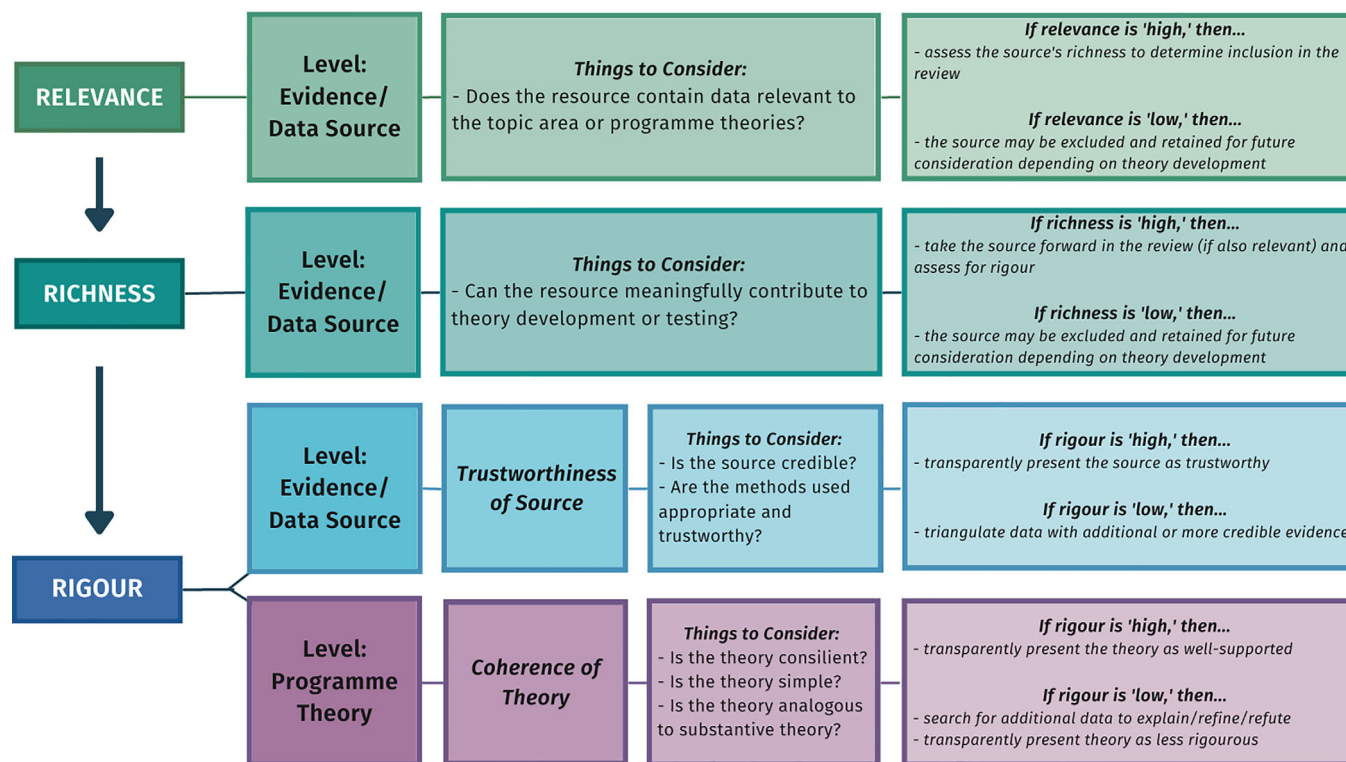


FIGURE 2 Proposed considerations for applying relevance, richness and rigour appraisals in a realist review. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

*trustworthiness of the evidence source and the coherence of the programme theory.* We suggest this appraisal criteria for rigour is sufficiently robust without adding the term 'plausibility' into the mix. To help assess these three R's and aid transparency in reporting, we unpacked these terms and offered a process by which reviewers could apply them (Figure 2).

Regarding trustworthiness, there is a pervasive positivist classification of evidence suggesting that some evidence is more trustworthy than others. For example, typical evidence classification describes randomised clinical trials (RCTs), experimental studies, and meta-analyses of RCTs at the top of the hierarchy, while experts' opinions are at the bottom.<sup>33</sup> Based on this, the positivist understanding is that: Level of evidence + Quality of evidence = Strength of evidence.<sup>34</sup> This implies that the information obtained from highly systematised meta-analyses of RCTs, experimental studies and RCTs contribute more trustworthy evidence. This notion of a hierarchy of evidence does not apply to realist reviews. In realist reviews, all sources of evidence are considered to be of equal standing provided they contribute to the development of mechanism-based explanatory theories. Different sources of information make different contributions to theory development in realist reviews. For example, RCTs can contribute to establishing a pattern of behaviour, while qualitative studies contribute to identifying

mechanisms, contexts and their causal relationships.<sup>12</sup> Realist reviews do not support a hierarchy of evidence based on research design. Therefore, judgements of trustworthiness are applied at the 'data level' (i.e. is the data credible?).

Assessing rigour at the programme theory level would involve an assessment of the resulting theory's explanatory coherence. This can be done by determining whether the theory offers a better explanation of the available data than its rivals.<sup>35–37</sup> Thagard<sup>36</sup> describes a theory's explanatory coherence as: explaining a greater range of the data, being simple, and aligning to credible and existing theories. To assist reviewers in conceptualising and applying rigour in a realist review we have provided a box of questions and answers (see Box 1).

The snapshot of realist reviews published in 2021 suggests that researchers may be more comfortable with assessments of relevance and to some extent, richness. However, rigour assessments seem elusive based on this proposed two-pronged approach (data source level and theory level) described in Figure 2. RAMESES 'Excellent' quality standards category notes that 'selection and appraisal demonstrate sophisticated judgements of relevance and rigour within the domain'.<sup>21</sup> This emphasises the researchers' role in selecting and appraising evidence should follow structure, content and clear logic, but will also require insights and judgement from the team. Even with provided guidelines, evidence appraisals in realist

**BOX 1 Rigour Q&A.**

**Q1:** Is rigour assessed at the level of the evidence source or the programme theories developed?

**A:** *Rigour should be assessed at both the evidence source and programme theory levels.*

**Q2:** What does ‘rigour’ look for?

**A:** *At the evidence source level, we are asking if the data are **trustworthy**. This may consider the methodological process and credibility of the source. At the theory level, we are asking if the theory is **coherent**. A coherent theory is consilient (explains the data), simple (makes few assumptions), and analogous to substantive theory (aligns with existing credible theories).*

**Q3:** What is the impact or influence of doing a rigour assessment?

**A:** *The purpose of assessing rigour during a realist review is to ensure that the data and theory are rigorous so that the resulting recommendation(s) can inform evidence-based practice. When the rigour of the evidence source is low (i.e. the data are not trustworthy), reviewers can overcome this by triangulating the data with additional and/or more credible sources by revisiting previously excluded sources or conducting a new search for evidence. When the rigour at the theory level is low (i.e. the theory is not coherent), reviewers can search for additional evidence to further explain, refine, or refute this theory to make it more rigorous. If the rigour cannot be improved through additional data or if this is beyond the scope of the review, reviewers are urged to be transparent and present the theory as having ‘less’ rigour. This can be followed with suggestions for further primary research to redress the gap in available evidence. However, if a review is intended to inform policy/practice, caution should be given in reporting any recommendations based on non-rigorous theory.*

**Q4:** What is the relationship between rigour at the evidence source level and the programme theory level?

**A:** *Rigorous data (trustworthy) does not necessarily equate to a rigorous theory (coherent). Less rigorous data or theory may call for additional evidence searches, as described in Q3.*

**Q5:** How do I be transparent with my methods for evidence appraisals, especially when considering a limited publication word count?

**A:** *Be explicit about the considerations given to both the data and theory-levels in appraisals and what was considered. Utilise boxes/tables/illustrations and supplementary materials to show how data was triangulated or supported where necessary.*

reviews may not always look the same, demonstrated by the first-hand accounts of two of the authors in Box 2.

## 5 | RECOMMENDATIONS

### 5.1 | Assessing relevance

1. Follow processes similar to the inclusion/exclusion criteria of ‘traditional reviews’ by considering the document’s relevance to the topic area, theory, or both.
2. Relevance criteria may change throughout the review depending on how theories develop.

### 5.2 | Assessing richness

1. This may happen after or in parallel to relevance but is likely to be more subjective. The purpose of richness is to

ensure the included documents provide a significant level of depth to meaningfully contribute to theory building.

2. It is worth cautioning reviewers on what it means to apply a quantitative rating system to articles to assess their richness. While a ‘positivist’ scale could be considered incompatible with realist epistemology, such scales may still be useful in helping reviewers make sense of a large amount of evidence that is common in a synthesis.

### 5.3 | Assessing rigour

1. As a realist review aims to develop explanatory theories, judgements of rigour must consider the data source as well as the theory level (Figure 2).
2. For assessments of rigour at the data level, those conducting realist reviews should be mindful that all information sources are deemed to

## BOX 2 Authors' reflections on developing, applying, and reporting relevance, richness and rigour criteria.

### • Author A: The challenge of determining approaches to richness and rigour (11)

*When determining the criteria for selecting sources to be included in my realist review and assessing them, I struggled to find detailed examples in the literature of how to go about this. I developed relevance criteria as I would for a traditional systematic review, and built off of this for 'richness' criteria by asking more specific questions about the level of depth or detail the source provided for each of the topic areas of my inclusion criteria.*

*My process for considering rigour was not as straightforward. I originally used the JBI checklists to help contextualise the trustworthiness of data sources and documented the responses to each question for each source. As my knowledge and understanding of realist epistemology and the purpose of rigour assessments in a realist review developed, I moved to reflect on questions around the credibility of the source (which I did independently) and on the coherence of the theories the CMOCs informed (which I did through gauging an expert advisory committee's feedback on specific questions relating to the extracted CMOCs). However, one of the limitations of my realist review is that it was challenging to document this thought process as effectively and transparently as I would have liked to. While it was much clearer to document my relevance and richness criteria and where/why documents did or did not meet this, in the future I would plan to develop and document more explicit considerations for rigour geared towards my review (at **both** the evidence source and theory level).*

### • Author B: Adapting rigour criteria to assess non-traditional data sources (20)

*I found the questions posed in existing quality appraisal checklists and tools were difficult to answer for non-academic data sources like newspapers, blogs or extracts from social media. Furthermore, I found the answers became perfunctory and did not help me advance my reasoning. I was left with a 'so what?', 'what do I do now?' As a result I created my own criteria for assessing rigour in non-academic sources which was guided and adapted from Miles, Huberman and Saldana (1994).<sup>38</sup> This criteria focused on the reliability and validity of the data. Importantly, to my mind, it allowed for extracts of data from potentially untrustworthy sources to be included in the review if their addition was deemed to add value to theory development and the bias it introduced was balanced or countered by additional data.*

- *Look for author effects and bias—does this effect/bias undermine the credibility of what is being said? If so, what do you plan to do e.g. exclude/include?—If include, how are you going to redress the balance?*
- *Check for representativeness*
- *Get feedback from participants—a key component of realist review is to sense check with stakeholders*
- *Triangulating*
- *Look for ulterior motives*
- *Look for deception*
- *Check against alternative accounts—in realist review terms, I took this to mean look for rival theory, disconfirming case, counterfactuals, outliers etc.*
- *Sense check with research team and peers—As well as discussing the data and theory development with my research team, I presented a lot of my early work at training conferences to see if my inferences were logical or were too much of a leap.*

*I liked this criteria because it gave me an idea of what to do if the answer to the question was a yes or a no – For example:*

*Q: Is there author bias in this blog?*

*A: Yes*

*Q: Is it still worth including, if yes, why and what are you going to do about it?*

*A: Yes, this person has been living with pain for 20 years and wants the NHS to improve its service. Their opinions are valid and they have a good way of expressing themselves, I want to include the data. I am going to sense check with stakeholders and find other sources to refine/refute/reinforce.*



**BOX 3 To ‘tool’ or not to ‘tool’?**

*In the process of developing this manuscript, the authors discussed a number of ideas around the processes of how realist reviewers can go about conducting assessments for rigour—at both the data and theory level. A particular area of discussion, which has also been debated in the RAMESES forum,<sup>39</sup> was the use of existing critical appraisal tools to assess the trustworthiness of data. This box presents the arguments for and against the application of existing tools or checklists in realist reviews and invites the reader to also contribute to the discussion.*

**Argument for using existing tools/checklists:**

- Realist reviews call for assessing the quality/trustworthiness of the data, existing tools/checklists could be helpful by providing initial structures, support or ways to think about the data source
- Existing tools would not be the only component of assessing rigour, but could contribute to one aspect of looking at the source and data from different perspectives
- However, not all components of the tool might be necessary, and that they do not necessarily need to be applied and reported upon as they are in traditional reviews (in this way, realist reviewers can make the tool fit for purpose)

*→ In realist approaches, we do not reject survey data because they are positivist, we use them to help support our understanding. It is therefore not **what tools we use, but how we use them**. We would not reject or rank documents in a realist review based on the tool (like one might in a positivist review), but use the tool as one approach to help understand the trustworthiness of the data.*

**Argument against using existing tools/checklists:**

- Not epistemologically aligned with realist approaches
- Not fit for purpose for realist reviews as existing quality appraisal tools provide an overall assessment, rather than considering ‘nuggets’ or specific pieces of data
- Existing appraisal tools and checklists do not exist for non-traditional data sources, which can be equally valuable in a realist review

*→ Instead, descriptive assessments of trustworthiness should be made explicit through throughout the review and clear in the write-up of findings, with clear implications of how this impacts the rigour of the theory.*

***In conclusion, there are multiple ways in which realist reviewers could assess trustworthiness. What is important is that reviewers are clear on how they appraised this trustworthiness and how it influenced their review.***

be of equal standing and consideration in realist reviews.

- a. While a number of published realist reviews applied existing critical appraisal tools, reviewers should be cautioned against excluding sources in a realist review based on ‘traditional’ hierarchies of evidence (see Box 3).
- b. We recommend reviewers draw on their own research training and judgement to determine if the data are trustworthy and the source credible. These decisions and assessments should be documented transparently, which could be done by designing ‘bespoke tools’ or lists of guiding questions (as seen in Box 2), which could include

questions or considerations posed by existing tools/checklists.

3. Furthermore, descriptive assessments of trustworthiness at the data level should be made explicit throughout the review and clear in the write up of findings, with clear implications of how this impacts the rigour of the theory (i.e. *Was more evidence needed to support the theory because this particular piece of evidence was low in terms of trustworthiness?*).
4. In order to assess rigour at the theory level, we recommend the realist reviewer considers the explanatory coherence of the programme theory<sup>36</sup> and transparently reports which theories are well-supported (more rigorous) and less well-supported.

## 6 | CONCLUSION

We recognise that realist reviews need to be operationalised differently, and what works for some may not work for others. However, there is a need for greater transparency and clarity in the application and reporting of relevance, richness, and rigour appraisals in realist reviews. As demonstrated by this snapshot, only a third of the publications provided detail on the rigour assessments they conducted, while less than a fifth of papers reported on how richness was determined. Conducting appropriate evidence and theory appraisals within realist reviews will support higher more robust findings, thus having greater application and influence on policy and service delivery. It should be noted that in the process of writing this manuscript, the authors had a number of discussions and different opinions on the best ways forward. As a result, we hope that by discussing these initial considerations for evidence appraisals in realist reviews, we can open up this debate around the best approaches to the entire research community and not stifle innovation or connoisseurship in realist reviews.

### AUTHOR CONTRIBUTIONS

Sara Dada, Sonia Dalkin, Brynne Gilmore, Rebecca Hunter, Ferdinand C. Mukumbang contributed to the drafting and revisions of this article and have approved it for publication.

### CONFLICT OF INTEREST STATEMENT

The authors have no relevant financial or non-financial competing interests to report and no conflicts of interest.

### ACKNOWLEDGMENT

Open access funding provided by IReL.

### DATA AVAILABILITY STATEMENT

All data considered are made available through the manuscript and/or supplementary material.

### ORCID

Sara Dada  <https://orcid.org/0000-0003-3910-1856>

Sonia Dalkin  <https://orcid.org/0000-0002-3266-5926>

Brynne Gilmore  <https://orcid.org/0000-0003-4496-9254>

Rebecca Hunter  <https://orcid.org/0000-0001-7837-7260>

Ferdinand C. Mukumbang  <https://orcid.org/0000-0003-1441-2172>

### REFERENCES

- Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review - a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. 2005;10(1\_suppl):21-34.
- Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research. *Implement Sci*. 2012;7(1):33.
- Jagosh J. Realist synthesis for public health: building an ontologically deep understanding of how programs work, for whom, and in which contexts. *Annu Rev Public Health*. 2019;40:361-372.
- Boyko JA, Riley BL, Willis CD, et al. Knowledge translation for realist reviews: a participatory approach for a review on scaling up complex interventions. *Health Res Policy Syst*. 2018;16(1):101.
- Ashworth HC, Dada S, Buggy C, Lees S. The importance of developing rigorous social science methods for community engagement and behavior change during outbreak response. *Disaster Med Public Health Prep*. 2021;15:685-690.
- Linsley P, Howard D, Owen S. The construction of context-mechanisms-outcomes in realistic evaluation. *Nurse Res*. 2015;22(3):28-34.
- Pawson R, Tilley N. *Realistic Evaluation*. Sage; 1997.
- Jagosh J. Retroductive theorizing in Pawson and Tilley's applied scientific realism. *J Crit Realism*. 2020;19(2):121-130.
- Mukumbang FC. Retroductive theorizing: a contribution of critical realism to mixed methods research. *J Mixed Methods Res*. 2021;17:15586898211049847.
- Klingberg S, Adhikari B, Draper CE, et al. Engaging communities in non-communicable disease research and interventions in low-and middle-income countries: a realist review protocol. *BMJ Open*. 2021;11(7):e050632.
- Dada S, De Brún A, Banda EN, Bhattacharya S, Mutunga Z, Gilmore B. A realist review protocol on communications for community engagement in maternal and newborn health programmes in low- and middle-income countries. *Syst Rev*. 2022;11(1):201.
- Mukumbang FC, Kabongo EM, Eastwood JG. Examining the application of Retroductive theorizing in realist-informed studies. *Int J Qual Methods*. 2021;20:16094069211053516.
- Cooper C, Lhussier M, Shucksmith J, Carr SM. Protocol for a realist review of complex interventions to prevent adolescents from engaging in multiple risk behaviours. *BMJ Open*. 2017;7(9):e015477.
- Mukumbang FC, De Souza D, Liu H, et al. Unpacking the design, implementation and uptake of community-integrated health care services: a critical realist synthesis. *BMJ Glob Health*. 2022;7(8):e009129.
- Pawson R. *Evidence-Based Policy: a Realist Perspective*. Sage; 2006.
- Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. *BMC Med*. 2013;11(1):21.
- Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC Med Res Methodol*. 2013;13(1):118.
- Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist synthesis: an introduction. ESRC res methods. *Dent Prog*. 2004;2.
- Wong G, Greenhalgh T, Westhorp G, Pawson R. Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (realist and meta-narrative evidence syntheses - evolving standards) project. *NIHR Journals Library*. 2014;2(30).
- Hunter R, Gorely T, Beattie M, Harris K. Realist review. *Int Rev Sport Exerc Psychol*. 2022;15(1):242-265.
- Wong G, Greenhalgh T, Westhorp G, Pawson R. *Quality Standards for Realist Syntheses and Meta-Narrative Reviews*. RAMESES; 2014:24.

22. Pawson R. Digging for nuggets: how 'bad' research can yield 'good' evidence. *Int J Soc Res Methodol*. 2006;9(2):127-142.
23. Epstein DS, Zemski A, Enticott J, Barton C. Tabletop board game elements and gamification interventions for health behavior change: realist review and proposal of a game design framework. *JMIR Serious Games*. 2021;9(1):e23302.
24. Morton T, Wong G, Atkinson T, Brooker D. Sustaining community-based interventions for people affected by dementia long term: the SCI-Dem realist review. *BMJ Open*. 2021; 11(7):e047789.
25. Ulrich H-S, Kohler E, Spallek J, Richter M, Clauß D, Mlinarić M. Explaining psychosocial care among unaccompanied minor refugees: a realist review. *Eur Child Adolesc Psychiatry*. 2022;31: 1857-1870.
26. Pearce C, Wong G, Kuhn I, Barclay S. Supporting bereavement and complicated grief in primary care: a realist review. *BJGP Open*. 2021;5(3).
27. Calderón-Larrañaga S, Milner Y, Clinch M, Greenhalgh T, Finer S. Tensions and opportunities in social prescribing. Developing a framework to facilitate its implementation and evaluation in primary care: a realist review. *BJGP Open*. 2021;5(3).
28. Waldron C, Cahill J, Cromie S, et al. Personal electronic Records of Medications (PERMs) for medication reconciliation at care transitions: a rapid realist review. *BMC Med Inform Decis Mak*. 2021;21(1):307.
29. Price T, Wong G, Withers L, et al. Optimising the delivery of remediation programmes for doctors: a realist review. *Med Educ*. 2021;55(9):995-1010.
30. Grünwald LM, Duddy C, Byng R, Crellin N, Moncrieff J. The role of trust and hope in antipsychotic medication reviews between GPs and service users a realist review. *BMC Psychiatry*. 2021;21(1):390.
31. Kinsey D, Lang I, Orr N, Anderson R, Parker D. The impact of including carers in museum programmes for people with dementia: a realist review. *Arts Health*. 2021;13(1):1-19.
32. Wong G. *Data Gathering in Realist Reviews: Looking for Needles in Haystacks. Doing Realist Research*. Sage; 2018:131-146.
33. Poe S, White KM. *Johns Hopkins Nursing Evidence-Based Practice: Implementation and Translation*. Sigma Theta Tau; 2010.
34. Melnyk BM, Fineout-Overhold E. *Evidence-Based Practice in Nursing & Healthcare: A Guide to Best Practice*. Lippincott Williams & Wilkins; 2022.
35. Bhaskar R. *Scientific Realism and Human Emancipation*. Routledge; 2009.
36. Thagard P. Explanatory coherence. *Behav Brain Sci*. 1989;12(3): 435-467.
37. Ward T, Clack S, Haig BD. The Abductive theory of method: scientific inquiry and clinical practice. *Behav Chang*. 2016; 33(4):212-231.
38. Miles MB, Huberman AM. *Qualitative data analysis: An expanded sourcebook*. Sage; 1994.
39. *Three Rs not two (Rigour/Relevance)? RAMESES*. RAMESES forum; 2019.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Dada S, Dalkin S, Gilmore B, Hunter R, Mukumbang FC. Applying and reporting relevance, richness and rigour in realist evidence appraisals: Advancing key concepts in realist reviews. *Res Syn Meth*. 2023; 14(3):504-514. doi:[10.1002/jrsm.1630](https://doi.org/10.1002/jrsm.1630)