

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



Authors: Sara Dada, Sonia Dalkin, Brynne Gilmore, Rebecca Hunter, Ferdinand C. Mukumbang
Presentation By: Sara Dada (sara.dada@ucdconnect.ie)

17 October 2023
Realist Hive



UCD School of Nursing, Midwifery and Health Systems
UCD College of Health and Agricultural Sciences

Overview

Background &
Purpose

1

Overview of
Current Practice

3

Initial Reflections &
Future Directions

5

2

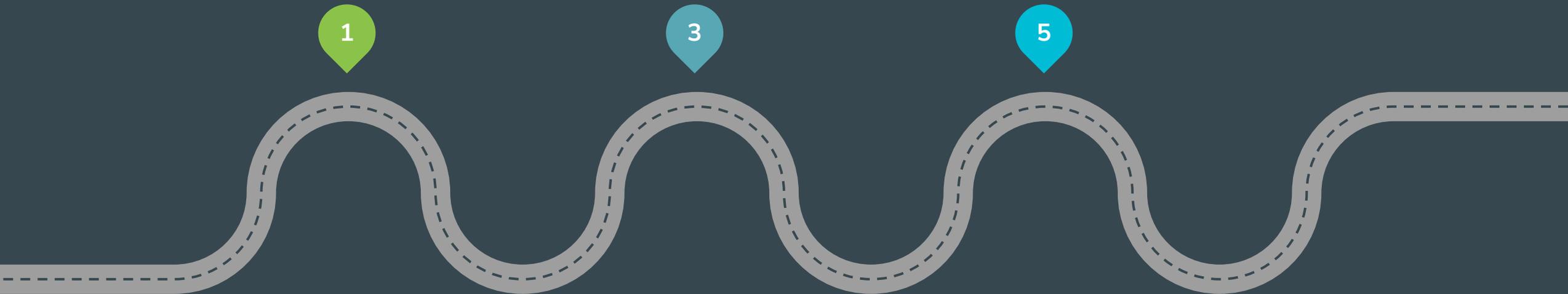
Approach

4

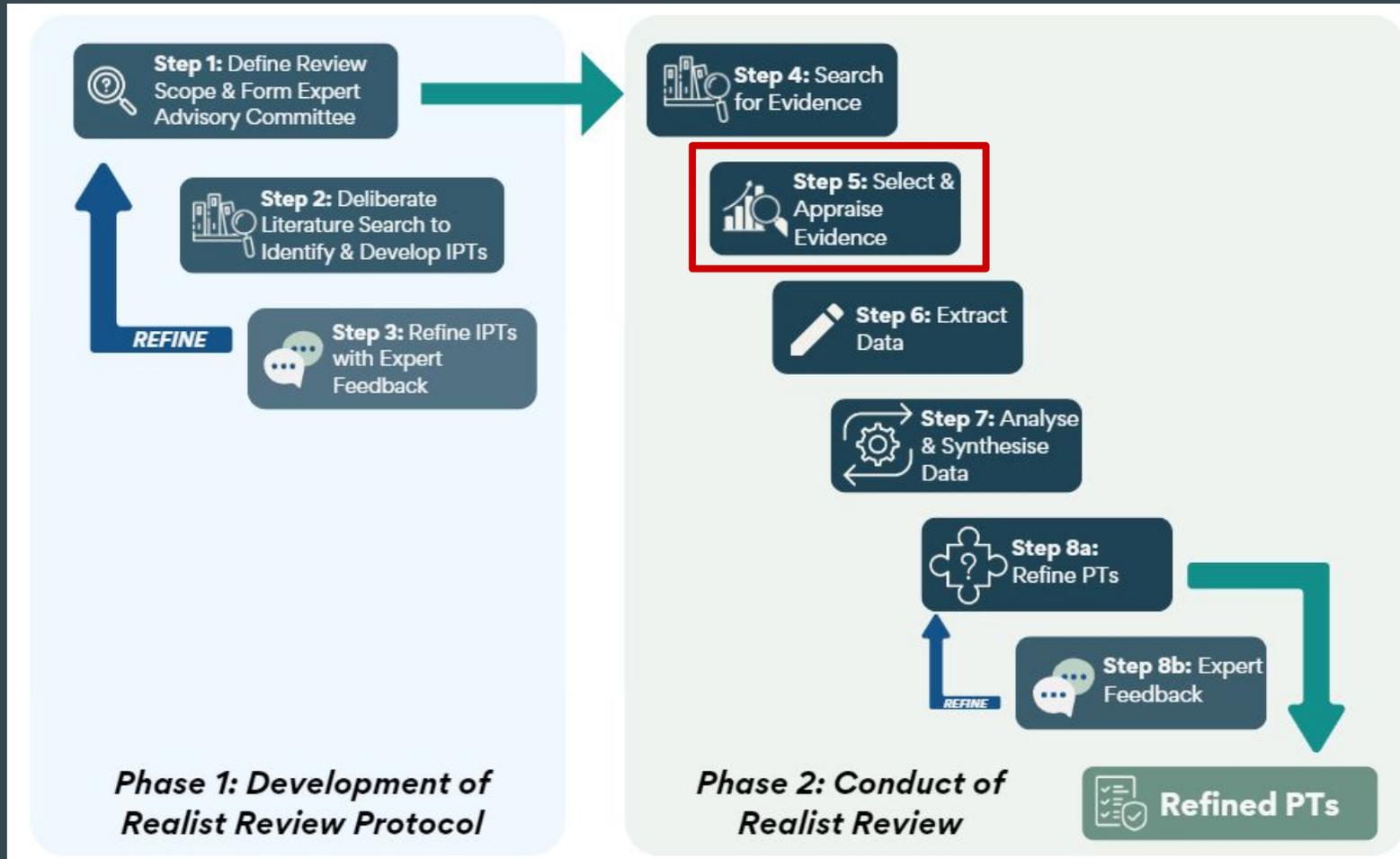
Recommendations:
Tips & Tricks

6

Q&A



Background: What is the realist review?



- Form of secondary analysis that synthesises evidence, *focused on explaining generative causation* - or *why/how* an intervention works (or doesn't), rather than *if* it works.
- Similarly to traditional reviews, evidence is screened and appraised - this is the step we are focusing on today

Background: What is evidence appraisal?

- Realist reviews differ from a traditional review by:
 - Encouraging use of data from a wide variety of sources
 - Inclusion based on ability to contribute to theory building or testing
 - Iterative process of screening and appraising
- In realist reviews, we often read about data sources being screened or appraised by relevance, richness and rigour

Relevance

Defined by RAMESES as “*whether [the data] can contribute to theory building and/or testing*”

- Constantly negotiated in realist reviews, as theory develops and changes
- Recommended that reviewers retain searches and results to revisit
- Can result in large amounts of ‘relevant’ sources, but their depth and contributions might vary

Richness

Booth *et al.* suggested evidence be appraised for ‘Richness’ to assess whether sources can meaningfully contribute to theory -

- **Conceptual richness:** degree of theoretical & conceptual development that explains how an intervention is expected to work
- **Contextual thickness:** sufficient detail that enables reader to i) establish what is occurring in the intervention and wider context & ii) infer whether findings can be transferred to other people, places, situations or environments

Rigour

Defined by RAMESES as “*whether the method used to generate that particular piece of data is credible and ‘trustworthy.’*”

- Multiple terms used within prominent guidelines: *credible, plausible, believable, trustworthy, coherent*
- Since range of sources encouraged, potential difficulty navigating across paradigms + non-academic sources
- General debate and lack of consensus on how to assess for rigour

While there are publication standards and guidelines for the conduct of realist reviews, published reviews often provide **minimal detail** regarding how they have **selected and appraised included evidence**. Our goal was to discuss some of the challenges faced in this process and to **provide pragmatic suggestions** for the realist reviewer.

Approach: What we did and why

- Problem: Lack of clarity around evidence appraisals, especially for rigour, in realist research
- Goal: To provide a snapshot of current practice to understand how appraisals are being done, and use this to provide examples and guidance for future reviews
- Approach: Conducted a basic search on PubMed for ‘realist reviews/syntheses’ AND ‘health systems’ published in 2021
 - All realist reviews/syntheses included
 - Reviewed and extracted information relating to how authors reported relevance, richness and rigour

Findings

- **73** published reviews identified
 - **67 (91%)** explicitly stated relevance criteria
 - **57 (78%)** acknowledged conducting some rigour assessment
 - **28 (38%)** of all papers described this process
 - Richness was the least described: **33 papers (45%)** acknowledged doing some form of richness assessment
 - but only **13 papers (18%)** provided an explanation of how
-

Relevance

Findings

- Mirrors inclusion/exclusion criteria approach used in traditional systematic reviews; typically provided in the main text or supplementary file
- Applied in two ways:
 - 1) Relevant to topic area/content of interest
 - 2) Evidence for theory development, refinement or testing

Significance

- The large majority of reviews report relevance considerations and provide details
- Limited contention or confusion within assessing for relevance and clarity in reporting

Relevance

examples from published reviews

1. [Epstein et al. 2021](#):

“Studies were expected to have adequate relevance to build the program theory Any paper that was not directly assessing a tabletop game intervention or not measuring a behavior change outcome was excluded at this point.”

2. [Grünwald et al. 2021](#):

Table 2 Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Adults (age 18 and above)	Service users currently under section (Mental Health Act, Forensic, Community Treatment Order) or currently in crisis or studies discussing Crisis services (Home Treatment Team etc)
Diagnosis of Psychosis, schizophrenia, psychosis like symptoms (SMI)	Animal studies
Medication reviews, care and treatment of service users diagnosed with SMI	Physical health reviews only, which do not include factors around treating SU or have medication reviews alongside
Published after 1954 (year the first antipsychotic was introduced)	Studies discussing prescription of non-antipsychotic medications
Published in English language	Studies from low- and middle-income countries
All study methodologies	Studies discussing the prevalence, and treatment of side effects by adding other (non-antipsychotic) medications
Prescription of antipsychotic medication in primary care	Studies discussing the prevalence or validity of a diagnosis of severe mental illness Off – label prescribing Excluded later: <ul style="list-style-type: none">• Studies investigating bipolar disorder• Clozapine

Richness

Findings

- Description of how richness was applied was often interwoven with relevance (*'how relevant'*)
- When reported, different approaches:
 - *Examples:* 'thick/thin' continuum, categories of high versus low relevance, 'traffic light' system, 5 point scale or ranking systems

Significance

- Many reviews apply *some* richness considerations, even if they aren't explicitly calling it this
- When it is explicitly applied, there is limited reporting and transparency of how this is done/how judgements are determined

Richness

examples from published reviews

1. [Calderon-Larrañaga et al. 2021:](#)

“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 theory would have been problematic.”

Richness

examples from published reviews

2. [Waldron et al. 2021:](#)

“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”

Rigour

Findings

- Reviewers often used scales created by own research team (n=10) **OR** existing validated quality assessment tools such as MMAT (n=9), CASP (n=7), Cochrane (n=4), or others (n=6)
 - All tools reported except for one (AACODS) are applied exclusively to peer-reviewed literature
- Rigour interpreted in two ways:
 1. Methodological conduct/quality of the included source → **data/evidence level**
 2. Strength of evidence (e.g. CMOCs) included in the review → **theory level**

Significance

- Overall, rigour not applied consistently
- Limited reporting of how judgements determined
- Limited reporting on influence of rigour assessment on next steps of review or implications for data/findings
- RAMESES and others caution on the use of formal checklists, but even the guidelines use multiple descriptions/approaches to rigour (credible, plausible, believable, trustworthy and coherent)

Rigour

examples from published reviews

1. Multiple examples of using formalised checklists
 2. [Grünwald et al. 2021](#) used checklists to assess the included document 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).”*
 3. [Morton et al. 2021](#):
“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?”
-

Ongoing Discussions & Debate around Rigour

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 consistent (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.

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,³⁹ 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.

Summary

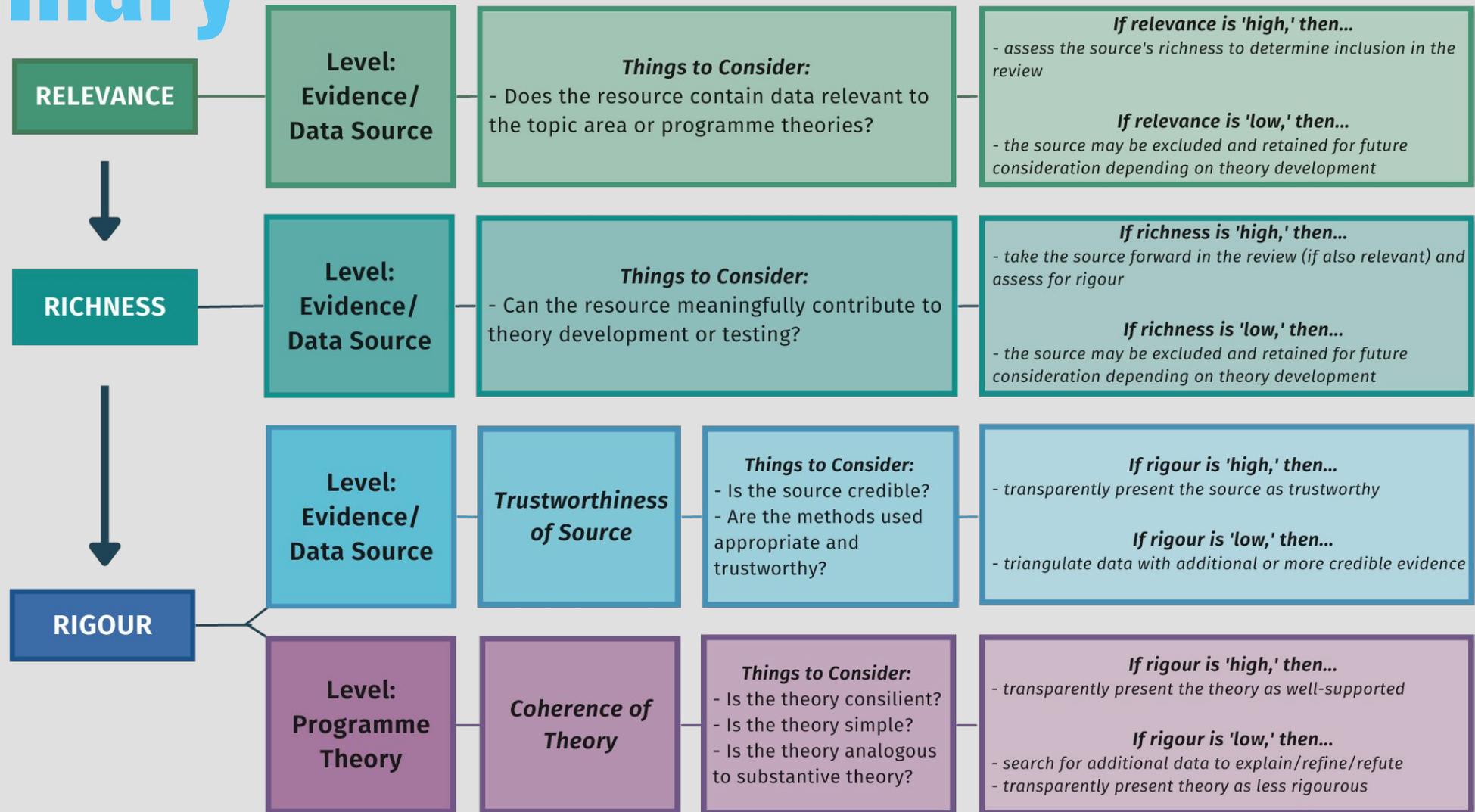


Figure 2 from [Dada et al. \(2023\) Applying and reporting relevance, richness and rigour in realist evidence appraisals: Advancing key concepts in realist reviews](#)

Recommendations: Tips & Tricks

Assessing Relevance:

- Follow process similar to traditional review's inclusion/exclusion criteria, considering relevance to topic, theory or both
- Criteria may change throughout review, retain records for iterative searching & appraising

Assessing Richness:

- Can occur in parallel or after relevance; likely more subjective but aims at ensuring documents provide significant level of depth to contribute to theory development
- Caution on applying 'positivist' rating systems, though may be helpful in making sense of large number of sources

Recommendations: Tips & Tricks

Assessing Rigour:

- No evidence hierarchies applied in realist reviews → Caution on excluding sources based on existing critical appraisal tools or evidence hierarchies
- More transparent assessments should be provided in write-up of findings, **including** how rigour impacts the review and theories
- Assessments should consider evidence/data source level **and** theory level
 - **At data source level:** consider the trustworthiness and credibility of source ([Wong 2018](#)), draw on own judgements, bespoke tool, or get support from existing tool (with caution)
 - **At theory level:** consider explanatory coherence (consilient, simple, analogous to substantive theory) of the theory and transparently report which theories are more or less supported

Recommendations: Bringing it all together

Experiences & reflections from applying these insights to a realist review: [Dada et al. 2023](#)

Table 1. Relevance, richness and rigor

Relevance	Richness	Rigor
<p>Include:</p> <ol style="list-style-type: none"> 1. Any study design/article type, except for reviews/meta-analyses 2. Documents ‘about CE being used for MNH programmes, AND’ 3. Describes programmes in an ‘LMIC, AND’ 4. Describes ‘the CE communication activities, processes or tools’ that were used <p>Exclude:</p> <ol style="list-style-type: none"> 1. Conference proceedings, studies that lack full-text, reviews/meta-analyses (secondary data), OR 2. Documents that are not about MNH programmes or outcomes, OR 3. Documents about programmes in high-income countries, OR 4. Do not describe CE communication activities, processes or tools, or only describe CE that is conducted for the purpose of research 	<p>High:</p> <p>Makes several contributions towards theory development. This includes, but is not limited to, insights related to generative causation and/or components of context, mechanisms and outcomes. There is a rich description of processes and/or context so that regardless of quality (rigor), there is sufficient content to build, refine and refute PTs</p> <p>Moderate:</p> <p>Makes one or two contributions towards theory development. Regardless of quality (rigor), content can contribute to inferences relating to PTs</p> <p>Low:</p> <p>Makes little contribution towards theory development and/or results or evidence lack credibility. There is a limited description of potential contexts, mechanisms or outcomes that could contribute to CMOCs and/or theory refining</p> <p>None:</p> <p>While relevant to the review subject matter, papers make no contribution to theory development. It does not include any description of the processes or any content relevant to the initial PTs, or potential contexts, mechanisms or outcomes</p>	<p>The trustworthiness of the data source and coherence of the theory it informed was considered to contextualize the contributions of the evidence. While this did not involve the use of a formalized appraisal tool or checklist, this included the researcher reflecting on and considering the feedback from the expert advisory committee on questions such as:</p> <ul style="list-style-type: none"> • Is the information plausible? • Are the CMOCs extracted justifiable? • Are the findings put forward coherent? <p>While no documents were excluded based on rigor, this assessment contributes to building credible arguments to test and refine the PTs</p>

The processes of evidence appraisals applied in this review.

Reflections & Future Directions

- Likely **no 'one-size-fits-all'** approach
 - *Evidence selection & appraisal should make sense and be appropriate for the review*
 - **More transparency in reporting** how evidence appraisals were done, and what (if any) influence they had on the review, findings and implications
 - Continued debate and discussion around the best approaches for evidence appraisals, *especially in relation to rigour* - **to tool or not to tool?**
-

1. **Can the RRR method be used to triage papers for screening not just for assessing evidence?**
 - a. Adapt screening and appraisal process based on what's appropriate/feasible for the review

2. **What composition of a review team is recommended for a realist review and what training is available?**
 - a. Vary in composition and size - some include stakeholders or PPI representatives as authors; what's most appropriate depends on purpose/scope of review (and feasibility/resources)
 - b. [RAMESES for standards/guidelines](#)
 - c. Number of different *paid* modules/trainings (Oxford, Charles Darwin University, CARES) available

3. **Are quality appraisals mandatory for realist reviews? i.e use of quality appraisal tools**
 - a. *"An appraisal of the contribution of any section of data (within a document) should be made on two criteria: Relevance and Rigor"* ([Wong et al. 2013](#))
 - i. We suggest this can be done by drawing on own judgements, using bespoke tools, or getting support from existing tools (with caution)

Questions

Acknowledgements

- Thank you to the co-authors of this realist review for their contribution to the project and this presentation
- Slides by Brynne Gilmore & Sara Dada
- Manuscript reference:
Dada, S., Dalkin, S., Gilmore, B., Hunter, R. and Mukumbang, F.C., 2023. *Applying and reporting relevance, richness and rigour in realist evidence appraisals: Advancing key concepts in realist reviews*. Research synthesis methods, 14(3), pp.504-514.

Received: 1 November 2022 | Revised: 8 February 2023 | Accepted: 26 February 2023
DOI: 10.1002/jrsm.1630

RESEARCH IN BRIEF

Research Synthesis Methods WILEY

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

Sara Dada^{1,2} | Sonia Dalkin^{3,4} | Brynne Gilmore^{1,2} | Rebecca Hunter⁵ | Ferdinand C. Mukumbang⁶

¹UCD 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
²School of Nursing Midwifery and Health Systems, University College Dublin, Dublin, Ireland
³Faculty of Health and Life Sciences, Northumbria University, Newcastle Upon Tyne, UK
⁴Fuse (The Centre for Translational Research in Public Health), Northumbria University, Newcastle Upon Tyne, UK
⁵Department of Nursing and Midwifery, University of Highlands and Islands, Inverness, UK
⁶Department of Global Health, University of Washington, Seattle, Washington, USA

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

Correspondence

References

- Dada, S., Dalkin, S., Gilmore, B., Hunter, R. and Mukumbang, F.C., 2023. Applying and reporting relevance, richness and rigour in realist evidence appraisals: Advancing key concepts in realist reviews. *Research synthesis methods*, 14(3), pp.504-514.
- Grünwald, L.M., Duddy, C., Byng, R., Crellin, N. and Moncrieff, J., 2021. The role of trust and hope in antipsychotic medication reviews between GPs and service users a realist review. *BMC psychiatry*, 21, pp.1-14.
- Epstein, D.S., Zemski, A., Enticott, J. and Barton, C., 2021. Tabletop board game elements and gamification interventions for health behavior change: realist review and proposal of a game design framework. *JMIR Serious Games*, 9(1), p.e23302.
- Calderón-Larrañaga, S., Milner, Y., Clinch, M., Greenhalgh, T. and Finer, S., 2021. Tensions and opportunities in social prescribing. Developing a framework to facilitate its implementation and evaluation in primary care: A realist review. *BJGP open*, 5(3).
- Waldron, C., Cahill, J., Cromie, S., Delaney, T., Kennelly, S.P., Pevnick, J.M. and Grimes, T., 2021. Personal Electronic Records of Medications (PERMs) for medication reconciliation at care transitions: a rapid realist review. *BMC Medical Informatics and Decision Making*, 21(1), pp.1-17.
- Morton, T., Wong, G., Atkinson, T. and Brooker, D., 2021. Sustaining community-based interventions for people affected by dementia long term: the SCI-Dem realist review. *BMJ open*, 11(7), p.e047789.
- Wong, G., 2018. Data gathering in realist reviews: looking for needles in haystacks. *Doing realist research*. London: SAGE, pp.131-45.
- Dada, S., Aivalli, P., De Brunún, A., Barreix, M., Chelwa, N., Mutunga, Z., Vwalika, B. and Gilmore, B., 2023. Understanding communication in community engagement for maternal and newborn health programmes in low-and middle-income countries: a realist review. *Health Policy and Planning*, p.czad078.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. and Pawson, R., 2013. RAMESES publication standards: realist syntheses. *BMC medicine*, 11, pp.1-14.

Thank you!



UCD School of Nursing, Midwifery and Health Systems
UCD College of Health and Agricultural Sciences

Contact Information:
sara.dada@ucdconnect.ie
[@dadasara3](https://www.instagram.com/dadasara3)