Article:

Hunter, R., Booth, A. and Wood, L. (2026) Searching smarter, not harder: leveraging AI to enhance literature searches for theory-driven reviews—A methodological case study. BMC Medical Research Methodology. ISSN: 1471-2288

https://doi.org/10.1186/s12874-026-02814-3

UNIVERSITY OF LEEDS · University of Sheffield · UNIVERSITY of York

Searching smarter, not harder: leveraging AI to enhance literature searches for theory-driven reviews—A methodological case study

R. Hunter, A. Booth & L. Wood

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Searching Smarter, Not Harder: Leveraging AI to Enhance Literature Searches for Theory-Driven reviews. A methodological case study.

R. Hunter[1,2], A. Booth[3]. L. Wood[1]

[1]Department of Public Health and Sports Sciences, University of Exeter, Medical School Building, St Luke's Campus, Exeter, EX4 4QJ.

[2]School of Communities and Education, Northumbria University, Coach Lane Campus, Benton, Newcastle upon Tyne, NE7 7XA.

[3]School of Health and Related Research (ScHARR), The University of Sheffield, 30 Regent St, Sheffield City Centre, Sheffield, S1 4DA.

Dr Rebecca Hunter - https://orcid.org/0000-0001-7837-7260
Professor Andrew Booth - https://orcid.org/0000-0003-4808-3880
Dr Lianne Wood – https://orcid.org/0000-0003-1039-1642

Corresponding Author:

Dr Rebecca Hunter
Department of Public Health and Sports Sciences,
University of Exeter,
Medical School Building,
St Luke's Campus,
Exeter
EX4 4QJ, UK

Email: rebecca.j.hunter@northumbria.ac.uk

Abstract

Background:

Integrating artificial intelligence (AI) into literature searching has the potential to enhance research synthesis by improving the identification of conceptually rich or otherwise difficult-to-locate evidence. Theoretical or conceptual literature reviews, including realist reviews, often involve resource-intensive searches because they aim to trace nuanced ideas, mechanisms, or conceptual relationships across multiple sources. This case study illustrates the use of AI-powered tools to support and streamline such literature searching, using a realist review as an example.

Methods:

We applied AI tools—Scite and Undermind—in the context of a realist review to facilitate the identification of relevant studies. Seed papers and key informant papers guided the search, and a novel classification system (grandparent, parent, and child papers) was used to systematically organise studies for developing and refining theoretical constructs. Transparent screening procedures and decision-making frameworks were employed to ensure methodological rigour and reproducibility.

Results:

The integration of AI tools supported the retrieval of conceptually relevant literature and helped manage complex datasets. The classification system enabled structured organisation of studies, supporting iterative testing and refinement of theoretical constructs. The workflow demonstrated flexibility and adaptability, suggesting potential applicability beyond realist review

Conclusions:

Our findings suggest that AI-powered tools can support literature searching, particularly in identifying conceptually relevant studies. However, these tools do not replace the critical interpretive work required by researchers. Human judgement remains essential to assess relevance, evaluate nuanced concepts, and make informed decisions throughout the search process, with AI serving as a valuable adjunct rather than a substitute.

Keywords

Literature searches, review methodology, realist reviews, evidence appraisal, literature screening, artificial intelligence

1

2

1 **Background**

2 The advancement of artificial intelligence (AI) capabilities presents an opportunity to enhance
3 research processes, particularly in literature searching[1]. Capturing nuanced concepts, identifying
4 theoretically relevant papers, or tracking related studies can be time- and resource-intensive using
5 conventional methods. AI-powered tools offer opportunities to streamline these processes while
6 maintaining conceptual depth and rigour. This is particularly relevant for literature reviews with a
7 theoretical or explanatory focus, such as realist reviews.

8 Realist reviews aim to understand complex interventions by examining how context interacts with
9 underlying mechanisms to produce outcomes[2]. To do this, they often require broad, inclusive
10 searches to identify conceptually rich material to develop and test programme theories[3]. The "realist
11 search" must be systematic, transparent, and guided by established methodological principles yet also
12 iterative, fluid, and purposeful [4-7].

13 Instead of relying on pre-specified topic-based procedures, the realist search embraces multiple
14 diverse retrieval techniques throughout the review[4]. Berry picking[8], by which the searcher
15 'forages' using six search techniques: checking cited references: citation searching; searching relevant
16 journals; browsing proximate to items of interest; using subject-indexed databases; and author
17 searching[8] has enjoyed a renaissance in the specific realist guise of CLUSTER Searching.
18 Specifically, CLUSTER searching constitutes a systematic methodology for identifying a study
19 cluster (a group of related reports from a single project) by tracking Citations, Lead Authors,
20 Unpublished materials, Scholar searches, Theories, Early examples, and Related projects[5]. More
21 generally, Snowballing or Citation Chaining offers a method for following up citations of a key study
22 either backward (reviewing references) or forward (tracking subsequent citing articles) to find
23 contextually or theoretically proximate documents[7].

24 Beyond citation-based methods, realist searches often incorporate alternative sources and targeted
25 searches. Grey Literature searching seeks out non-peer-reviewed sources, such as policy documents,
26 reports, web-based evaluations, and manuscripts, to supply the necessary theoretical or contextual
27 detail missing from conventional literature[9]. More broadly, the use of general search engines (e.g.,
28 Google) to retrieve grey literature and discover programme theories has been suggested to be more
29 efficient than academic database searches for theory identification[9] while scholarly Search Engines
30 such as Google Scholar offer comprehensive coverage, citation analysis, or identification of full-text
31 documents[5]. Meanwhile, focused named Theory Searches (e.g., using a Behaviour, Health context,
32 Exclusions, Models or Theories—BeHEMoTh—template) specifically target literature that explicitly
33 mentions or discusses relevant mid-range theories[10].

34 While these techniques have proven valuable, they present practical challenges. Manual
35 implementation of multiple search strategies is labour-intensive and time-consuming[11]. Citation
36 chaining requires extensive manual tracking across multiple sources and document types[5].
37 Identifying semantically related concepts—particularly the nuanced theoretical constructs central to
38 realist reviews—remains difficult using keyword-based searching alone[12]. Assessing relevance
39 across large volumes of retrieved literature demands significant expertise and repeated judgement
40 calls[13]. These limitations can constrain the breadth and depth of evidence synthesis, particularly for
41 complex interventions where relevant insights may be scattered across diverse literatures.

42 Recent advances in AI and large language model (LLM)-assisted tools offer potential solutions to
43 these challenges. These tools can identify relevant studies through semantic understanding rather than
44 keyword matching, support automated relevance ranking based on theoretical alignment, and
45 streamline citation tracking and cross-referencing processes[14, 15]. AI applications are now capable
46 of automating multiple stages: generating tailored search strategies, conducting iterative searches that
47 adapt based on emerging findings, and directly identifying studies based on semantic similarity[14,
48 15].

49 However, the rapid development of innovative literature search tools—particularly those powered by
50 AI—has created an emerging need for contemporary methodological guidance to help researchers
51 effectively utilise these tools within established review frameworks. While AI capabilities offer clear
52 potential to address the practical challenges of realist searching, there is limited guidance on how to
53 integrate these tools while maintaining the systematic, transparent, and theoretically grounded
54 approach that realist methodology requires. This case study aims to address this need by providing a
55 detailed, replicable account of how AI-driven tools can be integrated into the realist review process,
56 ensuring that research methodologies evolve alongside technological advancements.

57 Although developed in the context of a realist review, the underlying process—systematically
58 extending searches to surface evidence that conventional strategies may overlook—has broader
59 applicability. Any literature review seeking to identify material not easily captured through standard
60 indexing or keyword-based searching could benefit from this approach. By clearly outlining the steps
61 taken to retrieve and select relevant studies, we seek to enhance transparency and replicability,
62 making it easier for others to follow. In doing so, this case study provides a detailed, replicable
63 account of how these advanced tools were integrated into the realist review process. The goal is not
64 only to demonstrate how these innovative methods can be used effectively but also to encourage other
65 researchers to adopt and apply these approaches in their own reviews, contributing to the ongoing
66 evolution of the field.

67 **Methods**

68 Realist review is a flexible approach to evaluating the literature, allowing for variation in how the

69 process is undertaken[16]. While authors may perform more or fewer steps than those outlined by

70 Pawson, his framework consists of six key stages: identifying the review question, searching for

71 empirical evidence, quality appraisal, data extraction, data synthesis, and dissemination of findings[2].

72 This study focuses specifically on steps two and three—searching for empirical evidence and

73 conducting quality appraisal—within the context of a realist review evaluating prehabilitation

74 programmes for individuals awaiting spinal surgery for neurogenic claudication (NC)[17]. We

75 describe the use of AI-powered tools, Scite and Undermind (among others available), to identify

76 primary studies for testing Context-Mechanism-Outcome Configurations (CMOCs) developed as part

77 of a programme theory. The following sections outline the search process and provide a transparent

78 and rigorous approach to screening results. We present this process, not as a prescriptive method or a

79 model of "best practice," but rather as a resource for other researchers to adapt to their own studies.

80 **Searching for empirical evidence**

81 Our search focused on identifying empirical studies to test, refine, and potentially refute the CMOCs

82 developed within an initial programme theory —a preliminary framework outlining the expected

83 relationships between context, mechanisms, and outcomes for prehabilitation in NC surgical

84 candidates. Further details on the development of the initial programme theory and the realist review

85 methods have been published elsewhere[18] . This objective shaped our search strategy and

86 influenced the selection of tools used. Guided by an experienced information specialist with expertise

87 in realist methodology (AB), the research team decided to use two AI-powered tools (Scite and

88 Undermind) with complementary functions. Scite was used to explore citation networks and analyse

89 how concepts were cited across the literature—identifying supporting, contrasting, or mentioning

90 evidence through its claim-analysis capabilities. Undermind was used to conduct semantic searches

91 that identified conceptually relevant papers through natural language processing, even when specific

92 terminology differed (see Box 1 for further details on Scite and Undermind). Using these tools

93 together enhanced our ability to retrieve studies that conceptually aligned with the CMOCs under

94 investigation   Although we chose to use Scite and Undermind specifically, , the methods we describe

95 are not dependent on these particular tools; the principles outlined here are broadly applicable and can

96 be operationalised using other similar platforms.

97 *Search strategy*

98 Scite's claim-analysis capabilities, which classify citations as 'supporting,' 'contrasting,' or

99 'mentioning' the claims of the cited work[19], were used to explore the core concepts within each

100 CMOC.   Targeted search queries, combining 'prehabilitation' with relevant keywords, were

101 constructed (see supplementary file 1 for an example). This approach enabled the research team to

102   analyse how concepts were used and referenced within the literature—specifically, the 'citation
103   context'—allowing the research team to identify supporting or contrasting evidence and assess the
104   credibility and application of these concepts within the prehabilitation literature.

105   To maintain precision in our search, we privileged retrieval of the most relevant studies rather than
106   expanding the search with numerous synonyms or alternative terms. For example, to test CMOC1, the
107   search strategy used was: Prehabilitation "waiting list" OR Prehabilitation "waiting period". We chose
108   these terms deliberately, as we considered it unlikely that a relevant full-text document would fail to
109   reference either of these terms if it were addressing prehabilitation in waiting lists or periods.
110   Although additional synonyms could have broadened our results, this focused strategy helped ensure
111   that the studies retrieved were directly relevant to our research question. Such an approach contrasts
112   with systematic reviews where a review team works painstakingly through a list of terms of increasing
113   futility.

114   Undermind generated detailed reports on specific research topics aligned to each CMOC. For our
115   search related to CMOC 1, we began by copying and pasting the CMOC into the initial query box of
116   Undermind, which then guided us through a series of answer-and-response prompts to clarify the
117   focus of our search. Through this iterative process, Undermind helped clarify the key concepts of
118   interest, such as the type of surgery and the desired outcomes. After refining these details, Undermind
119   generated a final search query for CMOC 1: '*The impact of prehabilitation on patient empowerment
120   and stress reduction in individuals awaiting orthopaedic surgeries'*. Based on this, Undermind
121   produced a comprehensive report for CMOC 1, summarising relevant literature related to key
122   concepts such as patient empowerment, stress reduction, multimodal programmes, and orthopaedic
123   prehabilitation. The report also included curated citations, helping to streamline the identification of
124   contextually relevant studies. (see supplementary file 2 for an example of a report).

Scite: Claim verification and citation analysis
Claim Verification: Assesses scientific claims by analysing how they have been supported, contradicted, or mentioned across the literature.   **Citation Analysis:** Tracks how individual papers have been cited, offering insight into their influence and relevance.   **Research Tool Integration**: Compatible with Google Scholar, PubMed, and Web of Science.   **Output:** Provides *contextual citation insights* to enhance the interpretation of evidence.
Undermind: Advanced literature searching for complex or interdisciplinary topics

- **Iterative Search Process**: Refines and expands search strategies in real time to uncover both central and peripheral literature.
- **Semantic Understanding:** Uses natural language processing to grasp the meaning behind research questions and extract conceptually relevant studies.
- **Personalised Recommendations:** Adapts search results to match a researcher's specific interests and goals.
- **Output:** Yields a broader and more nuanced evidence base tailored to the review's evolving needs.

125   Box 1: Overview of Scite and Undermind - purpose and capabilities[20]

126

Initial Screen
CMOC 1

Records identified by Scite
(n=7)

Records identified by Undermind
(n=23)

Records Excluded
(n=13)
Undermind score
<50%

Remaining papers
(n=10)

All 10 papers checked to confirm agreement

Key papers identified by Undermind
(n=5)

Records Excluded
(n=3)

Papers included guided by RRR framework
(n=2)

Papers proceeding to full text Relevance screen
(n=9)
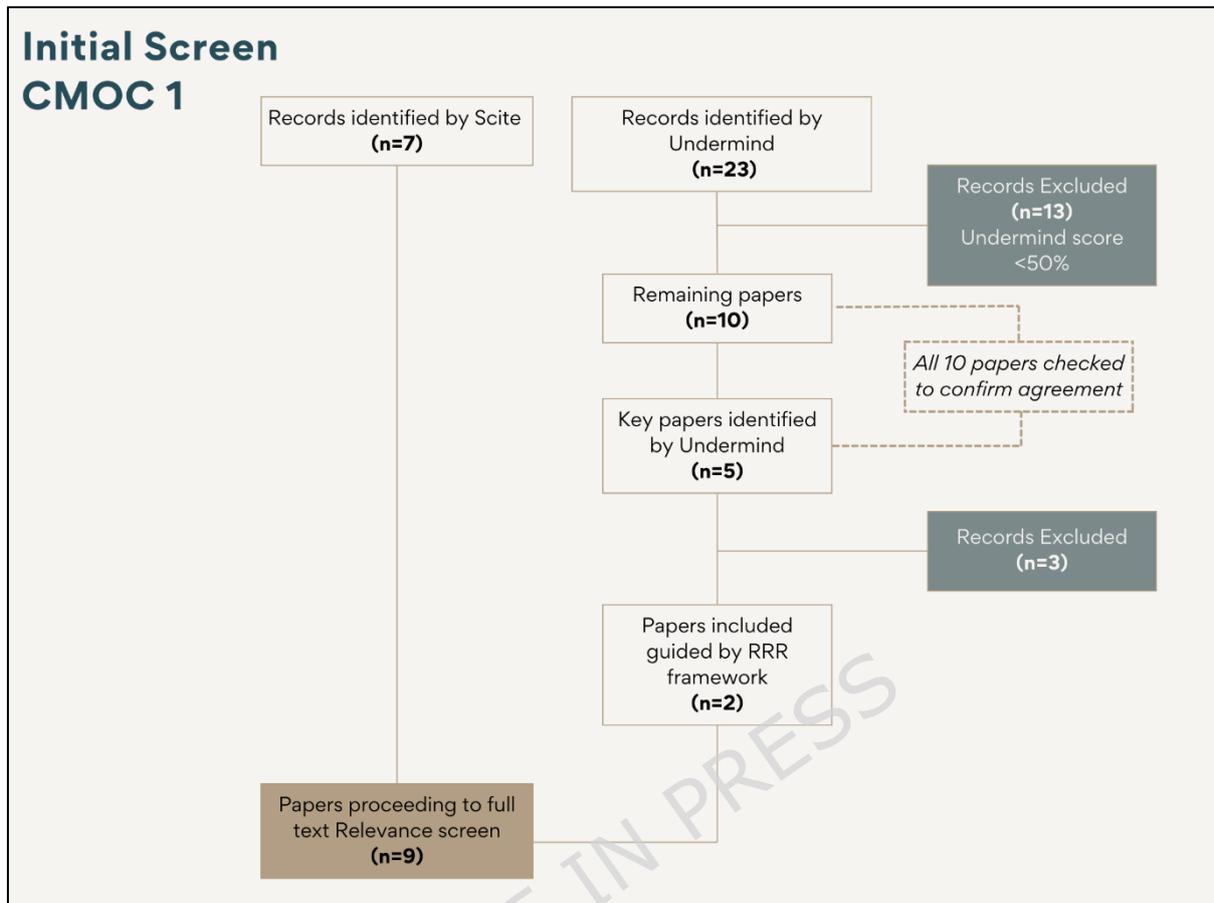
127   *Initial screening*

128   Figure 1: An initial screen of search results for CMOC 1

129   Having retrieved a set of potentially relevant papers from both Scite and Undermind, the next step

130   involved systematically screening these results to assess their relevance to the CMOCs. To reduce the

131   volume of papers and make the dataset manageable, preliminary screening of the literature results was

132   conducted by a single researcher (RH). Although this stage was not formally quality assessed, it was

133   guided by broad criteria for assessing relevance, richness, and rigour (RRR)[21]. The RRR framework

134   is widely used in realist research to evaluate studies across these three dimensions rather than focusing

135   solely on methodological rigour as traditional appraisal tools do. This recognises that valuable

136   insights ('nuggets') contributing to theory development can be found even in studies that may be

137   considered methodologically weak by traditional standards. The Undermind search engine assigned a

138   percentage score to each paper, indicating how closely it matched the research query. Papers scoring

139   below 50% were excluded from further consideration. Among those scoring above 50%, Undermind

140   highlighted 'top scoring' papers it considered key. The research team reviewed these to verify

141   alignment with the CMOC and then applied the RRR framework to determine which papers should

142   proceed to full-text review.

143   This initial screening process evokes what Bates described as 'berrypicking'—a model of real-world

144   information seeking where queries evolve over time and researchers gather relevant material

145   iteratively[8]. Just as a forager scans multiple bushes, choosing only the most promising berries, this

146   stage involved scanning AI-ranked results and selectively "picking" studies based on both algorithmic

147   cues and the RRR framework. 'Berrypicking' aligns well with the interpretive, non-linear nature of

148   realist reviews, where studies are selected not through exhaustive retrieval but through theoretically

149   informed, context-sensitive filtering[5]. The research team concluded that the risk of exclusion errors

150   at this stage was minimal. Any borderline decisions would be captured in subsequent full-text

151   screening phases, and the grandparent-parent-child approach (described below) meant that

152   conceptually relevant literature could still be retrieved through citation chaining even if initially

153   missed at this stage.

154   In our review, the number of papers retrieved from Scite for each CMOC was generally small (1–7

155   papers) due to the conceptual specificity of each CMOC. As such, all retrieved papers were taken

156   directly to full relevance assessment. Reading Scite's citation statements—brief extracts from the full

157   text highlighting the relationship between a cited study and the citing paper—helped confirm the

158   relevance of these studies to the CMOC. In scenarios where Scite returns larger sets of papers, citation

159   statements could support a staged prioritisation approach that could be applied consistently across all

160   CMOCs, enabling researchers to focus first on mechanism-rich studies while maintaining consistency

161   and rigour. Even if larger numbers of papers were retrieved, it is plausible that reviewing and selecting

162   relevant studies would be faster using the citation statements than relying solely on titles and abstracts

163   in traditional searches. This, however, remains conjectural and warrants further investigation.

164   **Application Example: Initial screening of results for CMOC 1**

165   In the case of CMOC 1, Scite produced seven results —considered a manageable number—which

166   were taken forward directly without the need for initial screening. In contrast, Undermind presented

167   23 results, of which 13 scored below the 50% threshold. After a brief review of titles and summaries,

168   these 13 papers were excluded. Of the remaining ten, Undermind identified five papers as key. While

169   these five were prioritised, the researcher reviewed all ten to validate Undermind's assessment and

170   ensure no relevant studies were overlooked. The RRR framework was then applied to the five key

171   papers resulting in two being selected for full-text relevance screening. These two papers were then

172   combined with the seven identified by Scite, resulting in a total of nine papers to be taken forward to a

173   full text relevance screen (See figure 1 for illustration).

174   **Seed and key Informant papers**

175   To support the iterative nature of the review and maintain a transparent record of potentially relevant

176   evidence, studies were categorised as *seed papers* or *key informant papers*. A *seed paper* refers to a

177   secondary study that was not included in the analysis itself but cited primary studies that may be

178 relevant to one or more CMOCs. The use of this term was informed by the Terminology, Application,
179 and Reporting of Citation Searching TARCiS statement, which defines "seed references" as relevant
180 articles known beforehand and used as a starting point for citation searches[22]. Adopting this label
181 allowed for systematic tracking of secondary studies that served as gateways to further primary
182 research. *Key informant papers*, on the other hand, were primary studies that were directly relevant to
183 more than one CMOC within the programme theory (See box 2 for a definition of papers). This
184 coding enabled the research team to ensure that studies with the potential to inform multiple CMOCs
185 were flagged for further analysis, ensuring a thorough and connected synthesis across the programme
186 theory.

Paper Type	Definition/Decision Criteria
Seed paper	A **secondary study** not included in the review itself but which cites primary studies that may be relevant to the programme theory's CMOCs
Key informant paper	A **primary study** that is relevant to more than one CMOC
Grandparent Paper	**Primary** or **secondary** studies identified in the initial screen as being potentially relevant to the CMOC being tested. - If the grandparent paper is a **primary study** and meets the relevance criteria, it is included in the second formal screening phase for Relevance, Richness, and Rigour (RRR). - If the grandparent paper is a **secondary study**, it is not included in the review. - If the grandparent paper is a **secondary study** that references potentially relevant **primary studies** informing the CMOC being tested or other CMOCs, it is colour-coded as a *seed paper* and tagged with hashtags for each CMOC it may inform. - If the grandparent paper is a **primary study** that does not directly inform the CMOC being tested but is relevant to other CMOCs in the programme theory, it is also categorised as a *key informant* and similarly tagged.
Parent paper	- A **primary study** identified from a grandparent paper that is considered potentially relevant for the CMOC being tested or for other CMOCs within the programme theory. - If a parent study is relevant to more than one CMOC, it is coded as a **key informant paper** and assigned hashtags for each CMOC it is considered potentially relevant to
Child paper	A primary study identified from an included parent paper that is considered potentially relevant for the current context-mechanism-outcome configuration (CMOC) being tested.

187 Box 2: Definition of paper types
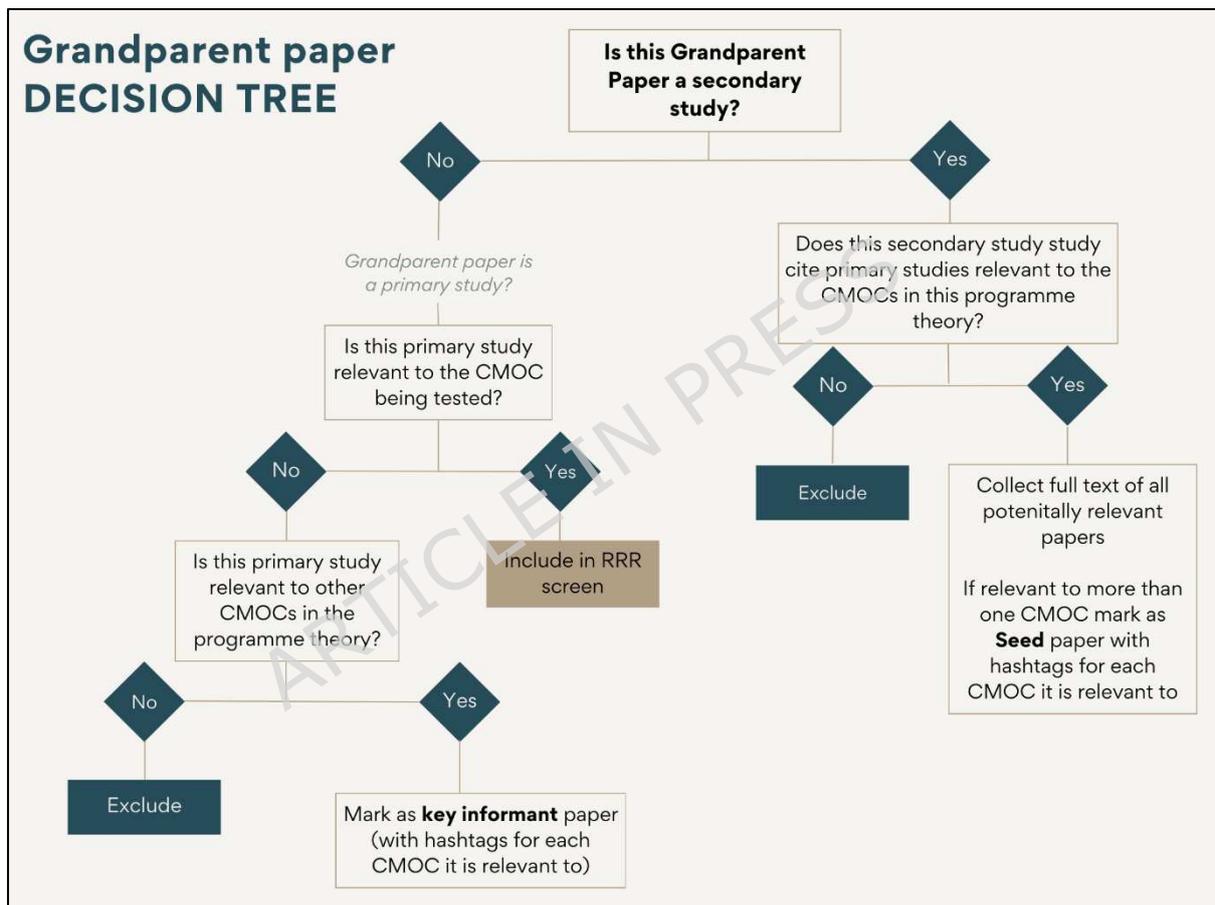
188 *Relevance Screen*

189 After the initial screen, papers identified from the search results of Scite and Undermind as potentially
190 relevant were categorised as 'grandparent papers' and subjected to a full-text relevance screen. This
191 process involved assessing whether the studies directly addressed key concepts and variables within
192 the CMOC being tested, as well as evaluating their potential to inform other CMOCs. As outlined in
193 the criteria below (see Box 3), studies were classified based on their relevance to the CMOC being
194 tested, with additional considerations for those that could inform multiple CMOCs.

Include	• **Primary studies** (qualitative or quantitative) that are directly relevant to the CMOC being tested, meaning the study explicitly addresses concepts or variables within the CMOC. • **Primary studies that are relevant** to the CMOC being tested but also identified as relevant for other CMOCs. **Studies with dual relevance** will be included in the current CMOC analysis and coded as a **"key informant"** for future re-evaluation with other CMOCs. • **Primary studies not relevant** to the current CMOC, **but relevant to other CMOCs**: These studies, will be coded as **key informants for other CMOCs** and retained for future analysis.
Exclude	• **Primary studies** that are not related to the CMOC being tested or any other CMOC within the programme theory. • **Secondary studies** that are not relevant to the CMOC being tested or any other CMOC within the programme theory.
Inform	• **Seed papers**: Secondary studies (e.g., reviews, meta-analyses) that cite primary studies potentially relevant to the CMOC being tested or other CMOCs within the programme theory. These papers will be reviewed to identify relevant primary studies. • **Key informant papers**: Primary studies that are not directly relevant to the CMOC being tested but may be relevant to other CMOCs within the programme theory. These papers will be flagged for future re-evaluation. If a study becomes relevant to a CMOC already tested, it will be re-evaluated for inclusion.

195 Box 3: Relevance Screening and Categorisation Criteria for CMOC Analysis: Include, Exclude, and Inform

196 **Grandparent papers**

197   Since our literature search was specifically designed for searching for empirical evidence for testing

198   CMOCs[2], only primary studies were considered for inclusion at this stage. Relevant grandparent

199   studies (studies retrieved from the AI systems) that were primary studies advanced to RRR screening.

200   Grandparent studies classified as secondary studies, retrieved similarly via the AI systems, were not

201   includable in the review in their own right but were mined for any potentially relevant primary

202   studies.   If relevant primary studies were identified, the grandparent study was colour coded to

203   indicate its status as a 'seed paper'. This process continued until all relevant grandparent studies for

204   the CMOC under testing had been screened. Figure 2 illustrates the decision tree to guide the

205   researcher in determining the appropriate course of action regarding grandparent papers.


206   Figure 2: Decision tree for evaluating Grandparent paper

207

208 **Application example: Grandparent paper relevance screen CMOC 1**

209 Using CMOC 1 as an example, nine papers were identified from the initial screen as potentially

210 relevant to CMOC 1 (see Figure 3). These were read in full and screened for relevance both for

211 CMOC 1 and for any other CMOCs in the programme theory. Six of the nine papers were considered

212 not relevant and were excluded (see Box 3 for exclusion criteria). One paper was a secondary study

213 which referenced potentially relevant primary studies, and thus it was coded as a seed paper. Two

214 papers were primary studies that were not relevant to the current CMOC being tested but were

215 considered relevant to other CMOCs in the programme theory, and so they were coded as key

216 informant papers.
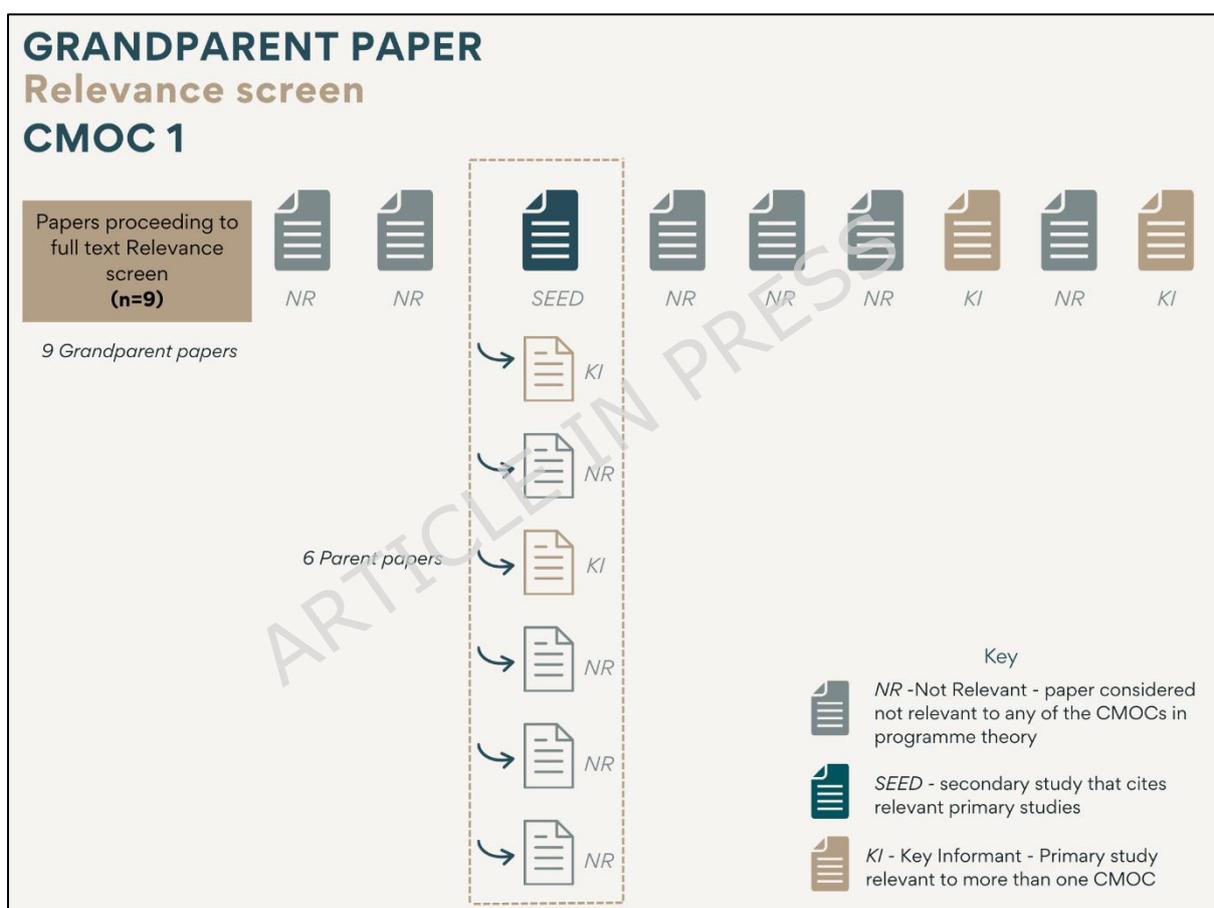
217 Figure 3: Overview of relevance screen for CMOC 1

218 **Parent papers**

219 Relevant primary studies identified from the grandparent papers were then, in turn, termed *parent*

220 *papers*. Each parent paper was read in full to assess its relevance to the CMOC being tested (See Box

221 3). If a parent paper was relevant to the current CMOC, it was advanced to the RRR screening phase.

222 In some cases, a parent paper was also relevant to other CMOCs within the broader programme

223 theory; in these instances, it was both progressed to RRR screening and marked as a *key informant*

224 *paper* for future analysis. If a parent paper was not relevant to the current CMOC but held potential

225   relevance for other CMOCs, it was designated solely as a *key informant paper*, to be revisited during

226   the testing of those CMOCs. Finally, if a parent paper was found to be irrelevant to both the current

227   CMOC and all other CMOCs, it was excluded from further analysis. Figure 4 provides a visual

228   illustration of the decision tree used to guide the assessment of parent papers, outlining the pathways

229   for inclusion, exclusion, or designation as key informant papers based on their relevance to the

230   CMOC under review and/or the broader programme theory.

231   Figure 4: Decision tree for screening and categorising parent papers
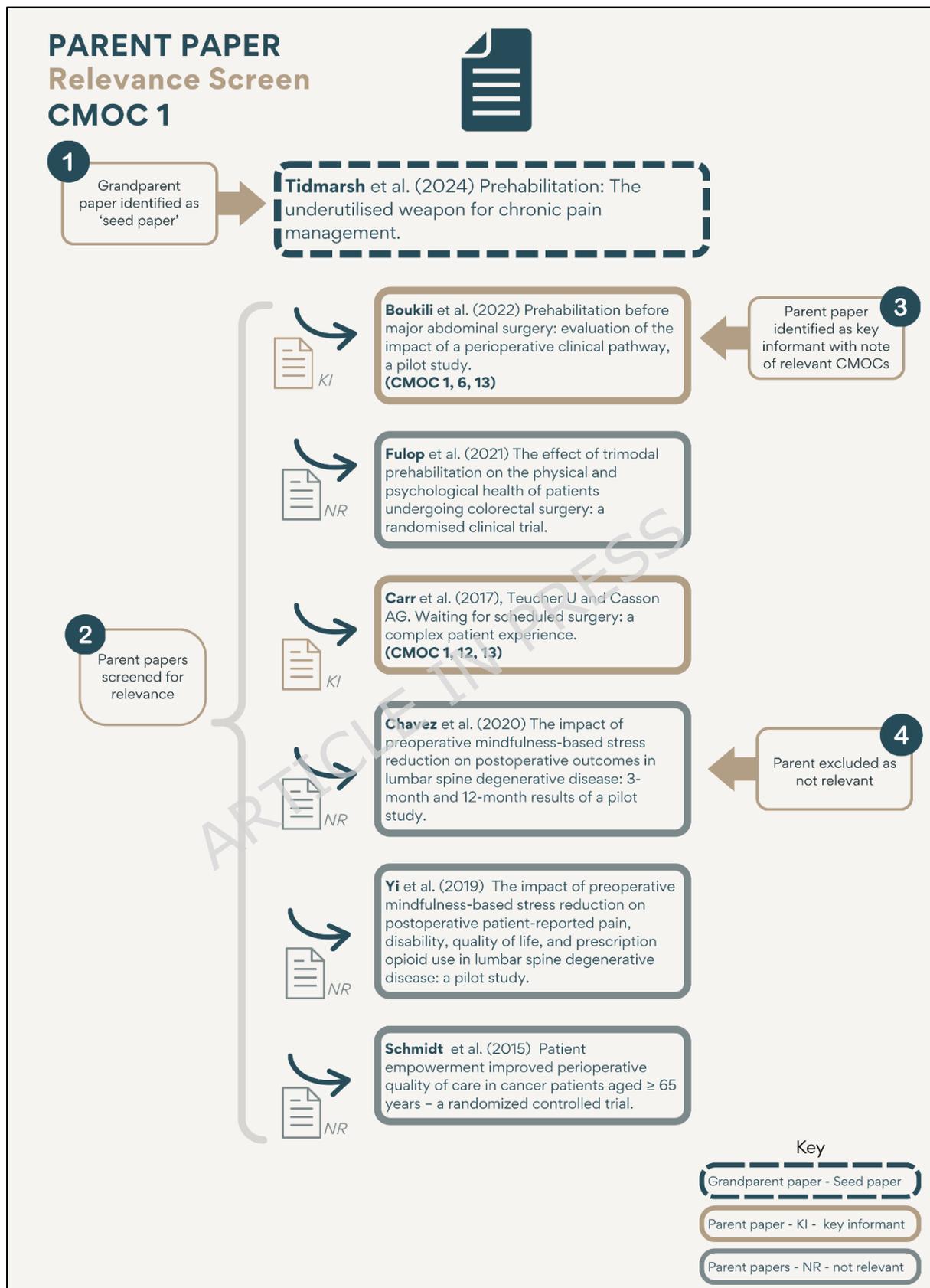

232

PARENT PAPER
Relevance Screen
CMOC 1

1 Grandparent paper identified as 'seed paper'

Tidmarsh et al. (2024) Prehabilitation: The underutilised weapon for chronic pain management.

Boukili et al. (2022) Prehabilitation before major abdominal surgery: evaluation of the impact of a perioperative clinical pathway, a pilot study.
(CMOC 1, 6, 13)
KI

3 Parent paper identified as key informant with note of relevant CMOCs

Fulop et al. (2021) The effect of trimodal prehabilitation on the physical and psychological health of patients undergoing colorectal surgery: a randomised clinical trial.
NR

2 Parent papers screened for relevance

Carr et al. (2017), Teucher U and Casson AG. Waiting for scheduled surgery: a complex patient experience.
(CMOC 1, 12, 13)
KI

Chavez et al. (2020) The impact of preoperative mindfulness-based stress reduction on postoperative outcomes in lumbar spine degenerative disease: 3-month and 12-month results of a pilot study.
NR

4 Parent excluded as not relevant

Yi et al. (2019) The impact of preoperative mindfulness-based stress reduction on postoperative patient-reported pain, disability, quality of life, and prescription opioid use in lumbar spine degenerative disease: a pilot study.
NR

Schmidt et al. (2015) Patient empowerment improved perioperative quality of care in cancer patients aged ≥ 65 years – a randomized controlled trial.
NR

Key
Grandparent paper - Seed paper
Parent paper - KI - key informant
Parent papers - NR - not relevant

Figure 5: Parent relevance screen for CMOC 1

233   **Application example: Parent paper relevance screen CMOC 1**

234   From the relevance screen of the nine grandparent studies identified for CMOC 1, one was
235   categorised as a seed paper—a secondary study that cited potentially relevant primary studies. From
236   this seed paper, six primary studies were identified and termed 'parent papers'. These parent papers
237   were read in full to assess their relevance using the same criteria applied during the grandparent
238   screening stage. Of the six, four were excluded as not relevant to the current CMOC or any other
239   CMOCs. The remaining two were found to be relevant both to CMOC 1 and to other CMOCs within
240   the broader programme theory. These two parent papers were therefore included for RRR screening
241   and simultaneously coded as key informant papers to be revisited in future CMOC analyses (see
242   figure 5).

243   When a primary study is identified as relevant to more than one CMOC, it may seem redundant to
244   code it as a key informant paper. It could be argued that, for expediency, the paper should go straight
245   through to the RRR screening phase, tagged with hashtags for each relevant CMOC—bypassing the
246   "key informant" categorisation. However, revisiting the paper when focusing on each new CMOC
247   offers an opportunity to reassess its relevance in light of the current focus and the evolving data being
248   analysed. Re-reading at this point allows researchers to refine their understanding and verify whether
249   the study's initial relevance remains valid or whether new insights have emerged that might affect its
250   inclusion within another CMOC.

251   The act of revisiting a study is analogous to re-interviewing a participant in qualitative research after
252   having spoken to additional participants. Insights gained from the other interviews may influence how
253   you interpret the second conversation with the same person, allowing for a deeper understanding or a
254   different perspective based on the new information[23]. Similarly, when revisiting a study in the
255   context of a new CMOC, the interpretation of its relevance may evolve as it is considered alongside
256   the findings from other studies. While this process may seem time-consuming, it tends to become
257   increasingly efficient with each iteration. Since the study has already been reviewed, subsequent
258   readings are generally lighter-touch providing an opportunity to add nuanced insights and annotations.

259   Revisiting a study when focusing on a new CMOC not only provides an opportunity to refine
260   interpretations but also helps sensitise researchers to the emergence of demi-regularities—context-
261   dependent patterns that reveal how mechanisms operate across different settings[24]. However, while
262   searching for these demi-regularities, it is important to remain cautious of confirmation bias —the
263   tendency to focus on evidence that supports an emerging idea while overlooking information that
264   might contradict or complicate it[25]. This can lead to a reinforcing feedback loop where each item of
265   supporting evidence strengthens belief in the prevalent idea and heightens sensitivity to similar
266   patterns. This mirrors the 'streetlamp effect': searching only where the light is shining, rather than
267   where the most meaningful evidence may lie[26]. Realist researchers are actively engaged in looking
268   for rivalry and counterfactuals, which are fundamental for theory development[27]. By deliberately

269 seeking alternative explanations and examining how different CMOCs might play out, researchers can

270 refine their theories and ensure that they aren't simply confirming pre-existing assumptions. While the

271 challenge of confirmation bias cannot be fully eliminated, being aware of the broader aim of revisiting

272 studies—testing, challenging, and refining emerging theories—helps mitigate this risk. This approach

273 ensures that the review remains rigorous, balanced, and open to new insights, ultimately leading to a

274 robust understanding of the relationships between context, mechanisms, and outcomes.

Child papers

276 In some cases, primary studies referenced within included parent papers—referred to as child studies

277 (second stage referrals)—may be potentially relevant to the CMOC under testing. The screening

278 process for child papers follows the same procedure as parent papers. Decisions on whether to screen

279 and include child studies can be made by the research team, typically guided by the extent of available

280 data and the need for further evidence to test the CMOC in question. In our review, where the

281 literature on nutrition in prehabilitation programmes for NC surgical candidates was relatively sparse,

282 references from a few child studies were included. However, given time constraints and limited

283 resources, a pragmatic decision was made not to extend the review to the child level for the remaining

284 CMOCs, as the prevalent literature base was deemed sufficient. Authors thus face an ongoing decision

285 on how many iterations (generations) are required, both in general, and for particular CMOCs. In

286 doing so they need to decide whether to prioritise standardisation across CMOCs (following up the

287 same number of levels for each CMOC) or to favour customisation to the specific needs of each

288 CMOC (with different levels for each).

Quality assessment

290 Quality assurance checks were applied at the parent study stage only, as this represented the point at

291 which exclusion decisions posed the greatest risk to theory development. Formal quality assurance

292 checks were not conducted at earlier stages of the screening process, including on records excluded by

293 the AI-powered searches at the grandparent stage, as these stages functioned as relevance-based

294 identification filters rather than points of quality appraisal. Potential gaps in retrieval were instead

295 addressed through citation and reference checking of included papers.

296 At the parent study stage, we implemented a quality assurance check (see figure 6 and the box

297 highlighted as 'relatively high risk'). Since decisions about which papers to advance to RRR

298 screening were based on judgments of relevance, ensuring a consistent and rigorous approach was

299 important. Given that we were trialling this method, we adopted a cautious approach by having a

300 second screener review 50% of the records excluded by the first reviewer for each CMOC. One

301 disagreement was identified during this process, which was resolved through discussion and

302 consensus. While this percentage may seem high, it proved feasible, as a total of 18 papers across all

303 CMOCs were excluded at this stage, meaning the second screener only needed to assess nine papers.

304   This additional check helped strengthen the reliability of our process. During this process, the second
305   screener noted the importance of keeping a copy of all CMOCs readily available whilst screening to
306   ensure that papers were assessed not only for relevance to the CMOC being tested but also for
307   potential relevance to other CMOCs within the broader framework.

308   Figure 6 uses CMOC 1 as an example to illustrate the flow of studies included in the review, starting
309   with the grandparent studies retrieved through the AI-powered searches using Scite, Undermind, and
310   key informant papers. The diagram traces the path to the parent studies, highlighting key points at
311   which studies were excluded from the review. Two boxes with dashed outlines indicate these
312   exclusion points. The second exclusion box labelled 'relatively high risk,' marks the stage where a
313   second reviewer was introduced for quality appraisal, as described earlier in the methodology. Such
314   an approach harmonises with recent approaches to quality assessment, particularly for rapid review
315   methods, where previous arbitrary models of double checking are increasingly being replaced with
316   strategies specifically targeting stages considered of higher risk[28]. Similar flow diagrams were
317   created for all 14 CMOCs included in the programme theory, ensuring consistency and transparency
318   throughout the study selection process.
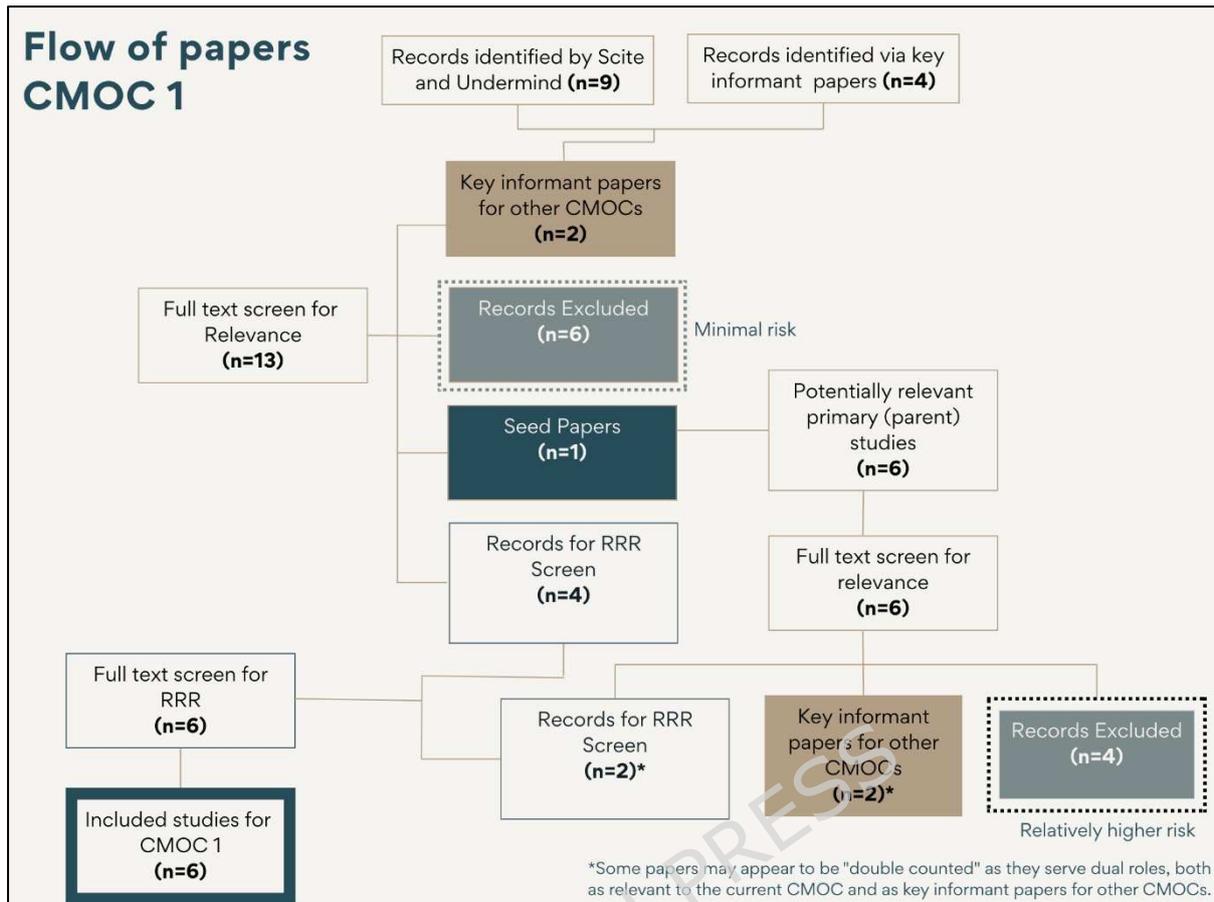
319 Figure 6: Study selection flow for CMOC *1* with exclusion points highlighted

320 **Relevance, Richness and Rigour appraisal**

321 The RRR screening process served as a transparency step rather than a strict eligibility filter. By this

322 stage, all papers had already undergone an initial relevance screen, ensuring they were relevant to one

323 or more CMOCs. As a result, there were no exclusions, and the RRR screen focused on assessing the

324 confidence with which conclusions could be drawn from each paper. Each paper was evaluated

325 against three independent criteria: relevance, richness, and rigour, with scores of high, moderate, or

326 low assigned to each (see supplementary file 3). While papers could theoretically be excluded if they

327 scored low across all three criteria, in practice, the RRR screen primarily aimed to clarify the strength

328 of the evidence and highlight any limitations. For example, papers with high relevance but lower

329 richness or rigour were still included, but their limitations were explicitly acknowledged to ensure

330 transparent and nuanced interpretation of the findings. This approach aligns with the guidance

331 provided by Dada et al[21], which emphasises transparency in the evaluation of evidence to ensure

332 clarity regarding the confidence in findings and conclusions drawn from studies.

333 **Discussion**

334 This case study demonstrates how AI-powered tools can enhance literature searching by retrieving

335 conceptually rich and contextually relevant studies that might be overlooked using conventional title-

336 and abstract-based searches. Using tools such as Scite and Undermind within a structured

337 workflow— including seed papers, key informant papers, and a grandparent–parent–child

338 classification system— allowed us to improve the conceptual quality of the retrieved studies in terms

339 of rigour, relevance, and richness. Although developed for a realist review, this approach can be

340 applied more broadly to literature reviews where hard-to-find or conceptually rich evidence is

341 important. By extending searches beyond conventional strategies, AI-assisted methods can

342 complement human judgement and support the identification of studies that inform theory

343 development and refinement.

344 Conventional keyword searches often rely on the researcher's vocabulary aligning with that used in

345 titles, abstracts, or indexing terms. This limitation can hinder the discovery of relevant literature that

346 uses alternative language or conceptual framing. In contrast, AI-powered tools like Undermind and

347 Scite use natural language processing (NLP) to assess full-text similarities and conceptual

348 relationships. This allowed us to retrieve literature aligned with the theoretical underpinnings of our

349 CMOCs—even when key terms were absent—thereby improving conceptual retrieval beyond the

350 capabilities of traditional search strategies. Importantly, these tools were deployed within tightly

351 defined conceptual boundaries (i.e., specific CMOCs), rather than broader population–intervention–

352 comparison–outcome (PICO) queries, making them particularly suited to realist inquiry. For example,

353 when testing CMOC 1—which proposed that being listed for surgery provides relief and

354 empowerment for patients who have been advocating for their health, leading to increased motivation

355 to engage in prehabilitation—the AI tools retrieved studies examining patient experiences of surgical

356 waiting lists, the psychological impact of having to 'advocate' for themselves, and motivation for pre-

357 operative preparation. These papers were conceptually aligned with our CMOC but did not

358 necessarily use terms like 'empowerment' or 'relief' in their titles or abstracts.

359 AI-powered tools enhanced our ability to identify relevant papers for testing our CMOCs.

360 Conventional keyword searches rely on specific terms and phrases input by the researcher matching

361 either those used by the author in their title or abstract or those used by an indexer to enhance

362 retrieval. Not only does this limit the search to the researcher's knowledge and their associated

363 vocabulary but it also misses greater retrieval opportunities offered by the richness of full text

364 retrieval. As a result, these searches may miss conceptually related literature that uses different

365 terminology or phrasing. In contrast, tools such as Undermind and Scite employ natural language

366 processing techniques to identify papers that align conceptually with the search query, even when the

367 exact terms or synonyms are not present. By drawing on a broader knowledge base, these tools offer a

368 nuanced approach that extends beyond keyword searches. This capability, broadening retrieval but

369 only within very tight search parameters (the language associated with a CMOC, not a broader PICO

370 review question) was particularly valuable for our realist review, as we sought literature to test

371 concepts, identify mechanisms, and understand important contextual factors.

372 While AI tools like Scite and Undermind improved the conceptual precision of our screening process,

373 they did not replace the interpretive work required in a realist review. Their ability to identify

374 semantically relevant papers and provide summarised insights helped surface data that may have been

375 overlooked using traditional keyword searches. However, the task of making sense of these findings

376 —through retroductive and abductive reasoning[29] remained firmly with the research team. This

377 interpretive work was essential in testing, refining, and refuting our CMOCs. For instance, when

378 papers discussed patients wanting to 'take an active role' during the waiting period, we had to interpret

379 whether this genuinely evidenced our theorised mechanism of empowerment, or whether it reflected a

380 different psychological process entirely. Our findings therefore support the view increasingly

381 reflected in the literature: AI is not a substitute for human judgement but a valuable adjunct to it.

382 Studies such as Ge et al.[30] and Gwon et al.[31] highlight similar dynamics in other evidence

383 synthesis contexts, where AI accelerates routine tasks like screening but requires human oversight for

384 rigour and contextual interpretation. Feng et al.[32] further emphasise this by demonstrating that AI

385 tools often achieve high recall (i.e., successfully identifying most relevant papers), but can vary in

386 precision—reinforcing the need for critical human appraisal. As such, the integration of AI tools can

387 enhance the scope and sensitivity of realist reviews, but only when grounded in thoughtful, theory-

388 driven human interpretation.

389 *Challenges and Limitations*

390 While this manuscript is based on a single case study, using a detailed, single-case application to

391 introduce and demonstrate a new methodological approach is a common practice. Much of the

392 methodological evidence for information retrieval in systematic reviews is built on individual case

393 studies rather than large comparative studies[33-35] and this is recognised as a standard approach for

394 advancing search methodology. Nonetheless, there are several challenges and limitations in this case

395 study that are worth noting, and the approach requires further testing and evaluation in other contexts

396 and with other research teams to assess its broader applicability and robustness.

397 AI-powered tools operate within the constraints of their underlying algorithms, and the lack of

398 transparency in the coding of these algorithms may introduce biases. For instance, the relevance

399 scoring could reflect limitations or preferences embedded in the training data or algorithmic design

400 potentially leading to certain types of evidence being emphasised over others. Additionally, these

401 tools may have limited access to certain databases or publication types, such as grey literature or very

402 recent publications, which could result in gaps in the literature covered. Human reviewers are

403 encouraged to be reflexive about their influence on research processes, AI systems are not – although

404 some researchers have begun experimenting by scripting AI systems to identify their own limitations!

405 Furthermore, as these tools are relatively new, limited evidence exists for their performance against

406 librarian-led search strategies, especially in the context of realist reviews. Realist reviews prioritise

identifying conceptually relevant literature over exhaustive searches[5], so the effectiveness of AI-powered tools should be evaluated based on their ability to capture studies that meaningfully contribute to programme theory development, using criteria such as relevance, richness, and rigour[21]. While this study did not conduct a head-to-head comparison with traditional or baseline literature review methods using the same keywords, the value of the AI-assisted approach lies in retrieving conceptually rich and contextually relevant studies, prioritising quality over quantity. Empirical evaluation against conventional searches remains a potential avenue for future research. Despite the promise of AI-powered tools, they still require a degree of methodological expertise to use effectively. Researchers with a solid understanding of realist methodology—or working within a team with such experience—may be better equipped to formulate appropriate search queries and critically assess the conceptual relevance of retrieved studies. AI-powered tools are a useful adjunct to traditional search methods, helping to streamline the process. However, we would argue that they cannot replace the nuanced judgement and interpretation essential to realist review. Specifically, the time overhead required to oversee and validate their otherwise credible-looking output should not be overlooked.

While AI-powered tools can save time in searching and assembling the literature, time spent categorising papers may offset these gains. The multiple screening stages—such as categorising grandparent, parent, and child papers—demand significant researcher time and effort. However, this challenge arose primarily because this approach was novel and developed iteratively; with clearer instructions in place, it is likely to be quicker and more efficient in future applications. Our classification system and decision trees will require ongoing refinement and validation to ensure their continued credibility across different contexts.

A limitation regarding accessibility stems from our decision to use AI tools which require paid subscriptions for full functionality. However, the use of specialised, commercial software is a standard practice in modern academic research (e.g., licensed statistical or reference management packages). Furthermore, many features of Scite and Undermind could be used in the free or 'lite' versions of these applications, thereby mitigating the access barrier for resource-constrained researchers.

Finally, a key limitation of this review was the absence of quality assurance checks during the initial and grandparent screening stages. However, at the parent screening stage—where the risk of excluding rich and relevant studies was considered greater—we introduced a quality check, with a second reviewer screening 50% of excluded studies. While arguably risk averse, this step reflects a commitment to maintaining rigour in the screening process. Other research teams may wish to consider introducing similar checks earlier, depending on the stage at which key decisions are made and the level of confidence in the tools used.

Conclusion

441

442 This case study offers a methodological resource for researchers seeking to integrate and make the

443 most of AI-powered literature search tools in support of realist reviews. Searching for nuanced

444 concepts to build explanatory theories is fundamental to realist synthesis, but it often involves

445 navigating large volumes of literature in pursuit of elusive 'mechanisms' or key contextual factors.

446 By demonstrating the application of tools such as Scite and Undermind, this case study shows how AI

447 can support targeted, conceptually driven searches—enabling researchers to access rich and relevant

448 data. However, as with any tool, their value lies in how they are used. This is especially true for AI-

449 powered searches, which are still relatively new. We are continuing to learn how best to use these

450 tools effectively, and their role is to complement—not replace—the judgment and interpretive

451 thinking of the researcher.

452 The methods outlined in this case study offer a practical framework for selecting, organising, and

453 appraising search results in a transparent and robust way, contributing to the ongoing development of

454 realist review methodology. While not intended as a definitive blueprint, this case study offers a

455 foundation for others to build upon, test, and refine as methodological practices continue to evolve.

456 **List of abbreviations**

457 AI   Artificial Intelligence

458 CMOC   Context-Mechanism-Outcome configuration

459 RRR   Relevance, Richness, Rigour

460

461 # Declarations

462 **Ethical approval**

463 Not applicable

464 **Consent for publication**

465 Not applicable

466 **Availability of data and materials**

467 The authors confirm that key materials supporting the findings of this study are provided within the

468 article and its supplementary materials. Additional flow diagrams and related data are available from

469 the authors upon reasonable request

470 **Competing interests**

471   The authors declare that they have no competing interests

Author contributions

482   LW conceptualised the original realist review that informed this study. RH led the literature review

483   from which the current methodology was developed, with RH and AB jointly designing the

484   methodological approach. RH also created the system for organising selection and appraisal flow

485   diagrams in consultation with AB. AB designed the literature search, and RH interpreted the results in

486   collaboration with AB and LW. RH and AB co-wrote the initial draft of the manuscript, which was

487   subsequently reviewed and edited by AB and LW. All authors read and approved the final manuscript.

491

492

493   1.   Roth S, Wermer-Colan A. Machine learning methods for systematic reviews: a rapid scoping
494   review. Delaware Journal of Public Health. 2023;9(4):40.
495   2.   Pawson R. Realist synthesis: new protocols for systematic review. Evidence-based policy.
496   2006:73-104.
497   3.   Wong G. Data gathering in realist reviews: looking for needles in haystacks. Doing realist
498   research London: SAGE. 2018:131-45.
499   4.   Booth A, Briscoe S, Wright JM. The "realist search": A systematic scoping review of current
500   practice and reporting. Research Synthesis Methods. 2019;11(1):14-35.
501   5.   Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for
502   cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex
503   interventions: case study (CLUSTER). BMC medical research methodology. 2013;13:118-.
504   6.   Booth A, Wright J, Briscoe S. Scoping and searching to support realist approaches. Doing
505   realist research. 2018:147-66.
506   7.   Duddy C, Roberts N. Identifying evidence for five realist reviews in primary health care: A
507   comparison of search methods. Research Synthesis Methods. 2022;13(2):190-203.
508   8.   Bates MJ. The design of browsing and berrypicking techniques for the online search interface.
509   Online review. 1989;13(5):407-24.
510   9.   Coleman S, Wright JM, Nixon J, Schoonhoven L, Twiddy M, Greenhalgh J. Searching for
511   Programme theories for a realist evaluation: a case study comparing an academic database search and
512   a simple Google search. BMC Medical Research Methodology. 2020;20(1).
513   10.   Booth A, Carroll C. Systematic searching for theory to inform systematic reviews: is it
514   feasible? Is it desirable? Health Information & Libraries Journal. 2015;32(3):220-35.
515   11.   Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to
516   conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ
517   open. 2017;7(2):e012545.
518   12.   Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review-a new method of systematic
519   review designed for complex policy interventions. Journal of health services research & policy.
520   2005;10(1_suppl):21-34.
521   13.   O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study
522   identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4(1):5.
523   14.   Scherbakov D, Hubig N, Jansari V, Bakumenko A, Lenert LA. The emergence of large
524   language models (llm) as a tool in literature reviews: an llm automated systematic review. arXiv
525   preprint arXiv:240904600. 2024.
526   15.   Wu S, Ma X, Luo D, Li L, Shi X, Chang X, et al. Automated literature research and review-
527   generation method based on large language models. National Science Review. 2025;12(6):nwaf169.
528   16.   Hunter R, Gorely T, Beattie M, Harris K. Realist review. International Review of Sport and
529   Exercise Psychology. 2022;15(1):242-65.
530   17.   Hunter R, Booth A, Lamb SE, Williamson E, Hendrick P, Sahota O, et al. Prehabilitation for
531   lumbar spinal stenosis: understanding mechanisms and contexts for enhanced engagement-a realist
532   review. Age and Ageing. *In press*.
533   18.   Hunter R, Booth A, Lamb S, Williamson E, Hendrick P, Sahota O, et al. Prehabilitation for
534   lumbar spinal stenosis: understanding mechanisms and contexts for enhanced engagement—a realist
535   review. Age and Ageing. 2025;54(10):afaf311.
536   19.   Brody S. Scite. Journal of the Medical Library Association: JMLA. 2021;109(4):707.
537   20.   Labcritics. Undermind.ai Review: AI-Driven Revolution in the Literature Review: Labcritics,
538   Lab Tech Review & Marketplace; 2024 [Available from: https://labcritics.com/undermind-ai-review-
539   ai-driven-revolution-in-the-literature-review/.
540   21.   Dada S, Dalkin S, Gilmore B, Hunter R, Mukumbang FC. Applying and reporting relevance,
541   richness and rigour in realist evidence appraisals: advancing key concepts in realist reviews. Research
542   synthesis methods. 2023;14(3):504-14.
543   22.   Hirt J, Nordhausen T, Fuerst T, Ewald H, Appenzeller-Herzog C. Guidance on terminology,
544   application, and reporting of citation searching: the TARCiS statement. bmj. 2024;385.

545   23.   Goyes DR, Sandberg S. Trust, nuance, and care: Advantages and challenges of repeat
546   qualitative interviews. Qualitative Research. 2024:14687941241246159.
547   24.   Jagosh J, Macaulay AC, Pluye P, Salsberg J, Bush PL, Henderson J, et al. Uncovering the
548   benefits of participatory research: implications of a realist review for health research and practice.
549   Milbank Q. 2012;90(2):311-46.
550   25.   Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. Review of
551   general psychology. 1998;2(2):175-220.
552   26.   Kaplan A. The conduct of inquiry: Methodology for behavioural science: Routledge; 2017.
553   27.   Jagosh J. Realist synthesis for public health: building an ontologically deep understanding of
554   how programs work, for whom, and in which contexts. Annual review of public health.
555   2019;40(1):361-72.
556   28.   Booth A, Sommer I, Noyes J, Houghton C, Campbell F. Rapid reviews methods series:
557   guidance on rapid qualitative evidence synthesis. BMJ evidence-based medicine. 2024;29(3):194-200.
558   29.   Jagosh J. Retroductive theorizing in Pawson and Tilley's applied scientific realism. Journal of
559   Critical Realism. 2020;19(2):121-30.
560   30.   Ge L, Agrawal R, Singer M, Kannapiran P, De Castro Molina JA, Teow KL, et al. Leveraging
561   artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges.
562   Syst Rev. 2024;13(1):269.
563   31.   Gwon YN, Kim JH, Chung HS, Jung EJ, Chun J, Lee S, Shim SR. The use of generative AI
564   for scientific literature searches for systematic reviews: ChatGPT and Microsoft Bing AI performance
565   evaluation. JMIR Medical Informatics. 2024;12:e51187.
566   32.   Feng Y, Liang S, Zhang Y, Chen S, Wang Q, Huang T, et al. Automated medical literature
567   screening using artificial intelligence: a systematic review and meta-analysis. Journal of the American
568   Medical Informatics Association. 2022;29(8):1425-32.
569   33.   Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of
570   supplementary search techniques and recommendations in review methodology handbooks: a
571   methodological review. Syst Rev. 2017;6(1).
572   34.   Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to
573   literature searching in systematic reviews: a literature review of guidance and supporting studies.
574   BMC Medical Research Methodology. 2018;18(1).
575   35.   Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured
576   methodological review. Syst Rev. 2016;5(1).

577

578