

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sustaining community-based interventions for people affected by dementia long term: The SCI-Dem realist review
AUTHORS	Morton, Thomas; Wong, Geoff; Atkinson, Teresa; Brooker, Dawn

VERSION 1 – REVIEW

REVIEWER	Joseph Sawyer University College London, Marie Curie Palliative Care Research Department
REVIEW RETURNED	18-Dec-2020

GENERAL COMMENTS	<p>To the authors,</p> <p>Thank-you for opportunity to read this article and indeed for all your efforts in putting what appears to be an enormous amount of work together. It is a fantastic paper on a topic of great importance and relevance. You have assimilated and comprehensively analysed a huge amount of data to give a set of findings and recommendations that are both practical and novel in this area. You are also to be commended for such a creative approach in presenting your vast data. The supplementary files provide great depth and transparency to your workings allowing the reader to read the paper at several different levels. The methodological description is comprehensive and the involvement of key stakeholders throughout adds great value to the results. CMOCs follow a clear pattern and understanding of what constitutes a context, mechanism and outcome in relation to the context of this study. The clarity here helps guide the reader through what can be quite a vast topic.</p> <p>My only suggestions relate to the figures in supplementary file 4: I am unsure as to what the colours within each individual graphic refer to, there appears to be a pattern e.g. green, brown and black but I am not sure what this represents. A key for this would therefore be helpful to the reader.</p> <p>Congratulations on such a comprehensive piece of work.</p>
-------------------------	---

REVIEWER	Rebecca Palm University Witten Herdecke Faculty of Health, Department of Nursing Sciences
REVIEW RETURNED	30-Dec-2020

GENERAL COMMENTS	<p>Methods: Page 7 line 28: Please also define outcomes (are outcomes measured on the level of people with dementia or their relatives, are process measures also considered as outcomes?)</p>
-------------------------	--

	<p>Page 8 line 15: Could you please give more information on the initial program theory and how this was developed? Did you develop CMO configurations for the IPT?</p> <p>Supplementary files: The description of the search syntax is very brief. Usually, for every database an own syntax is used because the databases use different terms as key words (e.G. MeSH in PubMed). They syntax does not specify if or which defined keywords were used. Please explain this. If database defined key words were not used for the search this may have consequences for the search results. Please discuss this issue.</p> <p>The selection process was carried out mainly by one author. This should be considered as a limitation and discussed.</p> <p>Methods: You exclude interventions for people with severe dementia. This aspect should be made clear in the background and research question, maybe also in the title or abstract.</p> <p>I would also recommend to give a reason why interventions for people with severe dementia are excluded. I can imagine that the sustainability of interventions for people with severe dementia is even more difficult than for those with milde to moderate dementia.</p> <p>Methods: The analysis was carried out by only one author. This should be considered as a limitation.</p> <p>Methods: Please add an overview of your coding scheme. It would be of interest, which codes were defined based on the IPT and which were developed inductively. A description of the IPT would in general be of value.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

- *“I am unsure as to what the colours within each individual graphic [in Supplementary File 4] refer to, there appears to be a pattern e.g. green, brown and black but I am not sure what this represents. A key for this would therefore be helpful to the reader.”*

As requested, we have added the follow text as a key to the text colours used in what was Supplementary File 4 (Now Supplementary File 5):

*“Text colour key: **Black text** = sub-outcome directly impacting upon “getting/keeping”
Green text = likely to help sub-outcome **Red text** = likely to hinder sub-outcome”*

Reviewer 2:

- *“Page 7 line 28: Please also define outcomes (are outcomes measured on the level of people with dementia or their relatives, are process measures also considered as outcomes?)”*

We have added a line to clarify what can be identified as “outcomes” at this point in the manuscript, as follows:

“... outcomes can be “either intended or unintended and can be proximal, intermediate, or final”²⁷ and in this review refer to any identifiable result (of the interaction between contexts and mechanisms) that can directly have a bearing on an

intervention's ability to sustain long-term.”

- *“Page 8 line 15: Could you please give more information on the initial program theory and how this was developed? Did you develop CMO configurations for the IPT?”*

AND

“Please add an overview of your coding scheme. It would be of interest, which codes were defined based on the IPT and which were developed inductively. A description of the IPT would in general be of value.”

An account of how the IPT was developed and what form it took has been added at this point in the manuscript, with the following text:

“This model began as two diagrams (one regarding engagement, one regarding sustainability) drawn up by TM and TA by batching issues raised at the March workshop, and possible links between them. These diagrams were then discussed, altered and added to iteratively over four months as new stakeholder input became available (these can be seen in Supplementary file 1). These diagrams were speculative so kept deliberately broad and fluid in focus, as a work in progress. Detailed analysis of possible context-mechanism-outcome configurations (CMOCs) was not considered appropriate at this stage, as: 1) Not enough data had been gathered; 2) This would be both labour intensive and too limiting for a model whose purpose was only as a steering guide to inform the review proper, yet to be undertaken.”

In addition a new Supplementary file (now Supplementary file 1) has been added and referred to in this part of the *Methods* section (*Step 1: Locating existing theories*), so it is clear what elements the IPT included.

The top-level headings from these diagrams were used to create deductive coding categories for NVivo analysis. Also in Supplementary file 1 is an overview of the NVivo categories/coding scheme. The categories shown are top-level “parent” codes, not shown is that each has multiple “children” codes. Codes that are not mirrored in the headings of the IPT diagrams will be inductive codes, created in response to the data content. I have now mentioned this under *Steps 3 and 4: Article selection, data extraction and organization* where the NVivo analysis is discussed in the *Methods* section, with the following text:

“(An overview of top-level ‘parent’ codes can also be seen in Supplementary file 1; deductive codes can be identified in that they mirror the headings of the initial model diagrams).”

- *“The description of the search syntax is very brief. Usually, for every database an own syntax is used because the databases use different terms as key words (e.g. MeSH in PubMed). They syntax does not specify if or which defined keywords were used. Please explain this. If database defined key words were not used for the search this may have consequences for the search results. Please discuss this issue.”*

Search terms were kept uniform across all databases and searching was carried out by looking for the occurrence of these terms within the title, abstract and key words of documents in each database. If a database did not allow for this, the strategy was altered slightly to the closest option (e.g. in ProQuest this was searching everywhere in a document except full text; in PubMed this was by carrying out three separate searches by title content, by abstract content and key word content, then combining the results).

We used this strategy instead of using each database's own defined keywords because the types of intervention we wished to include were not only very diverse but often informally run, without a common agreed terminology. Using too narrowly-specified terms would have resulted in an unmanageably voluminous list of possible key words, without necessarily locating better targeted results. A long list of terms used by stakeholders and in extant literature was refined through piloting searches in multiple databases, including and excluding each term to note the difference in results in terms of whether it broadened or narrowed the number of relevant hits. However, our search needed to be broad and inclusive in the first instance as, due to the lack of standard terminology and the atypical (i.e. not efficacy/effectiveness-focused) nature of our research question, manual screening would be key in determining relevance.

It should also be noted that formal searching in a realist review differs from a standard Cochrane-style realist review in a number of ways:

- It seeks to include a wide diversity of sources, study and document types containing data in enough abundance to build a robust theory, rather than comprehensively locating every study of a certain type on a certain question.
- It seeks to locate relevant information within documents that themselves may not share the same research questions (in this case unlikely to).
- The formal search may be supported by multiple supplementary and informal searches and stakeholder input, and as such is unlikely to be perfectly replicable.

Therefore, aside from "Dementia", which our pilot searches found to be a standard, universal key term that caught all of our participant population without unhelpfully limiting it, we found using a database's defined key terms was often unhelpfully limiting and misleading in searches regarding the data we were trying to find. The downside of this was that we had to accept a higher ratio of irrelevant search hits which then had to be excluded through manual screening of title and abstract.

We have added a clarification to the search strategy outline in what was Supplementary file 1 (now Supplementary File 2) with the following text:

"Search terms were kept uniform across all databases and searching was carried out by looking for the occurrence of these terms within the title, abstract and key words of documents in each database. If a database did not allow for this, the strategy was altered slightly to the closest option (e.g. in ProQuest this was searching everywhere in a document except full text; in PubMed this was by carrying out three separate searches by title content, by abstract content and key word content, then combining the results)."

We have also added a few lines of discussion in the *Methods* section under the heading *Step 2: Search for evidence*, with the following text:

"Search terms were kept uniform across all databases and searching was carried out by looking for the occurrence of these within the title, abstract and key words of documents (or nearest equivalent) in each database. Database-specific defined keywords were not used as the types of intervention were not only very diverse but often without a common agreed terminology, hence using too narrowly-specified terms would have resulted in an unmanageably voluminous list of possible key words, without necessarily locating better targeted results, and could be limiting and misleading. In addition the nature of this review's research question is atypical in that it does not have an efficacy/effectiveness focus in common with many of its sources of data, hence manual screening was key in determining relevance. A disadvantage

of this was that we had to accept a higher ratio of irrelevant search hits which then had to be excluded through manual screening of title and abstract.”

- *“You exclude interventions for people with severe dementia. This aspect should be made clear in the background and research question, maybe also in the title or abstract. I would also recommend to give a reason why interventions for people with severe dementia are excluded. I can imagine that the sustainability of interventions for people with severe dementia is even more difficult than for those with mild to moderate dementia.”*

Those in advanced stages with severe dementia are more likely to be receiving at least some standard help via the local health and social care pathway due to high and acute needs, and also less likely to be living independently in the community. Hence place-based community interventions tend to serve those in the earlier stages, but are also more likely to be grass-roots/thirdsector/volunteer run rather than driven by health or social care authorities. Hence these in particular tend to have problems sustaining long-term due to the lack of official support and recognition. It is in the earlier stages that it often falls to family and friends to provide all care with no universal, consistent and standard intervention offered.

Though, of course, on a continuum, those with living with milder forms of dementia at earlier stages have very different needs to those with more severe forms at more advanced stages, so interventions for each demographic are quite distinct: substantially different in their aims, their content, how they are delivered and even where they can be delivered. By their nature, community-based interventions where people come together to meet outside of their home are likely to serve those who are towards the start of their dementia journey rather than those at an advanced stage, and are distinct from more acute care. A few lines on this has been added under *Step 2: Search for evidence* in the *Methods* section, with the following text:

“Interventions exclusively for those with severe dementia at advanced stages were excluded as these were not the focus of this review. As those with severe dementia have high needs and are less likely to be living independently in the community, by their nature community-based interventions where people meet outside of their home are likely to serve those who are towards the start of their dementia journey rather than those at an advanced stage, and are distinct from more acute care.”

“Post-diagnosis” will generally refer to people in the earlier stages (though this is of course not the case in every instance of diagnosis) so this has been clarified in the abstract. The introductory two paragraphs already talk about “Improving provision of early, post-diagnosis support” and gaps in provision of “support following a diagnosis of dementia”, but *“for those with less severe symptoms”* has been added.

- *“The selection process was carried out mainly by one author. This should be considered as a limitation and discussed.”*

AND

“The analysis was carried out by only one author. This should be considered as a limitation.”

The practice of one researcher (in this case TM) carrying out the bulk of article screening, article selection relevance and rigour assessment and data analysis, with a second researcher (in this case TA) independently checking 10% at each stage for consistency, is common in realist review/synthesis research. In addition to this, there was regular discussion throughout with the whole team – in particular DB and GW overseeing project progress and methodological integrity respectively – which included discussion and input into all of the above stages. Nevertheless, that the above tasks were carried out mainly by only one

researcher can be seen as a limitation, as in Cochrane-style systematic reviews double-screening by two reviewers independently is recommended for greater reliability of results. However, it should be noted that realist review is a theory-driven, interpretive approach hence does not have quite the same aims as a Cochrane-style systematic review;³⁰ i.e. the aim is to develop an evidence-informed theory rather than a comprehensive summation of all research data available on a particular research question. The following lines discussing this have been added under *Strengths and limitations* in the *Discussion* section:

“The practice of one researcher carrying out the bulk of article selection and data analysis, with a second researcher independently checking 10% at each stage for consistency (along with regular input and discussion with other members of the research team) is common in realist review, but nevertheless can be seen as a limitation, as in Cochrane-style systematic reviews double-screening by two reviewers independently is recommended for greater reliability of results. However, it should be noted realist review is a theory-driven interpretive approach with significant differences to more traditional forms of systematic review;³⁰ i.e. the aim is to develop an evidence-informed theory rather than a comprehensive summation of all research data available on a particular research question.”

VERSION 2 – REVIEW

REVIEWER	Rebecca Palm University Witten Herdecke Faculty of Health, Department of Nursing Sciences
REVIEW RETURNED	10-Mar-2021
GENERAL COMMENTS	<p>Dear authors,</p> <p>the manuscript is now very transparent in its methods which is important because realist reviews are still a rather unknown methodology (at least outside UK and commonwelath countries). Although the manuscript is acceptable for publication in my opinion, I would still like to comment on two points: I can understand that just one reviewer screened the large number of publications, because other is almost not feasible given the constrained staff resources in research and the high numbers of records. But nevertheless I think that a review process that was conducted by two researchers is always superior to a review process conducted by one researcher - and not only in Cochrane reviews. Because realist reviews have different aims than Cochrane reviews does not exempt them from the same quality requirements. The chance to exclude an important paper is smaller with two reviewers and this is also important for realist reviews.</p> <p>The other apsect is that an advanced literature search that uses database defined key terms (e.g. for the term dementia) may be more precise without excluding important papers. Search filters that are tested for reliability and that also allow a narrow search could also be benficiary. And again - this has in my opinion nothing to do with the review type. Advanced literature searches are not reserved for Cochrane reviews but are important for all kinds of reviews, irrespective of their aim.</p>

VERSION 2 – AUTHOR RESPONSE

In response to Reviewer Report:

With regards to point one (one researcher screening the bulk of records in a formal search):

We agree a missing a study is indeed very important in an aggregative review (e.g. a Cochrane review) because, to put it crudely, such reviews seek out and add up all the numbers from a study for data analysis. So, missing a large and important study because of screening or searching deficiencies may introduce significant threats to the validity of the findings for an aggregative review. This is, however, not the case in a realist review, which is a configurational type of review and has a more explanatory focus. In other words, in realist reviews data are used in a 'configurational way' to produce theories that explain causation. The explanatory powers of a theory are not necessarily dependent on having to find each and every piece of data that would support the theory [see: Data gathering for realist reviews: Looking for needles in haystacks. Wong G. In: Emmel N, Greenhalgh J, Manzano A, Monaghan M, Dalkin S, editors. Doing Realist Research. London: Sage, 2018]. As such, it is much less likely that threats to the validity of the findings of a realist review would result from the occasional omissions from single screening of the citations and/or less than comprehensive searches. On the issue of errors introduced from single screening, what might matter in terms of threats to the validity of our findings might come from systematic errors, which is the reason for 10% checking at various stages in the screening and quality assessment process, to ensure a consistent and standardised approach has been taken. It is because of the reasons we have given above that it is commonly accepted practice in realist reviews that screening process we have adopted is used [see: Does therapeutic writing help people with long-term conditions? Systematic review, realist synthesis and economic considerations. Nyssen OP, Taylor S, Wong G, Steed L, Bourke L, Lord J, Ross C, Hayman S, Field V, Higgins A, Greenhalgh T, Meads C. Health Technology Assessment 2016;20; Interventions to improve antimicrobial prescribing of doctors in training (IMPACT): a realist review. Papoutsis C, Mattick K, Pearson M, Brennan N, Briscoe S, Wong G. Health Serv Deliv Res 2018;6(10); Interventions to minimise doctors' mental ill-health and its impacts on the workforce and patient care: the Care Under Pressure realist review. Carrieri D, Pearson M, Mattick K, Papoutsis C, Briscoe S, Wong G, Jackson M. Health Serv Deliv Res 2020;8(19)].

With regards to point two (on the use or otherwise of database-defined key terms):

We accept that using database-defined key terms may have helped us be more precise in locating papers to an extent, for example with regards to the term "Dementia". However we did not take this route - not so much because of the review type, but because of the unusual nature of the research question, in which we were looking for a "broad style" of intervention rather than a specific type, precluding the use of standardised terms as. While such interventions are identifiable by the gap they fill and the common functions they share, they still vary greatly not just in their surface aims but also in how they conceive of and present themselves (i.e. there are no universally accepted standardised terms that link them). It is possible however, that using some database-defined terms where possible, together with broader terms where not, could have streamlined our search and we will consider this if tackling a similar project in the future.