

## 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

<b>TITLE (PROVISIONAL)</b>	Protocol for a systematic review of preference-based instruments for measuring care-related outcomes and their suitability for the palliative care setting
<b>AUTHORS</b>	McCaffrey, Nikki; Al-Janabi, Hareth; Currow, David; Hofeman, Renske; Ratcliffe, Julie

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Margarita Corry School of Nursing & Midwifery, Trinity College Dublin, Ireland
<b>REVIEW RETURNED</b>	26-Apr-2016

<b>GENERAL COMMENTS</b>	<p>This protocol is well written and the authors are experienced in conducting systematic reviews. There is clear evidence of subject specific knowledge and expertise is evident throughout the protocol. The introduction provides the backdrop and clearly sets out the rationale for the review.</p> <p>Despite this I have a few points that need further consideration:</p> <ol style="list-style-type: none"> <li>1. It is now mid-April 2015 so the search date needs to be changed from 31st March 2016. I suggest that you state 'date of search' as that will enable you to ensure that all published literature at the time of searching is included.</li> <li>2. It may be beneficial to have a librarian with expertise in systematic review searching as part of the team/available to the team for the conduct of the search.</li> <li>3. Page 5, line 34 you write "Consequently, the general aim....." I think this would be better phrased with a definite aim by omission of the words 'the general'</li> <li>4. The heading 'objectives' is misleading as you present research questions (should this be review questions?) - I recommend that the content is phrased as objectives or the title is changed to 'review questions'. The use of the term 'suitable' in the 3rd question is vague. A more precise term would strengthen this question/objective and guide the criteria used for measurement/determination.</li> <li>5. On page 4, lines 52-54 you explain what is meant by the overall term 'Preference-based measures'. I feel it is important that you provide a definition of what is meant by 'preference-based measures' in the context of this review. I am unclear as to whether the search will cross all conditions across clinical areas or if it will be confined to palliative care conditions across clinical areas. From what is written on page 5 lines 38-41 it appears that it is confined to</li> </ol>
-------------------------	---

	<p>the palliative care setting but this needs to be clarified so that there is no misunderstanding from the reader's perspective.</p> <p>6. Selection process On page 7, you state that a randomly selected subset of citations 10% will be independently screened. It is common and best practice for all titles and abstracts to be independently screened for inclusion. Any deviation from that needs to be well justified.</p> <p>Who will randomly select the subset of citations and why 10% - and at what point will further duplicate reviewing be considered a requirement. You have Kappa &lt;4 in brackets but it is not clear if this is the point that further duplicate reviewing will be conducted and if so to what extent. I would suggest that within a 10% subset review agreement levels should be high (at least 0.8). Anything less that this would suggest that additional duplicate reviewing is required.</p> <p>7. Data collection Page 8, Line 8: You state that 'one reviewer will extract data independently'. It is good practice in the conduct of systematic reviews for at least two authors to extract data independently. A process should be explicated as to how discrepancies will be managed.</p> <p>Page 8 –line 28 you state that information on how the content of the instrument was developed, content validity ..... will be used to assess the quality and suitability of the instrument for the palliative care setting. I think that you need to have a set structure/approach based on pre-determined criteria for determining this 'suitability'.</p> <p>A randomly selected subset (25%) of included instruments will be selected for independent review. Who will select this subset? Why 25% (earlier you stated that small numbers of papers are expected)</p> <p>8. Data synthesis: A more detailed description of the approach that will be adopted to the narrative summary is required. While I acknowledge that there are few if any published methods for conduction a narrative summary details of how best practice guidelines will be adhered to can be included (see Green et al. 2001 Writing narrative literature review for peer reviewed journals: secrets of the trade, J. Sports Chiropr Rehab, 15(5): 5-19 for further details/inspiration on this issue). Are there any more advanced methods of data synthesis that could be used in this review?</p> <p>9. How will reporting bias be addressed?</p>
--	---

<b>REVIEWER</b>	Elizabeth Buckley Research Fellow University of South Australia Australia
<b>REVIEW RETURNED</b>	03-May-2016

<b>GENERAL COMMENTS</b>	This is a carefully planned and well written systematic review protocol that is very likely to yield results that will be useful to health economists and palliative care professionals. Some further clarity and justification around the critical appraisal process would provide
-------------------------	---

	<p>additional rigour to the systematic review.</p> <p><b>Abstract</b> The list of bibliographic databases to be searched includes both Medline and PubMed, as the primary component of PubMed is Medline, it would not be necessary to search both.</p> <p><b>Introduction</b> In the second paragraph, the notation for Australian currency/dollars is AUD and this should be used consistently when referring to Australian currency. In the description of health care costs, "...and between one third and one fifth of overall costs in the UK and Canada...." Does this refer to health care costs specifically?</p> <p><b>Methods</b> Selection process – "Secondly, the full text articles of potentially eligible studies will be retrieved and independently assessed for inclusion..." It would be helpful to clarify to what this is independent of? For example, is this independent of the first selection process? Or is it independent of the lead investigator. With the following sentence also referring to an independent selection process, it may be confusing to the reader.  Will the assessment of agreement between reviewers of articles be undertaken prior to resolving disagreement? Otherwise it may be a moot point once disagreement has been resolved.  The sentence referring to pearling appears to be a duplication of a previous sentence (Search Strategy section) and may also be redundant.</p> <p><b>Data Collection</b> Regarding the use of reporting checklists as critical appraisal tools, for example, CREATE, are they suitable for use as critical appraisal checklists - are the reporting items also able to measure study quality? Or do they just facilitate better reporting and subsequently, also facilitate the assessment of critical appraisal? If an item is reported, does that provide an assessment of the likelihood that study bias has been minimised?  In relation to the following statement, "Information on how the content of the instrument was developed, content validity and the populations used to test the instrument will be used to assess the quality and suitability of the instrument for the palliative setting." Will there be objective measures used to determine quality and suitability?</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

**Reviewer 1**

Reviewer Name: Margarita Corry

Institution and Country: School of Nursing & Midwifery, Trinity College Dublin, Ireland  
Competing Interests: none

*This protocol is well written and the authors are experienced in conducting systematic reviews. There is clear evidence of subject specific knowledge and expertise is evident throughout the protocol. The introduction provides the backdrop and clearly sets out the rationale for the review.*

**Response:** Thank you for the positive feedback and constructive comments which have been addressed below.

*Despite this I have a few points that need further consideration:*

1. *It is now mid-April 2015 so the search date needs to be changed from 31st March 2016. I suggest that you state 'date of search' as that will enable you to ensure that all published literature at the time of searching is included.*

**Response:** Thank you for this helpful suggestion. The search date has been removed given the uncertainties of the exact date the search will be conducted.

2. *It may be beneficial to have a librarian with expertise in systematic review searching as part of the team/available to the team for the conduct of the search.*

**Response:** Thank you for this suggestion. The Flinders University Health Sciences Liaison librarian will be available to the team for the conduct of the study.

3. *Page 5, line 34 you write "Consequently, the general aim....." I think this would be better phased with a definite aim by omission of the words 'the general'*

**Response:** The word "general" has been removed as recommended.

4. *The heading 'objectives' is misleading as you present research questions (should this be review questions?) - I recommend that the content is phrased as objectives or the title is changed to 'review questions'. The use of the term 'suitable' in the 3rd question is vague. A more precise term would strengthen this question/objective and guide the criteria used for measurement/determination.*

**Response:** The heading 'objectives' has been changed to 'review questions' as suggested.

The term 'suitable' has been removed and the 3rd question has been reworded (p6) to correlate with the assessment of suitability described in the data collection section (p9).

"3. Have preference-based instruments measuring care-related outcomes been validated and valued in the palliative setting?"

5. *On page 4, lines 52-54 you explain what is meant by the overall term 'Preference-based measures'. I feel it is important that you provide a definition of what is meant by 'preference-based measures' in the context of this review. I am unclear as to whether the search will cross all conditions across clinical areas or if it will be confined to palliative care conditions across clinical areas. From what is written on page 5 lines 38-41 it appears that it is confined to the palliative care setting but this needs to be clarified so that there is no misunderstanding from the reader's perspective.*

**Response:** All clinical areas will be included in step one of the search. The text (pp5-7) has been amended to clarify the clinical setting and what is meant by preference-based measure in the context of this review. Specifically, in relation to the review question, the text has been amended as follows:

"What preference-based instruments are available generally, i.e. in any clinical setting, to measure care-related outcomes?"

6. *Selection process: On page 7, you state that a randomly selected subset of citations 10% will be independently screened. It is common and best practice for all titles and abstracts to be independently screened for inclusion. Any deviation from that needs to be well justified.*

**Response:** Ideally, whilst all titles and abstracts should be independently screened for inclusion, a pragmatic decision was made to initially triple screen 10% of randomly selected citations based on the high volume of citations elicited from the pilot searches, low volume of likely inclusions (based on the knowledge of the authorship team), the expertise and experience of the primary reviewer in conducting systematic reviews (NM) and research resource constraints (see Goodrich, Kaambwa & Al-Janabi (2012)<sup>1</sup> as example).

*Who will randomly select the subset of citations and why 10% - and at what point will further duplicate reviewing be considered a requirement. You have Kappa <0.4 in brackets but it is not clear if this is the point that further duplicate reviewing will be conducted and if so to what extent. I would suggest that within a 10% subset review agreement levels should be high (at least 0.8). Anything less that this would suggest that additional duplicate reviewing is required.*

**Response:** NM will randomly select the subset of citations. Once duplicates have been removed, all remaining citations from the initial search will be allocated a unique identifying number in sequence. The subset of citations will be identified using the random number generator in Excel and corresponding citation.

A pragmatic decision was taken to triple screen 10% of citations based on the systematic review conducted by Goodrich, Kaambwa & Al-Janabi (2012)<sup>1</sup> and reasons outlined above.

Further duplicate reviewing was initially to be considered if the Kappa statistic is less than 0.4. Based on the reviewer's suggestion, the threshold has been amended to 0.8. If the inter-rater reliability is lower than this threshold an additional subset of citations (25%) will be double screened. If the inter-rater reliability is still lower than 0.8, all remaining citations will be double screened. This process has been revised in the manuscript (p8).

"A randomly selected subset of citations (10%) will be independently screened by two other members of the research team. Agreement between the reviewers will be assessed using the Kappa statistic to determine if further duplicate reviewing is required. If the inter-rater reliability is lower than 0.8, i.e. strong agreement,<sup>43</sup> an additional subset of citations (25%) will be independently screened. If the inter-rater reliability is still less than 0.8, all remaining citations will be independently screened. Any disagreements between reviewers will be resolved by consensus amongst the research team."

7. *Data collection: Page 8, Line 8: You state that 'one reviewer will extract data independently'. It is good practice in the conduct of systematic reviews for at least two authors to extract data independently. A process should be explicated as to how discrepancies will be managed.*

**Response:** Ideally, whilst data should be extracted independently by at least two reviewers, single-reviewer data extraction is not unprecedented in this field<sup>2 3 4</sup> and the limitations of single reviewer data extraction (potential bias) will be acknowledged in the discussion of the final results. Findings will need to be interpreted within this context.

Unfortunately, data extraction by two reviewers is unfeasible for this review due to research resource constraints and a pragmatic decision was made to take the stated approach. Where there is ambiguity, input will be sought from a second reviewer (RH). Two of the reviewers (HAJ & RH) will oversee data extraction to help ensure rigour. Once again, any disagreements between reviewers will be resolved by consensus amongst the research team. This process has been described in the manuscript (p9) as follows:

"One reviewer will extract data independently. Where there is ambiguity, input will be sought from a second reviewer (RH). Two of the reviewers (HAJ & RH) will oversee data extraction to help ensure rigour. Any disagreements between reviewers will be resolved by consensus amongst the research team."

Page 8 –line 28 you state that information on how the content of the instrument was developed, content validity will be used to assess the quality and suitability of the instrument for the palliative care setting. I think that you need to have a set structure/approach based on pre-determined criteria for determining this ‘suitability’.

**Response:** The degree of suitability of the instruments for palliative economic evaluations will be assessed using the International Society for Quality of Life Research minimum standards for patient-reported outcome measures, the checklist for reporting valuation studies of multi-attribute utility-based instruments and information on the development of the instrument in the palliative setting (abstract, p3). Further details about how this will be approached in a structured manner have been provided in the ‘Data Collection’ section as follows (p9):

“Information on how the contents of the instruments were developed, psychometric properties and valuation will be used to assess the suitability of the instruments for the palliative setting; ; instruments will be scored on whether the domains or dimensions were developed using input from informal caregivers of people receiving palliative care (yes/no) and whether each of the reporting checklist items has been evaluated for this population (if yes, then a score of one will be allocated), and a total score calculated.

*A randomly selected subset (25%) of included instruments will be selected for independent review. Who will select this subset? Why 25% (earlier you stated that small numbers of papers are expected)*

**Response:** Thank you pointing out this inconsistency. As small numbers of instruments are anticipated the second reviewer will independently assess the quality and suitability of ALL the instruments for the palliative setting. The manuscript has been updated accordingly (p9).

“Two reviewers will independently assess the suitability of the preference-based instruments...”

8. *Data synthesis: A more detailed description of the approach that will be adopted to the narrative summary is required. While I acknowledge that there are few if any published methods for conducting a narrative summary details of how best practice guidelines will be adhered to can be included (see Green et al. 2001 Writing narrative literature review for peer reviewed journals: secrets of the trade, J. Sports Chiropr Rehab, 15(5): 5-19 for further details/inspiration on this issue). Are there any more advanced methods of data synthesis that could be used in this review?*

**Response:** A more detailed description of the approach that will be adopted to the narrative summary has been added to the ‘Data synthesis’ section (p9), as follows:

“A narrative summary of the included studies and instruments will be provided in accordance with recent guidance from the Cochrane Collaboration®. Four tables summarising the main features of the included studies, instrument descriptions, psychometric properties and valuation will be presented (see ‘Data Collection’). Using these data, similarities and differences between the instruments will be described, including their suitability for measuring care-related outcomes in palliative care and possible reasons for variations between instruments will be explored, e.g. variability in study populations and method of content development. Recommendations for practice on instruments for measuring care-related outcomes in palliative care economic analyses will be given.”

As the numbers of instruments is anticipated to be small, more advanced methods of data synthesis will not be used.

9. *How will reporting bias be addressed?*

**Response:** The total number of preference-based instruments measuring care-related outcomes is anticipated to be small based on the reviewers’ knowledge of this field.

Consequently, to minimise reporting bias, the instrument development groups will be contacted to make sure no crucial information or developments have been missed (p9).

## **Reviewer: 2**

Reviewer Name: Elizabeth Buckley

Institution and Country: Research Fellow, University of South Australia, Australia Competing Interests: none

*This is a carefully planned and well written systematic review protocol that is very likely to yield results that will be useful to health economists and palliative care professionals. Some further clarity and justification around the critical appraisal process would provide additional rigour to the systematic review.*

### Abstract

*The list of bibliographic databases to be searched includes both Medline and PubMed, as the primary component of PubMed is Medline, it would not be necessary to search both.*

**Response:** Thank you for this advice. The abstract and methodology have been updated accordingly (p3 & p7).

### Introduction

*In the second paragraph, the notation for Australian currency/dollars is AUD and this should be used consistently when referring to Australian currency.*

**Response:** AUD\$ has been added to the text in the manuscript wherever Australian currency is mentioned. Thank you for spotting this error.

*In the description of health care costs, "...and between one third and one fifth of overall costs in the UK and Canada...." Does this refer to health care costs specifically?*

**Response:** This description refers to overall costs, i.e. health, social, informal care and out-of-pocket costs. The text has been altered to provide further clarity (p5).

### Methods

*Selection process – "Secondly, the full text articles of potentially eligible studies will be retrieved and independently assessed for inclusion..." It would be helpful to clarify to what this is independent of? For example, is this independent of the first selection process? Or is it independent of the lead investigator. With the following sentence also referring to an independent selection process, it may be confusing to the reader.*

**Response:** The text has been revised as follows, to aid clarity.

"Secondly, the full text articles of potentially eligible studies will be retrieved and independently assessed from the first selection process for inclusion in the review."

*Will the assessment of agreement between reviewers of articles be undertaken prior to resolving disagreement? Otherwise it may be a moot point once disagreement has been resolved.*

**Response:** Yes, the assessment of agreement between reviewers of articles will be undertaken prior to resolving disagreement and the text has been restructured to reflect this ordering.

“A randomly selected subset of citations (10%) will be independently screened by two other members of the research team. Agreement between the reviewers will be assessed using the Kappa statistic to determine if further duplicate reviewing is required. If the inter-rater reliability is lower than 0.8, i.e. strong agreement,<sup>43</sup> an additional subset of citations (25%) will be independently screened. If the inter-rater reliability is still less than 0.8, all remaining citations will be independently screened. Any disagreements between reviewers will be resolved by consensus amongst the research team.”

*The sentence referring to pearling appears to be a duplication of a previous sentence (Search Strategy section) and may also be redundant.*

**Response:** Thank you for pointing out this duplication. This duplicate sentence has been now been removed from the manuscript.

#### *Data Collection*

*Regarding the use of reporting checklists as critical appraisal tools, for example, CREATE, are they suitable for use as critical appraisal checklists - are the reporting items also able to measure study quality? Or do they just facilitate better reporting and subsequently, also facilitate the assessment of critical appraisal? If an item is reported, does that provide an assessment of the likelihood that study bias has been minimised?*

**Response:** The checklists facilitate better reporting, thereby aiding evaluation of the suitability of the preference-based instruments for measuring outcomes in palliative care. Instruments will be scored on whether each of the reporting checklist items has been evaluated for informal caregivers of people receiving palliative care (if yes, then a score of one will be allocated) and a total score calculated (p9).

*In relation to the following statement, “Information on how the content of the instrument was developed, content validity and the populations used to test the instrument will be used to assess the quality and suitability of the instrument for the palliative setting.” Will there be objective measures used to determine quality and suitability?*

**Response:** Thank you for raising this point. Instruments will be scored on whether the domains or dimensions were developed using input from informal caregivers of people receiving palliative care (yes/no) and whether each of the reporting checklist items has been evaluated for this population (if yes, then a score of one will be allocated), and a total score calculated (p9)

#### **References**

1. Goodrich K, Kaambwa B, Al-Janabi H. The inclusion of informal care in applied economic evaluation: a review. *Value Health* 2012;**15**(6):975-81.
2. Bhattarai N, McMeekin P, Price C, et al. Economic evaluations on centralisation of specialised healthcare services: a systematic review of methods. *BMJ Open* 2016;**6**(5).
3. Kigozi J, Jowett S, Lewis M, et al. Estimating productivity costs using the friction cost approach in practice: a systematic review. *The European Journal of Health Economics* 2016;**17**(1):31-44.
4. Wittenberg E, Prosser LA. Disutility of illness for caregivers and families: a systematic review of the literature. *Pharmacoeconomics* 2013;**31**(6):489-500.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Margarita Corry Trinity College Dublin, Ireland
<b>REVIEW RETURNED</b>	13-Jun-2016

<b>GENERAL COMMENTS</b>	<p>The authors have addressed most of the concerns highlighted in the original review of the protocol and this will enhance the quality of the review. I feel that a few of the issues noted have not been fully addressed.</p> <ol style="list-style-type: none"> <li>1. The fact that the librarian will be available to the reviews needs to be explicated in the actual protocol itself and should include the role that he/she will have eg. Searching etc.</li> <li>2. Page 5, review questions point 3: What do you mean by “valued in palliative care setting?”</li> <li>3. Page 6, search strategy: It is common to indicate the span of the data search eg. Databases from inception to date of search.</li> <li>4. The view that data extraction by two reviewers is unfeasible is not acceptable. Quality systematic reviews require independent data extraction. If the reviewers were following a recommended framework for conducting a systematic review eg. Cochrane guidelines or other accepted approach, independent data extraction would be a requirement. Reporting this as a limitation is not acceptable given that there are five authors doing the review. In addition, on page 7 the authors have indicated that they expect the numbers to be small based on their knowledge of the field. It is therefore difficult to understand how independent data extraction could be onerous.</li> </ol>
-------------------------	--

<b>REVIEWER</b>	Elizabeth Buckley Research Fellow, School of Health Sciences, University of South Australia
<b>REVIEW RETURNED</b>	31-May-2016

<b>GENERAL COMMENTS</b>	All queries that were raised in the initial peer review have been well and adequately addressed by the authors. Thank you.
-------------------------	--

### VERSION 2 – AUTHOR RESPONSE

Thank you for your comments and the opportunity to submit the revised manuscript for publication in your Journal. We have addressed the outstanding reviewer’s comments and indicated where changes have been made in the manuscript.

In summary, the manuscript has been revised to address the reviewers’ points as follows:

1. The involvement and role of the librarian has been explicated in the protocol (p6);
2. The third review question has been expanded to aid clarity (p5);
3. The span of the data search has been provided (p7);
4. The data extraction process has been amended as advised. Two reviewers will now extract the data independently (NM;RH). These details have been added to methods section (p8);

We appreciate the constructive comments and trust the revisions address the points raised. A more

detailed response to the reviewers' comments has been included in the submitted documents.

### VERSION 3 - REVIEW

<b>REVIEWER</b>	Margarita Corry Trinity College Dublin, Ireland
<b>REVIEW RETURNED</b>	10-Jul-2016

<b>GENERAL COMMENTS</b>	Thank you for submitting this paper and addressing all of the concerns that I raised in my previous reviews. I wish you all the best in taking this work to full review stage.
-------------------------	--