

## 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> | Do physical therapists follow evidence-based guidelines when managing musculoskeletal conditions? A systematic review |
| <b>AUTHORS</b>             | Zadro, Joshua; O'Keeffe, Mary; Maher, Christopher   |

### VERSION 1 – REVIEW

|                        |                                       |
|------------------------|---------------------------------------|
| <b>REVIEWER</b>        | Derek Clewley<br>Duke University, USA |
| <b>REVIEW RETURNED</b> | 21-Jun-2019                           |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | This is a very good paper. Perhaps the only suggestion I would have is to add a discussion point regarding how the quality of the studies that were pulled in this SR could impact the results. The D&B was scored, but maybe something more about how this could influence the results. The only other minor comment is the number of tables, but I think this enhances the results and as long as journal is okay with the number then good to go. |
|-------------------------|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Nicholas Taylor<br>La Trobe University, Australia |
| <b>REVIEW RETURNED</b> | 10-Jul-2019                                       |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | <p>The concept of this systematic review is excellent, the scope is ambitious and the manuscript represents an enormous amount of work. However, the review as presented is difficult to follow. One difficulty is that the unit of measure is not always clear, whether percentages are percentage of physical therapists, percentage of patients, or percentage of treatments. Another difficulty is that the methods of data synthesis are not clear to the reader, whether counts were summed across studies at the patient/therapist/treatment level and presented as an overall percentage, or whether the percentage from the studies of a the specific treatment were expressed as a median percentage, or perhaps some other method like a meta-analysis of proportions? A further difficulty is whether or not a treatment was judged to be accepted was based on whether it was 'recent (line 146)' or 'well-recognised' (line 138). Was consideration given to using an instrument such as AGREE II to evaluate the quality of guidelines? While appreciating the scope of this project I think more attention to these issues will improve the clarity and readability of this review.</p> <p>Some further comments:<br/>Abstract, line 14: The use of the term 'treatment choices' is not meaningful given the variety of data sources used.<br/>Abstract, search, line 19: It is now more than 15 months since the search was completed. Also was consideration given to limiting the</p> |
|-------------------------|--|

|  |  |
|--|--|
|  | <p>search to a date such that the data reflected current practice? Is it meaningful to conclude that physical therapists from a study that may have been completed many years ago do not follow current recommended treatment guidelines?</p> <p>Abstract, participants, line 21: 'other methods' should be explained or include examples.</p> <p>Abstract results, line 32: The summary results appear to combine percentage of physical therapists (54%) and percentage of treatments (63%). Since one value is based on an audit of treatments and the other on a survey of what physical therapists would do, it seems to be combining what actually happened with projections, with some potential confusion over whether the unit of measure/interest is the physical therapist or the treatment received by a patient.</p> <p>Introduction, line 87: The term 'treatment choices' may need to be further defined for clarity (see comments above).</p> <p>Methods, about line 94: Please also report other deviations to the registered protocol. E.g. using a modified version of Downs and Black checklist.</p> <p>Methods, line 99: The search strategy did not combine terms related to practice patterns'. The first concept searched terms related to the concept of 'low value care' which is very different. This should be acknowledged as a potential limitation.</p> <p>Methods, line 122: Please specify what proportions were extracted. Was it the proportion of physical therapists surveyed in that study? or the proportion of treatment sessions audited?</p> <p>Methods, line 128: Modified Downs and Black. Why was it a measure of quality that the aims of the study were clear – providing the study provided data relevant to the review question why was this relevant? Why evaluate accuracy of outcome measures, when the review is evaluating practice patterns not the effect of an intervention?</p> <p>Methods, line 175: The method of data synthesis is not clear. It is not certain what is meant by, taking 'an average of therapists' responses across vignettes of equal sample sizes'...</p> <p>Methods, line 190: It is not clear what the medians refer to, median of percentages of studies addressing a particular treatment? If so is this the best way to synthesise the data?...</p> <p>Methods, line 224: It needs to be clarified what the percentages referred to. For example if 1 study surveyed 10 PTs, another study surveyed 100 PTs and a third study surveyed 1 PT on behalf of a department of 10 PTs, how were these data handled? Were studies that surveyed larger numbers of PTs weighted?</p> <p>Results, line 265: Does 'people' refer to physical therapists or patients?</p> <p>Discussion, strengths and weaknesses, about line 328: To what extent do the results reflect clinical practice since older studies were included, surveys may have been subject to an expectation bias, and (as discussed by the authors) omissions in audits may have reflected reporting rather than practice? Related to this the generalisability of the results need to be discussed.</p> <p>Discussion, strengths and weaknesses, line 340: it needs to be acknowledged that the Care Track studies reported primary data collected with a common method in contrast to the systematic review.</p> <p>Discussion, meaning of the study: Could their authors discuss their findings against recommendations that evidence based practice should be based on patient values and clinical expertise as well as the best available evidence (at the time)?</p> |
|--|--|

|  |  |
|--|--|
|  | Tables e.g. Table 2: Does the column N refer to the number of studies, the number of physical therapists surveyed or the number of treatments audited? |
|--|--|

## VERSION 1 – AUTHOR RESPONSE

### **RESPONSE TO REVIEWERS' COMMENTS**

bmjopen-2019-032329: Do physical therapists follow evidence-based guidelines when managing musculoskeletal conditions? A systematic review

#### **REVIEWER #1**

##### **COMMENTS TO THE AUTHORS**

This is a very good paper. Perhaps the only suggestion I would have is to add a discussion point regarding how the quality of the studies that were pulled in this SR could impact the results. The D&B was scored, but maybe something more about how this could influence the results. The only other minor comment is the number of tables, but I think this enhances the results and as long as journal is okay with the number then good to go.

##### **AUTHORS' RESPONSE**

We thank the reviewer for their positive comments. We have now included a statement regarding study quality in the 'strengths and weaknesses of the study' section.

(Page 16, 1<sup>st</sup> paragraph)

Finally, most studies did not use an accurate assessment of treatment choices (n=55/94). However, we stratified our analysis by how treatment choices were assessed so the influence of having an accurate method of assessment is clear to readers.

#### **REVIEWER #2**

##### **COMMENTS TO THE AUTHORS**

The concept of this systematic review is excellent, the scope is ambitious and the manuscript represents an enormous amount of work. However, the review as presented is difficult to follow. One difficulty is that the unit of measure is not always clear, whether percentages are percentage of physical therapists, percentage of patients, or percentage of treatments.

##### **AUTHORS' RESPONSE**

The unit of measurement was dependent on how treatment choices were assessed in the original studies. Treatment choices assessed by surveys completed by physical therapists yielded data on the percentage of physical therapists that provide (survey without vignette) or would provide (survey with vignette) a particular treatment. Treatment choices assessed by audits of clinical notes, audits of

billing codes, treatment recording forms, clinical observation, or surveys completed by patients yielded data on the percentage of patients that received a particular treatment. We stratified all our analyses by these two approaches to measurement, and this has been outlined in the methods.

(Page 9, 2<sup>nd</sup> paragraph)

#### **2.4.1. Assessments of treatment choices**

Data on physical therapy treatment choices were divided into two main categories (and analysed separately) due to differences in how each category is interpreted:

#### **2.4.2 Treatment choices assessed by surveys completed by physical therapists (with or without vignettes)**

*Interpretation.* Surveys completed by physical therapists' yielded data on the percentage of physical therapists that provide (survey without vignette) or would provide (survey with vignette) a particular treatment for a condition they frequently treat.

...

#### **2.4.3. Treatment choices assessed by audits of clinical notes, audits of billing codes, treatment recording forms, clinical observation, or surveys completed by patients**

*Interpretation.* These assessment measures yielded data on the percentage of patients that received a particular physical therapy-delivered treatment in a single treatment session or throughout an episode of care (i.e. from initial consultation to discharge).

...

### **COMMENTS TO THE AUTHORS'**

Another difficulty is that the methods of data synthesis are not clear to the reader, whether counts were summed across studies at the patient/therapist/treatment level and presented as an overall percentage, or whether the percentage from the studies of a the specific treatment were expressed as a median percentage, or perhaps some other method like a meta-analysis of proportions?

### **AUTHORS' RESPONSE**

We have provided more detailed responses to the same concerns regarding our method of analysis in several places below. In summary, we used medians and interquartile ranges to provide an overall percentage for the proportion of physical therapy treatment choices that involved treatments that were recommended, not-recommended and had no recommendation across studies (see section 2.5). We did not perform a meta-analysis. Data were stratified by how treatment choices were assessed (surveys vs. audits of clinical notes). This means counts were summarised across studies at the therapist level when treatment choices were assessed by surveys, and at the patient level when assessed by audits of clinical notes (see section 2.4.1, 2.4.2, 2.4.3 and the response to comment directly above). Our main analysis summarised treatment choices by condition (see section 2.5.1); our secondary analysis summarised treatment choices by specific treatments provided for each condition (see section 2.5.2).

### **COMMENTS TO THE AUTHORS'**

A further difficulty is whether or not a treatment was judged to be accepted was based on whether it was 'recent (line 146)' or 'well-recognised' (line 138). Was consideration given to using an instrument such as AGREE II to evaluate the quality of guidelines? While appreciating the scope of this project I think more attention to these issues will improve the clarity and readability of this review.

### **AUTHORS' RESPONSE**

Given the large number of conditions investigated in this review, for pragmatic purposes we decided against using an instrument like AGREE II to identify high-quality clinical practice guidelines. We instead used guidelines that were well-recognised to the authorship team and the experts that we contacted (see Acknowledgements); all of which have published extensively in the field of musculoskeletal disorders. We believe that we selected the most recent and widely recognised guidelines for each condition. However, if the reviewer feels a particular guideline is missing from this review, we would be happy to consider including it.

### **COMMENTS TO THE AUTHORS'**

Abstract, line 14: The use of the term 'treatment choices' is not meaningful given the variety of data sources used.

### **AUTHORS' RESPONSE**

We believe 'treatment choices' is the most appropriate term given the data we collected. 'Treatment practices' is another phrase we considered, but we believe it is less specific to the individual treatment's physical therapists provide. We are happy to take suggestions on a more appropriate phrase from the reviewer if they have any.

### **COMMENTS TO THE AUTHORS'**

Abstract, search, line 19: It is now more than 15 months since the search was completed.

### **AUTHORS' RESPONSE**

Please see response to an earlier comment from the Editor. Here is the response:

We realise that we conducted the search over 12 months ago but do not feel this makes our search out of date. The recently updated AMSTAR checklist (A MeaSurement Tool to Assess systematic Reviews; see BMJ 2017;358:j4008 <http://dx.doi.org/10.1136/bmj.j4008>) considers searches conducted within 24 months of completion of a review up to date. Further to this, while there might be

a strong argument for updating a search of a systematic review of intervention studies (e.g. because early trials often have overly optimistic results), we feel this is less relevant for our review, particularly since we included 94 studies. Nevertheless, if the Editor feels strongly about having the search strategy as up to date as possible, we are happy to do so.

## **COMMENTS TO THE AUTHORS**

Also was consideration given to limiting the search to a date such that the data reflected current practice? Is it meaningful to conclude that physical therapists from a study that may have been completed many years ago do not follow current recommended treatment guidelines?

## **AUTHORS' RESPONSE**

Further to the argument regarding a 17-year time delay for the translation of research into practice (in response to an earlier comment), we did not limit the search date because we wanted to capture all available data on treatment choices. We have addressed a similar comment from the reviewer below. Here is the response:

We acknowledge that physical therapists' treatment choices may have changed over time so including older studies could limit the relevance of our findings. However, we do not believe this is an important limitation because many guideline recommendations have remained largely consistent overtime. For example, although some studies on treatment choices for low back pain are from 1994, a comparison of low back pain guidelines between 1994 and 2000 found a high degree of consistency of recommendations, such as advice to stay active and avoid bed rest (see Koes BW, et al. (2001). "Clinical guidelines for the management of low back pain in primary care: an international comparison." *Spine (Phila Pa 1976)* 26(22): 2504-2513). This is consistent with current low back pain guidelines. A similar argument could be made for knee pain and neck pain guidelines. Further, studies investigating physical therapists' treatment choices for conditions where guideline recommendations and systematic reviews have not been available until more recently were generally published more recently (see studies on the following conditions in Supplementary Table 4: acute ankle sprains, plantar fasciitis, orthopaedic conditions). We have included this point in the discussion.

(Page 16, 1<sup>st</sup> paragraph)

Third, physical therapists' treatment choices may have changed over time so including older studies could limit the relevance of our findings. Nevertheless, we do not believe this is an important limitation because many guideline recommendations have remained largely consistent overtime. For example, although some studies on treatment choices for low back pain are from 1994, a comparison of low back pain guidelines between 1994 and 2000 found a high degree of consistency of recommendations, such as advice to stay active and avoid bed rest (106). This is consistent with current low back pain guidelines.

## COMMENTS TO THE AUTHORS

Abstract, participants, line 21: 'other methods' should be explained or include examples.

## AUTHORS' RESPONSE

We have now included examples of 'other methods' in the abstract.

(Abstract)

Studies that quantified physical therapy treatment choices for musculoskeletal conditions through surveys of physical therapists, audits of clinical notes, and other methods (e.g. audits of billing codes, clinical observation) were eligible for inclusion.

## COMMENTS TO THE AUTHORS

Abstract results, line32: The summary results appear to combine percentage of physical therapists (54%) and percentage of treatments (63%). Since one value is based on an audit of treatments and the other on a survey of what physical therapists would do, it seems to be combining what actually happened with projections, with some potential confusion over whether the unit of measure/interest is the physical therapist or the treatment received by a patient.

## AUTHORS' RESPONSE

We agree that it is inappropriate to combine data on treatment choices assessed by surveys completed by physical therapists with data on treatment choices assessed by audits of clinical practice. We have not combined data in this way. Instead, we have stratified all our analyses based on these two units of measurement. This has been outlined in the methods and in response to an earlier comment.

(Page 9, 2<sup>nd</sup> paragraph)

### **2.4.2. Assessments of treatment choices**

Data on physical therapy treatment choices were divided into two main categories (and analysed separately) due to differences in how each category is interpreted:

### **2.4.3 Treatment choices assessed by surveys completed by physical therapists (with or without vignettes)**

*Interpretation.* Surveys completed by physical therapists' yielded data on the percentage of physical therapists that provide (survey without vignette) or would provide (survey with vignette) a particular treatment for a condition they frequently treat.

...

### **2.4.3. Treatment choices assessed by audits of clinical notes, audits of billing codes, treatment recording forms, clinical observation, or surveys completed by patients**

*Interpretation.* These assessment measures yielded data on the percentage of patients that received a particular physical therapy-delivered treatment in a single treatment session or throughout an episode of care (i.e. from initial consultation to discharge).

...

#### **COMMENTS TO THE AUTHORS**

Introduction, line 87: The term 'treatment choices' may need to be further defined for clarity (see comments above).

#### **AUTHORS' RESPONSE**

Please see response to an earlier comment regarding the use of the phrase 'treatment choices'.

#### **COMMENTS TO THE AUTHORS**

Methods, about line 94: Please also report other deviations to the registered protocol. E.g. using a modified version of Downs and Black checklist.

#### **AUTHORS' RESPONSE**

We have included other deviations to the registered protocol in the methods.

(Page 6, 2<sup>nd</sup> paragraph)

Other deviations to our registered protocol include using a modified version of the 'Downs and Black' checklist to rate study quality and changing the focus from 'high- and low-value care' to 'recommended and not-recommended care'.

#### **COMMENTS TO THE AUTHORS**

Methods, line 99: The search strategy did not combine terms related to practice patterns'. The first concept searched terms related to the concept of 'low value care' which is very different. This should be acknowledged as a potential limitation.



## **AUTHORS' RESPONSE**

We disagree. Terms related to 'practice patterns' included 'adherence to guidelines', 'guideline use', 'recommended care', etc. We acknowledge that we included terms related to 'low-value care'. These terms were included to capture studies relevant to the other review questions specified in our registered protocol and because we initially focused this review on 'high- and low-value care'. This has been made clear in the methods.

(Page 6, 2<sup>nd</sup> paragraph)

Due to the size of the review, other research questions in our registered protocol (including physical therapy treatment choices for cardiorespiratory and neurological conditions) will be addressed in separate manuscripts. Other deviations to our registered protocol include using a modified version of the 'Downs and Black' checklist to rate study quality and changing the focus from 'high- and low-value care' to 'recommended and not-recommended care'.

## **COMMENTS TO THE AUTHORS**

Methods, line 122: Please specify what proportions were extracted. Was it the proportion of physical therapists surveyed in that study? or the proportion of treatment sessions audited?

## **AUTHORS' RESPONSE**

The interpretation of the proportions that quantified physical therapists' treatment choices is outlined in sections 2.4 and 2.5 and has also been clarified in response to earlier comments. We have made this clear earlier in the methods.

(Page 7, 2<sup>nd</sup> paragraph)

One reviewer (JZ) independently extracted individual study characteristics (e.g. condition, country, participant demographics) and proportions that quantified physical therapy treatment choices (see sections 2.4 and 2.5).

## **COMMENTS TO THE AUTHORS**

Methods, line 128: Modified Downs and Black. Why was it a measure of quality that the aims of the study were clear – providing the study provided data relevant to the review question why was this relevant? Why evaluate accuracy of outcome measures, when the review is evaluating practice patterns not the effect of an intervention?

## **AUTHORS' RESPONSE**

A clear hypothesis/aim/objective is essential for all research studies and the item regarding whether the hypothesis/aim/objective of the study was clear appears in the original 'Downs and Black' checklist. We only removed items that were not relevant to our study design. Assessing accuracy of outcome measures was important because different studies used different methods to assess treatment choices and this could influence the interpretation of our findings. We considered the following assessments of treatment choices as 'accurate' as they likely correlate best with what happens in clinical practice: observation, audits of clinical notes, audits of billing codes, treatment recording forms and validated surveys. This has been outlined in the methods.

(Page 7, 3<sup>rd</sup> paragraph)

The methodological quality of included studies was assessed independently by two reviewers (JZ and MO) using a modified version of the 'Downs and Black' checklist. Any disagreements between the two reviewers were resolved through discussion. We modified the original 27-item 'Downs and Black' checklist (10) and selected eight items that were relevant to studies on treatment choices (Supplementary Table 2). For item eight, we considered the following assessments of treatment choices as 'accurate': observation, audits of clinical notes, audits of billing codes, treatment recording forms and validated surveys.

We also mentioned the proportion of studies that used an accurate assessment of treatment choices in the results and discussion.

(Page 13, 2<sup>nd</sup> paragraph)

The most common methodological limitations included failing to report that physical therapists who were prepared to participate were representative of the population from which they were drawn (n=88/94) and not using an accurate assessment of treatment choices (n=55/94).

(Page 16, 1<sup>st</sup> paragraph)

Finally, most studies did not use an accurate assessment of treatment choices (n=55/94). However, we stratified our analysis by how treatment choices were assessed so the influence of having an accurate method of assessment is clear to readers.

## **COMMENTS TO THE AUTHORS**

Methods, line 175: The method of data synthesis is not clear. It is not certain what is meant by, taking 'an average of therapists' responses across vignettes of equal sample sizes'...

## **AUTHORS' RESPONSE**

If physical therapists in a particular study had to outline their treatment choices for three vignettes (of the same condition), and the same number of physical therapists provided a response for each

vignette, we averaged the proportion of physical therapists that provided certain treatments across the vignettes. For example, consider three vignettes on acute low back pain where the proportion of physical therapists that provided advice to stay active was 30%, 40% and 50% for vignettes 1, 2, and 3, respectively. Assuming treatment choices were reported by the same number of physical therapists across these vignettes, we took the average of their responses (i.e. 40% of physical therapists provide advice to stay active for acute low back pain). If the number of physical therapists that completed each vignette differed, we only used data from the vignette that had the highest number of respondents. We are happy to include this explanation in the text if the Editor thinks it is necessary to do so.

## **COMMENTS TO THE AUTHORS**

Methods, line 190: It is not clear what the medians refer to, median of percentages of studies addressing a particular treatment? If so is this the best way to synthesise the data?...

## **AUTHORS' RESPONSE**

Medians were used to summarise the percentage of physical therapy treatment choices across studies that involved treatments that were recommended, not-recommended and had no recommendation. This has now been clarified in the methods.

(Page 10, 2<sup>nd</sup> paragraph)

### **2.5. Analysis**

We used counts and ranges to summarise study characteristics for each condition. We used medians and interquartile ranges (IQR) to summarise the percentage of physical therapy treatment choices that involved treatments that were recommended, not-recommended and had no recommendation across studies.

Justification for using medians to combine data is outlined in response to the next comment.

## **COMMENTS TO THE AUTHORS**

Methods, line 224: It needs to be clarified what the percentages referred to. For example if 1 study surveyed 10 PTs, another study surveyed 100 PTs and a third study surveyed 1 PT on behalf of a department of 10 PTs, how were these data handled? Were studies that surveyed larger numbers of PTs weighted?

## AUTHORS' RESPONSE

The interpretation of the percentage values is described in 2.4 and has been outlined in response to an earlier comment. See here:

(Page 9, 2<sup>nd</sup> paragraph)

### **2.4.3. Assessments of treatment choices**

Data on physical therapy treatment choices were divided into two main categories (and analysed separately) due to differences in how each category is interpreted:

#### **2.4.4 Treatment choices assessed by surveys completed by physical therapists (with or without vignettes)**

*Interpretation.* Surveys completed by physical therapists' yielded data on the percentage of physical therapists that provide (survey without vignette) or would provide (survey with vignette) a particular treatment for a condition they frequently treat.

...

#### **2.4.3. Treatment choices assessed by audits of clinical notes, audits of billing codes, treatment recording forms, clinical observation, or surveys completed by patients**

*Interpretation.* These assessment measures yielded data on the percentage of patients that received a particular physical therapy-delivered treatment in a single treatment session or throughout an episode of care (i.e. from initial consultation to discharge).

...

We did not weight studies by their sample size as we used medians and interquartile ranges (IQR) to summarise the percentage of physical therapy treatment choices that involved treatments that were recommended, not-recommended and had no recommendation across studies.

(Page 10, 2<sup>nd</sup> paragraph)

## **2.5. Analysis**

We used counts and ranges to summarise study characteristics for each condition. We used medians and interquartile ranges (IQR) to summarise the percentage of physical therapy treatment choices that involved treatments that were recommended, not-recommended and had no recommendation across studies.

## COMMENTS TO THE AUTHORS

Results, line 265: Does 'people' refer to physical therapists or patients?

## **AUTHORS' RESPONSE**

Physical therapists. This has now been revised.

(Page 13, 3<sup>rd</sup> paragraph)

The most common methodological limitations included failing to report that physical therapists who were prepared to participate were representative of the population from which they were drawn (n=88/94) and not using an accurate assessment of treatment choices (n=55/94).

## **COMMENTS TO THE AUTHORS**

Discussion, strength and weaknesses, about line 328: To what extent do the results reflect clinical practice since older studies were included, surveys may have been subject to an expectation bias, and (as discussed by the authors) omissions in audits may have reflected reporting rather than practice? Related to this the generalisability of the results need to be discussed.

## **AUTHORS' RESPONSE**

We acknowledge that physical therapists' treatment choices may have changed over time so including older studies could limit the relevance of our findings. However, we do not believe this is an important limitation because many guideline recommendations have remained largely consistent overtime. For example, although some studies on treatment choices for low back pain are from 1994, a comparison of low back pain guidelines between 1994 and 2000 found a high degree of consistency of recommendations, such as advice to stay active and avoid bed rest (see Koes BW, et al. (2001). "Clinical guidelines for the management of low back pain in primary care: an international comparison." *Spine (Phila Pa 1976)* 26(22): 2504-2513). This is consistent with current low back pain guidelines. A similar argument could be made for knee pain and neck pain guidelines. Further, studies investigating physical therapists' treatment choices for conditions where guideline recommendations and systematic reviews have not been available until more recently, were generally published more recently (see studies on the following conditions in Supplementary Table 4: acute ankle sprains, plantar fasciitis, orthopaedic conditions). We have included this point in the discussion.

(Page 16, 1<sup>st</sup> paragraph)

Third, physical therapists' treatment choices may have changed over time so including older studies could limit the relevance of our findings. Nevertheless, we do not believe this is an important limitation because many guideline recommendations have remained largely consistent overtime. For example, although some studies on treatment choices for low back pain are from 1994, a comparison of low back pain guidelines between 1994 and 2000 found a high degree of consistency of recommendations, such as advice to stay active and avoid bed rest (106). This is consistent with current low back pain guidelines.

We also reported that the use of surveys completed by physical therapists (in many included studies) is a limitation of this review as they might not be an accurate measure of treatment choices (due to expectation bias for example).

(Methods)

The methodological quality of included studies was assessed independently by two reviewers (JZ and MO) using a modified version of the 'Downs and Black' checklist. Any disagreements between the two reviewers were resolved through discussion. We modified the original 27-item 'Downs and Black' checklist (10) and selected eight items that were relevant to studies on treatment choices (Supplementary Table 2). For item eight, we considered the following assessments of treatment choices as 'accurate': observation, audits of clinical notes, audits of billing codes, treatment recording forms and validated surveys.

(Page 16, 1<sup>st</sup> paragraph)

Finally, most studies did not use an accurate assessment of treatment choices (n=55/94). However, we stratified our analysis by how treatment choices were assessed so the influence of having an accurate method of assessment is clear to readers.

## **COMMENTS TO THE AUTHORS**

Discussion, strengths and weaknesses, line 340: it needs to be acknowledged that the Care Track studies reported primary data collected with a common method in contrast to the systematic review.

## **AUTHORS' RESPONSE**

We have now outlined that the Care Track studies reported primary data collected and were not systematic reviews.

(Page 17, 1<sup>st</sup> paragraph)

A difference to our study is that the CareTrack studies used consensus of experts to judge the value of care; whereas we based this decision upon evidence-based practice guidelines and systematic reviews. Another difference is that the CareTrack studies only assessed healthcare decisions through audits of clinical notes; we used audit of clinical notes, surveys, vignettes, and clinical observation. Further, the Care Track studies reported primary data collected and were not systematic reviews.

## **COMMENTS TO THE AUTHORS**

Discussion, meaning of the study: Could the authors discuss their findings against recommendations that evidence based practice should be based on patient values and clinical expertise as well as the best available evidence (at the time)?

### **AUTHORS' RESPONSE**

We agree that evidence-based practice is based on high quality evidence, patient values and clinical expertise, and that the latter two sources of information help clinicians and patients to select the best option amongst a range of evidence-based options. This approach is entirely consistent with our analyses.

### **COMMENTS TO THE AUTHORS**

Tables e.g. Table 2: Does the column N refer to the number of studies, the number of physical therapists surveyed or the number of treatments audited?

### **AUTHORS' RESPONSE**

'N' refers to the number of studies as described in the footnotes of Table 2.

### **VERSION 2 – REVIEW**

|                        |                                  |
|------------------------|----------------------------------|
| <b>REVIEWER</b>        | Derek Clewley<br>Duke University |
| <b>REVIEW RETURNED</b> | 07-Aug-2019                      |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | This is a well-designed and well-done systematic review that is of high value, especially as more research begins to focus on implementation. This review will serve as a resource for those designing implementation projects. I recommend accept. |
|-------------------------|---|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Nicholas Taylor<br>La Trobe University, Australia |
| <b>REVIEW RETURNED</b> | 27-Aug-2019                                       |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | The manuscript has been improved and the explanations in the author responses are helpful. The authors have clarified that data on physical therapy treatment choices were divided into two main categories due to differences in how each category is interpreted (line 158). For this reason I think the results should be clearly reported separately in the Abstract as they are in the first lines of sections 3.2.1 and 3.2.2: 'The median percentage of physical therapists that provide treatments... and the median percentage of patients that received physical-therapy-delivered treatments...' |
|-------------------------|---|

|  |  |
|--|--|
|  | <p>For example, the second sentence of Abstract, results (line 30) should be replaced with: For musculoskeletal conditions, the median percentage of physical therapists who chose recommended treatments was 54% across 23 studies, and the median percentage of patients that received recommended physical-therapy delivered treatments was 63% across 8 studies. Similarly the first paragraph of the Discussion (lines 308-309) should be rewritten for clarity for the reader.</p> |
|--|--|

## VERSION 2 – AUTHOR RESPONSE

REVIEWER #1

COMMENTS TO THE AUTHORS

This is a well-designed and well-done systematic review that is of high value, especially as more research begins to focus on implementation. This review will serve as a resource for those designing implementation projects. I recommend accept.

AUTHORS' RESPONSE

We thank the reviewer for their positive comments.

REVIEWER #2

COMMENTS TO THE AUTHORS

The manuscript has been improved and the explanations in the author responses are helpful. The authors have clarified that data on physical therapy treatment choices were divided into two main categories due to differences in how each category is interpreted (line 158). For this reason I think the results should be clearly reported separately in the Abstract as they are in the first lines of sections 3.2.1 and 3.2.2: 'The median percentage of physical therapists that provide treatments... and the median percentage of patients that received physical-therapy-delivered treatments...'. For example, the second sentence of Abstract, results (line 30) should be replaced with: For musculoskeletal conditions, the median percentage of physical therapists who chose recommended treatments was 54% across 23 studies, and the median percentage of patients that received recommended physical-therapy delivered treatments was 63% across 8 studies. Similarly the first paragraph of the Discussion (lines 308-309) should be rewritten for clarity for the reader.

AUTHORS' RESPONSE

We thank the reviewer for their positive comments. We have revised the abstract and discussion accordingly.

(Abstract)



For musculoskeletal conditions, the median percentage of physical therapists who chose recommended treatments was 54% (n=23 studies; surveys completed by physical therapists) and the median percentage of patients that received recommended physical therapy-delivered treatments was 63% (n=8 studies; audits of clinical notes). For treatments not-recommended, these percentages were 43% (n=37; surveys) and 27% (n=20; audits). For treatments with no recommendation, these percentages were 81% (n=37; surveys) and 45% (n=31; audits).

(Page 15, 2nd paragraph)

Across all musculoskeletal conditions, 54% of physical therapists chose recommended treatments, 43% chose treatments that were not recommended and 81% chose treatments that have no recommendation (based on surveys completed by physical therapists). Based on audits of clinical notes, 63% of patients received recommended physical therapy-delivered treatments, 27% received treatments that were not recommended and 45% received treatments that have no recommendation.