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) | Behaviour-change interventions for the management of Raynaud’s Phenomenon: a systematic literature review |
| AUTHOR(S) | Daniels, Jo; Pauling, John; Eccleston, Christopher |

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Matthieu Roustit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ. Grenoble Alpes</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>01-Jul-2018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENERAL COMMENTS</th>
<th>Matthieu Roustit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ. Grenoble Alpes</td>
</tr>
<tr>
<td></td>
<td>01-Jul-2018</td>
</tr>
</tbody>
</table>

This is a well-conducted systematic review of behavior-change interventions in patients with Raynaud’s. Although most studies are old, and data are too scarce to conclude about the efficacy of behavioral interventions, the authors are to be congratulated for this work; it raises a number of issues and provides recommendations for the conduction of future trials. The methods are standard and have been published in BMJ Open and PROSPERO. Outcomes are relevant. I only have minor remarks.

- Was there any minimal pre-specified duration for intervention? This should be clearly stated in the methods (e.g. page 6, line 1-2)

- Table 1 could be improved. Study design might be included (e.g. crossover vs parallel). The word “period” (column 6) is commonly used for crossover studies, and suggests that all studies were crossover, which is not the case (RTS). Please add table legend with all abbreviations (BFB, EMG…)

- Overall, I found that interventions were not detailed enough. A description of “biofeedback”, and more details about underlying psychological models is needed.

- The issue of blinding probably has major influence on the results. Alternate designs such as randomized consent designs (i.e. adapted from Zelen's method) have been successfully used to assess behavioral interventions, and they might be relevant in this situation. This could be discussed or even suggested in Table 4.

- Page 13, line 25-26: “Third, small trials…” This is largely debated (how small is small?) and in the field of RP, especially when secondary to SSc, this would exclude most trials.
- Page 13, line 32; and Table 4, line 3. If there is a published consensus about the use of RCS, please add the reference. My feeling is that RCS has been largely debated recently. Moreover, to the best of my knowledge, RCS has never been properly validated in primary RP. If I am wrong, please add the appropriate references.

- Table 4: “Imaging modalities...” I am not certain that any available imaging technique can be considered as a reliable surrogate in clinical trials. Again, this should be supported by adequate references.

- Table 4: I agree that temperature, being the strongest predictor of RP attacks, should be measured and reported in all clinical trials. I would go further and recommend that it is systematically used as a covariate in statistical models.

- Page 13, line 22: what does “this technology” refer to?

- Table 3: the legend is missing (* ?). There are two dates for the “Melin & Fagerstrom” ref.

- Table 4: please add numbered ref for Bartholomew 2011. I could not find it in the list of references

- P4, line 16: “4digital” please correct the typo
Page 6, line 19.
The authors assessed the risk of bias according to the Cochrane handbook. They provide a Table with the evaluations across each domain. Moreover, they also provide an overall risk of bias for each study (Table 1). It is not clear to me how the overall score was obtained. Which were the parameters used to conclude that a study is on average ‘at high’ or ‘at low’ risk of bias? The worst score across the domains? If this was the case, I identify some discrepancies. For example, only RTS study was considered to be at low risk of bias. Checking in the Table 2 I see that four of seven items (random sequence allocation, allocation concealment, blinding of assessors and incomplete outcome reports) are at high risk of bias for this study.
Please explain how the overall risk of bias was built for each study in ‘Risk of bias’ section and justify why the study from RTS group only was considered at low risk in Result section.

Page 7, line 7 Study characteristics
Page 7, line 10. ‘All studies randomized participants to an active control’. However in study from Melin et al behavioural treatment was compared to placebo. Please clarify.
Page 7, line 12. ‘All studies indicated a non-active comparator…’ Again, this statement is not clear to me: a) the authors list only 6 studies with a potential non-active comparator; b) what does ‘intervention to isolate the effects of the active treatment’ mean? Are these studies versus no intervention?
I think this paragraph is hard to follow and should be better reorganized. For example it would be useful to highlight in the table the comparators and categorize them in active, non-active, no treatment, placebo…

Page 8. Table 1.
Guglielmi instead of Gugliemli (check this name also across the text).
‘N of patients’ column. Please clarify the meaning of 39 (36) in Guglielmi and 32 (30) in Surwit rows.
‘Outcome investigated’ column. Please put a short description in the legend of the terms ‘Antecedents’ and ‘Description’
‘Any effect (BC)’ column. ‘+’, positive effect’. I understand that the trial is positive, i.e. the treatment under investigation gives a statistically significant advantage, the primary point is met. Is it correct? Maybe it would be useful to better explain this in the legend. The same applies for < >.
Moreover, provide a legend with the meanings of all abbreviations.

Page 10, line 4-28 Interventions
Split the paragraph in 2 sections where it can be easily identified the description of the interventions and then that of the comparators (line 17-21 of the same page).
Page 10, line 46 Risk of bias
As above, please clarify why RTS group study was considered to be at low risk of bias.

Page 4, line 16. Delete ‘4’
6. Was there any minimal pre-specified duration for intervention? This should be clearly stated in the methods. For the purposes of including all relevant studies, there was no minimal pre-specified duration for interventions. We have stated this in the methods section under 'inclusion criteria', (pg 5, para 3, final sentence):
“Duration of intervention was not subject to a minimum criteria, providing the study within which the intervention was assessed fully met all inclusion criteria”.

7. Table 1 could be improved. Study design might be included (e.g. crossover vs parallel). The word “period” (column 6) is commonly used for crossover studies, and suggests that all studies were crossover, which is not the case (RTS). Please add table legend with all abbreviations (BFB, EMG…)

Study design has been inserted into table 1, the term ‘period’ has been replaced with ‘duration’. The abbreviations have been expanded to include the full terms, therefore a legend is no longer necessary.

8. Overall, I found that interventions were not detailed enough. A description of “biofeedback”, and more details about underlying psychological models is needed.

We have now expanded to include a more detailed description of biofeedback and underlying psychological models:
“Biological feedback, or ‘biofeedback’ interventions in RP are based on the notion that participants can be trained to voluntarily increase blood flow to the extremities, based on biological temperature feedback. Biofeedback interventions were similar in procedure (notwithstanding differences in duration/frequency). An overview of the procedural approach can be summarised herewith: following laboratory assessment, participants were invited to voluntarily increase their skin temperature unaided. Participants were then trained to increase skin temperature based on individualised biological feedback; through the use of physiological skin monitoring apparatus, changes in skin temperature were indicated through the use of acoustic sounds which increased in volume along with escalation in skin temperature [20, 25, 22, 26]; visual representations of changes in skin temperature such as a digital panel [23, 26, 27] or a moving light [22]. Participants were instructed to increase skin temperature based on this audio/visual feedback. Following completion of training, participants were invited to voluntarily increase finger temperature without the aid of biofeedback. One study did not describe the biofeedback intervention [21]. The ‘behavioural intervention’ [24, 28] reported a classical conditioning intervention directed at “weakening of the unconditioned link between cold and peripheral vasospasms” (p111, [24]). Subjects submerged both hands in water that was either +23°C (placebo group) or + 43 oC (active treatment intervention) and were instructed that the addition of a coloured substance to the water was an active treatment (placebo). Participants were informed that the warmer the water the better the penetration of the proposed ‘active’ (placebo) drug, seeking to weaken the link between cold temperatures and vasospasm in the active group. Expectation of treatment effect and self-monitoring were then measured. One study [25] delivered ‘cognitive stress management’ to 50% of each of the four groups in the participant sample (total n=32). This intervention consisted of reviewing symptoms, precipitants of episodes, cognitions during episodes and strategies employed to manage episodes.”

9. The issue of blinding probably has major influence on the results. Alternate designs such as randomized consent designs (i.e. adapted from Zelen’s method) have been successfully used to assess behavioral interventions, and they might be relevant in this situation. This could be discussed or even suggested in Table 4.

Thank you for your suggestion. We agree that blinding is an issue across the majority of psychological and behaviour change interventions as true blinding is unlikely to be achieved. As all studies included
have a behaviour-change intervention they were arguably all subject to a potential bias, however Zelen’s method would counter-balance some of the effect of this. To this end we have incorporated your suggestion into our table.

10. Page 13, line 25-26: “Third, small trials...” This is largely debated (how small is small?) and in the field of RP, especially when secondary to SSc, this would exclude most trials.

We agree this is lacking in clarity. We have altered the wording as follows (now pg 16 final para):

Third, smaller sample trials even when properly reported threaten precision of the effect estimates and introduce the possibility of unreliability; while smaller samples are not problematic per se, the majority of the studies included also demonstrate threats to reliability such as the absence of robust assessment measures and moderate to high risk of bias.

11. Page 13, line 32; and Table 4, line 3. If there is a published consensus about the use of RCS, please add the reference. My feeling is that RCS has been largely debated recently. Moreover, to the best of my knowledge, RCS has never been properly validated in primary RP. If I am wrong, please add the appropriate references.

We have added the following passage regarding this.

“The RCS diary is currently the preferred outcome measure for scleroderma-related RP clinical trials and has been endorsed by the Scleroderma Clinical Trials Consortium [SCTC] (Khanna et al. ARD 2008) and OMERACT (Merkel PA et al. 2003). Whilst not formally validated for primary RP, it has been successfully used in studies of mixed populations of patients (Chung et al. 2009). Recent work has highlighted limitations to the RCS diary and work is underway to develop novel patient-reported outcome instruments that might complement diary-based approaches in future clinical trials assessing the outcome of behaviour intervention on the severity and impact of RP (Pauling et al. Arthritis & Rheumatology 2018, Pauling et al. JSRD 2018, Baron et al. JSRD 2018)

Hence in Table 4 we have proposed:

“The RCS diary (or any future validated tools for assessing RP) should be employed as a standardised tool of choice in RP trials to allow for meaningful comparisons across treatment conditions and studies and gather relevant outcome data in one measure”

12. Table 4: “Imaging modalities...” I am not certain that any available imaging technique can be considered as a reliable surrogate in clinical trials. Again, this should be supported by adequate references.

We have modified this section in Table 4 and added references as requested. It now reads:

“Functional assessment of digital microvascular function (e.g. laser-derived imaging modalities or thermal imaging to assess digital vascular function) should be continued to be used as exploratory endpoints to triangulate with subjective measures, although further validation of non-invasive microvascular imaging techniques is necessary before they can be fully incorporated into the endpoint model of RP clinical trials”.

13. Table 4: I agree that temperature, being the strongest predictor of RP attacks, should be measured and reported in all clinical trials. I would go further and recommend that it is systematically used as a covariate in statistical models.

Table 4 has been altered to include your recommendation that it is used as a covariate.

14. Page 13, line 22: what does “this technology” refer to?

This sentence has been clarified to replace ‘technology’ with ‘behaviour change intervention’ as follows:
“There were four features of trial design and reporting which hampered any analysis of efficacy and safety of behaviour-change interventions in this review.” (pg 16, first sentence, final paragraph)

15. Table 3: the legend is missing (* ?). There are two dates for the “Melin & Fagerstrom” ref.

A legend has been inserted for ‘*’. The two dates refer to the original and follow up paper of the Melin & Fagerstrom paper. We have now made this more explicit with further detail inserted into the final paragraph of the ‘study characteristics’ section on page 7, and updated the tables to more clearly indicate this is a follow up paper:

“One study reported a 12 month follow-up of their original sample (1981)[24] in a much later, separate paper published in English (1996)[28].”

16. Table 4: please add numbered ref for Bartholomew 2011. I could not find it in the list of references

This has now been inserted in the reference list.

17. P4, line 16: “4digital” please correct the typo

This has now been corrected.

REVIEWER 2

No comments to address.

REVIEWER 3

18. Page 6, line 19. The authors assessed the risk of bias according to the Cochrane handbook. They provide a Table with the evaluations across each domain. Moreover, they also provide an overall risk of bias for each study (Table 1). It is not clear to me how the overall score was obtained. Which were the parameters used to conclude that a study is on average ‘at high’ or ‘at low’ risk of bias? The worst score across the domains? If this was the case, I identify some discrepancies. For example, only RTS study was considered to be at low risk of bias. Checking in the Table 2 I see that four of seven items (random sequence allocation, allocation concealment, blinding of assessors and incomplete outcome reports) are at high risk of bias for this study.

Please explain how the overall risk of bias was built for each study in ‘Risk of bias’ section and justify why the study from RTS group only was considered at low risk in Result section.

Thank you for drawing our attention to the discrepancy. Having carefully reviewed the RoB table and our ratings, a typographical error has been made, and now corrected. All other domains across all studies have also been checked. The table should now be consistent with the overall risk of assessment. Furthermore, we have expanded the risk of bias section to justify and explain overall risk of bias:

“As highlighted in the risk of bias assessment table (table 2), unclear reporting of random sequence generation, allocation concealment and blinding of participants were common limitations of the studies assessed. Overall risk of bias was assessed based on risk across the seven domains and the following considerations: due to the nature of psychologically based interventions, high risk of
performance bias is common because participants and therapists cannot be blinded adequately; incomplete and selective outcome reporting were deemed important in relation to the stated second study objective, to inform development of future treatment trials, therefore more weight is given to these domains; random sequence generation and allocation of concealment are considered pertinent and particularly important as regards interpreting the comparative efficacy and informing future research; uncertainty in several domains were regarded as susceptibility to high risk of bias and compromising study objectives; as indicated in the Cochrane handbook [18], high risk of bias across most domains is associated with over-estimates of effect, which would compromise the study objective to assess relative efficacy. Based on these particular considerations, the majority of the studies were rated as having a high risk of bias.

The lowest overall risk of bias was in the Raynaud’s Treatment Study group (RTS). This was the largest study and benefited from higher quality reporting typical of larger well controlled RCTs. The Gugliemli study [26] was a smaller, well reported but poorly controlled study in relation to random allocation methods and allocation of concealment; the randomisation process involved matching on age, frequency and duration of attacks and was lacking in transparency. All other studies indicated high risk of bias or uncertainty across the majority of domains and were thus considered overall high risk of bias.

19. Page 7, line 10. ‘All studies randomized participants to an active control’. However in study from Melin et al behavioural treatment was compared to placebo. Please clarify. This was an error, it now reads: ‘All studies randomised participants to an active treatment arm and control arm.’

20. page 7, line 12. ‘All studies indicated a non-active comparator…’ Again, this statement is not clear to me:

a) the authors list only 6 studies with a potential non-active comparator;

The wording has been altered to offer more clarity. The ‘non-active’ has been removed and the following sentences altered accordingly. It now reads:

“All studies randomised participants to an active treatment arm and at least one comparator arm. Three studies described one comparator treatment.[22-24] Five reported two comparator arms [20, 21, 25, 26, 27]. Behaviour change interventions used across the trials included Biofeedback (N=7) and a behavioural intervention (N=1). Three other active treatments formed part of the treatment trials: deep oscillation [21]; autogenic relaxation either alone as an active comparator treatment [20, 25] or combined with biofeedback [23]; and Nifedipine, a calcium channel blocker pharmacological intervention in tablet form (N=1) (see interventions section for detail). Only two studies used a no treatment condition as a control comparator.[21, 26]”

b) what does ‘intervention to isolate the effects of the active treatment’ mean? Are these studies versus no intervention?

This is a poorly phrased description referring to the use of comparator arms which were designed to compensate for placebo or potentially confounding effects of the treatment arm. This has been re-worded along with the changes relating to the previous comment, but for clarity is restated here:

“Five studies used interventions which were designed to compensate for the placebo effect or confounding effects of the active treatment arm: gymnastic hand exercises with similar levels of hand movement to perform biofeedback to perform biofeedback [21]; frontalis Electromyograph (EMG) to counterbalance effects of receiving physiological feedback from biofeedback [20,25,27]; autogenic relaxation where the active treatment arm was autogenic relaxation plus biofeedback [23].”
21. I think this paragraph is hard to follow and should be better reorganized. For example it would be useful to highlight in the table the comparators and categorize them in active, non-active, no treatment, placebo...

We have reorganised the paragraph, as detailed above with the purpose of offering more clarity. We have highlighted the specific arms of the trial in the table, and inserted a legend below the table.

22. Page 8. Table 1. Guglielmi instead of Gugliemli (check this name also across the text). ‘N of patients’ column.

This has been corrected.

23. Please clarify the meaning of 39 (36) in Guglielmi and 32 (30) in Surwit rows. ‘Outcome investigated’ column.

This has been changed to reflect only those who completed treatment and whose outcomes were reported.

24. Please put a short description in the legend of the terms ‘Antecedents’ and ‘Description’

Having reviewed the ‘study outcomes’ section (pg 7, first para) it was deemed appropriate to provide a fuller description of ‘antecedents’ and ‘description’ in this section; the table legends have been enhanced by several in number therefore quite busy – and it sits well in the ‘study outcomes’ section referring to the table. We hope you agree. It now states: “All studies targeting primary RP used diary-based approaches to assessing RP (predecessors of the RCS diary), including measures of severity/impact frequency, duration and severity of episodes, with some studies expanding further to include episodes antecedents (i.e. perceived triggers or events immediately prior to the RP episode), a description of the RP episode including symptoms (see table 1)".

25. ‘Any effect (BC)’ column. ‘+, positive effect’. I understand that the trial is positive, i.e. the treatment under investigation gives a statistically significant advantage, the primary point is met. Is it correct? Maybe it would be useful to better explain this in the legend. The same applies for < >. Moreover, provide a legend with the meanings of all abbreviations.

This is correct; clarification has been made in the detail as regards ‘< >’, ‘+’ and all abbreviations.

Page 10, line 4-28 Interventions

Split the paragraph in 2 sections where it can be easily identified the description of the interventions and then that of the comparators (line 17-21 of the same page).

This has now been separated into behaviour change interventions and controls.


This has now been deleted.

27. FORMATTING AMENDMENTS (if any)

Required amendments will be listed here; please include these changes in your revised version:

- Please embed your DATA SHARING STATEMENT in your main document file as shown in scholar one.

This has been done.
- Please provide another copy of your figures with better qualities and please ensure that Figures are of better quality or not pix-related when zoom in. NOTE: They can be in TIFF or JPG format and make sure that they have a resolution of at least 300 dpi and at least 90mm x 90m of width. Figures in PDF, DOCUMENT, EXCEL and POWER POINT format are not acceptable.

This has been provided and uploaded.

- We have implemented an additional requirement to all articles to include 'Patient and Public Involvement' statement within the main text of your main document. Please refer below for more information regarding this new instruction:

Authors must include a statement in the methods section of the manuscript under the sub-heading 'Patient and Public Involvement'.

This has been inserted.

**VERSION 2 – REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Michele Iudici</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHODS Team, CRESS, Paris, France</td>
<td></td>
</tr>
</tbody>
</table>

| REVIEW RETURNED         | 31-Oct-2018                        |

| GENERAL COMMENTS        | The authors have addressed all reviewer's comments. The paper has been improved. |