## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | The Scleroderma Patient-centered Intervention Network – Scleroderma Support group Leader Education (SPIN-SSLED) Program: Non-randomised Feasibility Trial |
| **AUTHORS** | Thombs, Brett; Dyas, Laura; Pépin, Mia; Aguila, Kylene; Carrier, Marie-Eve; Tao, Lydia; Harb, Sami; Malcarne, Vanessa; El-Baalbaki, Ghassan; Peláez, Sandra; Sauve, Maureen; Hudson, Marie; Platt, Robert; Patient Advisory Team, SPIN-SSLED |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Yvonne Zurynski<br>Australian Institute of Health Innovation<br>Macquarie University<br>Sydney, Australia |
| **REVIEW RETURNED** | 29-Apr-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | An interesting paper addressing an under-researched area.<br>A number of queries need to be dealt with.<br>Although semi-structured interviews were conducted with individual participants there is no explanation as to how the qualitative interview data were analysed. Please explain more clearly what was actuallt collected - PEMAT and qualitative data or just PEMAT? How was the qualitative data analysed?<br><br>Nine of the 10 participants had SSc - what about the 10th? what was their interest?<br>Table 1 - the Race/Ethnicity currently adds up to 110% - pls fix this.<br>Using statements such as " large efffect size" are not helpful - were any of the results significantly different pre- vs post? If significance testing was not undertaken this needs to be explained - why not? I would like to see a clearer and more detailed section on data analysis in the methods.<br>The discussion repeats the results in parts without adding additional information - the discussion should concentrate on the meaning and implication of the results.<br>Given that there were technical difficulties for some participants with hearing imparment, how will these be addressed in the main trial? What accommodations can be made?<br>Although this was a feasibility study, the small sample size and inability to provide data on statistical significance should be acknowledged as limitations or pls provide the relevant statistics. |

| | |
|---|---|
| **REVIEWER** | Ada Man<br>University of Manitoba<br>Canada |

| | |
|---|---|
| **REVIEW RETURNED** | 23-May-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a report on a feasibility trial, testing an educational program for current and future scleroderma support group leaders. Outcomes included feasibility, program quality, participant satisfaction, and participant personal outcomes. This is an important program to develop and will help improve overall scleroderma care.<br><br>The manuscript at times provides too much details, such as on page 12, describing entirely the development of one of the outcome measure, which has already been published (ref 46). Along the same lines, there are redundant points, such as explanation of why French participants were excluded, repeated multiple times in the paper. There are also a lot of details about the future RCT, while this report should be focusing on the current feasibility study. While these are only stylistic issues, it distracts from the focus of the paper and I would recommend the authors run through it a few more times to eliminate any unnecessary paragraphs.<br><br>A second issue I have is with using the SSGLSS as a major outcome measure in this study, and plans to use it as the primary outcome in the future RCT. Table 3 describes all the items in the scale. I would have expected, if the educational program has 13 modules, each addressing one or more learning objective, that the primary outcome of this educational intervention, would be to study whether the participants improved their knowledge in those learning objectives. As an analogy, if the videos were about 3 ways of cooking eggs, the learning assessment should be whether they now feel comfortable cooking eggs in those 3 ways, not so much all the ways in which they feel equipped at being a chef. The SSGLS certainly has items that were addressed by those modules, but some did not appear to be, such as item 1 financial resources. This would lower the score, but what does that mean? It means it probably just was not taught - not that the video educational program was not effective. I would recommend for the authors to consider this issue at least for the RCT.<br><br>I also wondered about the appropriateness of including emotional distress and physical function as outcome measures for the participants. There are many confounders that would affect these outcomes, especially physical function. How do the authors plan to control for these confounders. At face value/biological plausibility, there is not an obvious direct relationship between this type of educational program and overall emotional distress/physical function for a complex individual with many other life/health factors. Table 4 essentially shows minimal differences, yet the authors interpret these as moderate effect sizes (please include a reference to support this -Cohen's d typically refers to 0.5-0.8 as moderate (0.2-0.5 would be small)).<br><br>Since Standardized Mean Difference Effect Sizes appear to be the major outcome measure for the instruments used, please include a description of this and the interpretation of the effect sizes in the methods- data analysis section.<br><br>I am also unclear about including the planned trial outcome measures in this report. The objectives section states that the trial was to 1. evaluate feasibility 2. assess user satisfaction. In the |

| | |
|---|---|
| | discussion, the authors address these two objectives, but does not discuss the results of the planned trial outcome measures. The overall paper may be more focused and coherent if it only focused on the feasibility and user satisfaction. |

## VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

(1) An interesting paper addressing an under-researched area. A number of queries need to be dealt with.

We thank Reviewer #1 for this positive comment and hope that our responses have addressed her queries.

(2) Although semi-structured interviews were conducted with individual participants there is no explanation as to how the qualitative interview data were analysed. Please explain more clearly what was actuallt collected - PEMAT and qualitative data or just PEMAT? How was the qualitative data analysed?

We have clarified (Pages 12, Lines 17-19 to Page 13, Line 1) that “Individual semi-structured interviews were conducted with all participants via telephone upon completion of the 13 modules using items based on the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT) [45] and addressed topics related to usability, understandability, organisation, and clarity of the SPIN-SSLED program, including its videoconference-based delivery.” We have also clarified that all suggestions for improvements from the PEMAT interviews were recorded (Page 16, Lines 5-7), “Qualitative information via interviews and weekly reports by participants was collected, and all suggestions for changes to the program or trial methods that could be implemented prior to beginning a full-scale trial were recorded.” These are reported in Table 2.

(3) Nine of the 10 participants had SSc - what about the 10th? what was their interest?

We do not have information on why this person chose to become involved with a scleroderma advocacy and support organization. It is not uncommon that people without a disease lead support groups, and we have not addressed the reason why this person, specifically, became involved.

(4) Table 1 - the Race/Ethnicity currently adds up to 110% - pls fix this.

As described in the footnotes to Table 1 and as is common practice, participants could select more than one group.

(5) Using statements such as " large efffect size" are not helpful - were any of the results significantly different pre- vs post? If significance testing was not undertaken this needs to be explained - why not? I would like to see a clearer and more detailed section on data analysis in the methods.

Historically, researchers have sometimes proposed that pilot studies be used to estimate effect sizes and calculate sample sizes for full RCTs. However, it is recognized that using insufficiently powered pilot studies for this purpose often produces inaccurate and misleading estimates of effect.1-3 As described in the CONSORT extension for pilot and feasibility trials (see reference #33 in manuscript), the purpose of these kinds of trials is to assess key elements related to the feasibility of conducting a

full-scale trial. The purpose is not for hypothesis testing, and feasibility trial designs are not set up or powered for this. The CONSORT extension very explicitly recommends against conducting significance testing in feasibility trials. Thus, consistent with best practices, we have not conducted significance tests.

In the revised manuscript, we have edited to clarify this (Page 16, Lines 9-12), “Since the purpose of this feasibility trial was to evaluate feasibility and identify any modifications to the intervention or trial plan, the trial was not designed or powered to test hypotheses about outcomes. Thus, consistent with best practices,[33] hypothesis tests were not conducted, but effect sizes for pre-post differences are shown.”

1. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. Arch Gen Psychiatry. 2006;63(5):484-489.
2. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: The what, why and how. BMC Med Res Methodol. 2010;10:1.
3. Van Teijlingen ER, Rennie AM, Hundley V, Graham W. The importance of conducting and reporting pilot studies: The example of the Scottish Births Survey. J Adv Nurs. 2001;34(3):289-295.

(6) The discussion repeats the results in parts without adding additional information - the discussion should concentrate on the meaning and implication of the results. Given that there were technical difficulties for some participants with hearing imparment, how will these be addressed in the main trial? What accommodations can be made?

The discussion includes information about how we plan to address feedback from participants on the content of modules and the use of vignettes (Page 26, Lines 3-15). We have added information about assessing and addressing impairment-related issues to facilitate participation (Page 26, Lines 21-23), “In the full-scale trial, we will assess for hearing and any other impairments that might limit participation, and we will seek appropriate assistance to be able to provide adaptations to meet participant needs.”

(7) Although this was a feasibility study, the small sample size and inability to provide data on statistical significance should be acknowledged as limitations or pls provide the relevant statistics.

Historically, researchers have sometimes proposed that pilot studies be used to estimate effect sizes and calculate sample sizes for full RCTs. However, it is recognized that using insufficiently powered pilot studies for this purpose often produces inaccurate and misleading estimates of effect.1-3 As described in the CONSORT extension for pilot and feasibility trials (see reference #33 in manuscript), the purpose of these kinds of trials is to assess key elements related to the feasibility of conducting a full-scale trial. The purpose is not for hypothesis testing, and feasibility trial designs are not set up or powered for this. The CONSORT extension very explicitly recommends against conducting significance testing in feasibility trials. Thus, consistent with best practices, we have not conducted significance tests.

In the revised manuscript, we have edited to clarify this (Page 16, Lines 9-12), “Since the purpose of this feasibility trial was to evaluate feasibility and identify any modifications to the intervention or trial plan, the trial was not designed or powered to test hypotheses about outcomes. Thus, consistent with best practices,[33] hypothesis tests were not conducted, but effect sizes for pre-post differences are shown.”

1. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. Arch Gen Psychiatry. 2006;63(5):484-489.

2. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: The what, why and how. BMC Med Res Methodol. 2010;10:1.
3. Van Teijlingen ER, Rennie AM, Hundley V, Graham W. The importance of conducting and reporting pilot studies: The example of the Scottish Births Survey. J Adv Nurs. 2001;34(3):289-295.

Reviewer #2

(1) This is a report on a feasibility trial, testing an educational program for current and future scleroderma support group leaders. Outcomes included feasibility, program quality, participant satisfaction, and participant personal outcomes. This is an important program to develop and will help improve overall scleroderma care.

We appreciate Reviewer #2's enthusiasm for the program and recognition that it will improve overall scleroderma care.

(2) The manuscript at times provides too much details, such as on page 12, describing entirely the development of one of the outcome measure, which has already been published (ref 46). Along the same lines, there are redundant points, such as explanation of why French participants were excluded, repeated multiple times in the paper. There are also a lot of details about the future RCT, while this report should be focusing on the current feasibility study. While these are only stylistic issues, it distracts from the focus of the paper and I would recommend the authors run through it a few more times to eliminate any unnecessary paragraphs.

We adhered closely to the CONSORT statement for feasibility and pilot trials in setting up the manuscript. We do believe that the information about the outcome measure is important. In fact, in comment #3, Reviewer #2 expresses concern about it. We believe that an explanation of its development with patient support group leaders and that it was developed with the same information used to develop the intervention is important. We don't believe that this is adequately communicated by referencing another document in this case, but we will cut this material if the editor believes that this would improve the manuscript. We briefly note in the Methods that only English groups were held, and we describe why this was a limitation in the appropriate part of the Discussion. We believe that this is standard and expected. Similarly, the feasibility trial has informed our plans for the now funded full-scale trial, and we believe it is important to communicate the core plans for that trial. If the editor disagrees, we will regretfully cut that material.

(3) A second issue I have is with using the SSGLSS as a major outcome measure in this study, and plans to use it as the primary outcome in the future RCT. Table 3 describes all the items in the scale. I would have expected, if the educational program has 13 modules, each addressing one or more learning objective, that the primary outcome of this educational intervention, would be to study whether the participants improved their knowledge in those learning objectives. As an analogy, if the videos were about 3 ways of cooking eggs, the learning assessment should be whether they now feel comfortable cooking eggs in those 3 ways, not so much all the ways in which they feel equipped at being a chef. The SSGLS certainly has items that were addressed by those modules, but some did not appear to be, such as item 1 financial resources. This would lower the score, but what does that mean? It means it probably just was not taught - not that the video educational program was not effective. I would recommend for the authors to consider this issue at least for the RCT.

The SSGLSS and the SPIN-SSLED intervention were both developed based on input from scleroderma support group leaders on leadership challenges and training needs. We have added a note to this effect in the description of the measure (Page 14, Lines 8-10). All of the items reflect challenges that are addressed in the program. To ensure that this is clear, we have added Table 1, which provides a description of the content of each of the 13 training modules. Of note, our full-scale

trial was recently funded by the Canadian Institutes of Health Research and was ranked 1st of almost 50 funding proposals (Health Services Evaluation and Interventions Research); the committee emphasized the co-development of the SSGLSS and SPIN-SSLED Program as a strength.

(4) I also wondered about the appropriateness of including emotional distress and physical function as outcome measures for the participants. There are many confounders that would affect these outcomes, especially physical function. How do the authors plan to control for these confounders. At face value/biological plausibility, there is not an obvious direct relationship between this type of educational program and overall emotional distress/physical function for a complex individual with many other life/health factors. Table 4 essentially shows minimal differences, yet the authors interpret these as moderate effect sizes (please include a reference to support this -Cohen's d typically refers to 0.5-0.8 as moderate (0.2-0.5 would be small)).

We don't disagree about using physical function in the full-scale trial, and our full-scale trial protocol does not include that. We thank Reviewer #2 for that comment. We have edited the labels to read "small to moderate", and we have provided a reference (reference 58).

(5) Since Standardized Mean Difference Effect Sizes appear to be the major outcome measure for the instruments used, please include a description of this and the interpretation of the effect sizes in the methods- data analysis section.

The main outcomes are feasibility outcomes, as described in the objectives. Please see our responses to comments #5 and #7 to Reviewer #1 above. We have provided numerical results for outcomes for descriptive purposes only, as described in the analysis section and consistent with best-practices for feasibility trials. We thank Reviewer #2 for suggesting that we provide a reference for effect sizes, and we have added reference #58.

(6) I am also unclear about including the planned trial outcome measures in this report. The objectives section states that the trial was to 1. evaluate feasibility 2. assess user satisfaction. In the discussion, the authors address these two objectives, but does not discuss the results of the planned trial outcome measures. The overall paper may be more focused and coherent if it only focused on the feasibility and user satisfaction.

One of our feasibility outcomes described in the objectives was related to assessing the performance of the outcome measures. Thus, we have reported the scale and descriptive information on those outcomes. We intended to do this, it was important for our evaluation of feasibility to ensure that there were no surprises in using the measures, and this is, we believe, standard practice in this type of feasibility trial. Of note, Reviewer #1 wanted more information than what we provided on these outcomes. Consistent with best practices in feasibility trials, we have added a description of why we did not conduct hypothesis tests, and we have reported the outcomes that can be transparently reviewed by users of our feasibility trial report.

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Yvonne Zurynski<br>Australian Institute of Health Innovation, Macquarie University |
| **REVIEW RETURNED** | 30-Jun-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | An interesting feasibility trial and it will be good to see the larger trial when ready. This study highlights the importance of peer |

| | |
|---|---|
| | support organisations for patients with chronic conditions, not just Scleroderma.<br>This studuy may serve as a nice example for other similar studies of peer support leader training. |

| | |
|---|---|
| **REVIEWER** | ADA MAN<br>University of Manitoba<br>Canada |
| **REVIEW RETURNED** | 17-Jun-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | 1.The issue of using the SSGLSS as a major outcome measure in this study, and plans to use it as the primary outcome in the future RCT - this has not been adequately addressed. There is nothing inherently wrong with SSGLSS nor the SPIN-SSLED intervention. It seems the author has misunderstood my concern. My concern is using the SSGLSS to determine how well the SPIN-SSLED worked. Some of the items in SSGLSS are not addressed in the SPIN-SSLED intervention. So in plain language, it's not fair to judge the SPIN-SSLED using an outcome measure that measures more things than are taught. Perhaps the authors could consider modifying the SPIN-SSLED intervention so that it includes some of the items on the SSGLSS that had a low score (e.g. financial), or consider not using some items on the SSGLSS scale that the authors do not think are relevant. Or if if the author still decides to use it, address this as a limitation of the SSGLSS for judging the effectiveness of the SPIN-SSLED.<br><br>2. About the appropriateness of including emotional distress and physical function as outcome measures for the participants. Since the author does not disagree about physical function - is this stated in the revised version? And how about the plans to address confounders for the emotional distress in the planned RCT?<br><br>3. About including the planned trial outcome measures in this report. May I suggest to add objective number 3 as "assessing the performance of the outcome measures." Under objectives in the Abstract, and also at the end of the Introduction, there are only 2 objectives :1. evaluate feasibility 2. assess user satisfaction. It will make it easier to read and understand the paper. |

## VERSION 2 – AUTHOR RESPONSE

Reviewer #1:

(1) An interesting feasibility trial and it will be good to see the larger trial when ready. This study highlights the importance of peer support organisations for patients with chronic conditions, not just Scleroderma. This study may serve as a nice example for other similar studies of peer support leader training

We thank Reviewer #1 for her positive comments.

Reviewer #2

(1) The issue of using the SSGLSS as a major outcome measure in this study, and plans to use it as the primary outcome in the future RCT - this has not been adequately addressed. There is nothing

inherently wrong with SSGLSS nor the SPIN-SSLED intervention. It seems the author has misunderstood my concern. My concern is using the SSGLSS to determine how well the SPIN-SSLED worked. Some of the items in SSGLSS are not addressed in the SPIN-SSLED intervention. So in plain language, it's not fair to judge the SPIN-SSLED using an outcome measure that measures more things than are taught. Perhaps the authors could consider modifying the SPIN-SSLED intervention so that it includes some of the items on the SSGLSS that had a low score (e.g. financial), or consider not using some items on the SSGLSS scale that the authors do not think are relevant. Or if if the author still decides to use it, address this as a limitation of the SSGLSS for judging the effectiveness of the SPIN-SSLED.

Reviewer #2 is not correct that some of the items in the SSGLSS are not addressed in the SPIN-SSLED intervention. We do not know how this conclusion was drawn from the information in the manuscript. We have edited to avoid any confusion on this point (Page 14, Lines 10-11), “A strength of using the SSGLSS as the primary outcome measure is that both the intervention and the SSGLSS were designed to reflect training needs of SSc support group leaders, and the items of the SSGLSS all reflect material covered in the program.”

(2) About the appropriateness of including emotional distress and physical function as outcome measures for the participants. Since the author does not disagree about physical function - is this stated in the revised version? And how about the plans to address confounders for the emotional distress in the planned RCT?

If we understand, the reviewer is asking us to state that we agree with her and are not planning to include physical function as an outcome measure in the full-scale trial. We believe this is out of the scope of the present manuscript. The protocol for the full-scale trial has been accepted for publication in Trials, and we do not include physical function in that protocol. Similarly, the analysis plan for the full-scale trial does not belong in the present manuscript, which results of the feasibility trial as no analyses were done in the feasibility trial. Rather, it belongs in the protocol for the full-scale trial, where it is, indeed, described in detail.

(3) About including the planned trial outcome measures in this report. May I suggest to add objective number 3 as "assessing the performance of the outcome measures." Under objectives in the Abstract, and also at the end of the Introduction, there are only 2 objectives :1. evaluate feasibility 2. assess user satisfaction. It will make it easier to read and understand the paper.

We describe two main objectives, consistent with our trial registry and feasibility trial protocol. In the objectives statement in the last paragraph of the introduction (Page 7, Lines 5-9), the first objective, related to trial feasibility, including scientific aspects, and it is clearly stated that part of this is “assessing performance of outcome measures”. We have added this phrase into the feasibility objective in the abstract per reviewer suggestion. We do not believe that we should change our objectives post hoc but will transparently report the pre-defined objectives as required by all trial reporting standards.

**VERSION 3 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Ada Man<br>University of Manitoba<br>Canada |
| **REVIEW RETURNED** | 04-Oct-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed the concerns to the best of their ability and while I still find some items unclear, the authors have disagreed with me on this, and perhaps it is just me as one reader |

| | |
|---|---|
| | who find certain points unclear. The project itself is an important piece of work which should be published and continued upon. |