PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Pilot Randomized Controlled Trial of an Energy Management Program for Adults on Maintenance Hemodialysis: The Fatigue-HD Study
AUTHORS	Farragher, Janine F.; Ravani, P; Manns, Braden; Elliott, Meghan; Thomas, Chandra; Donald, Maoliosa; Verdin, Nancy; Hemmelgarn, Brenda

VERSION 1 – REVIEW

REVIEWER	Farragher, Janine F.; Ravani, P; Manns, Braden; Elliott, Meghan; Thomas, Chandra; Donald, Maoliosa; Verdin, Nancy; Hemmelgarn, Brenda
REVIEW RETURNED	10-Jun-2021

GENERAL COMMENTS	Farragher et al. conducted a pilot randomized clinical trial with the objective of determine the feasibility and value of conducting a randomized controlled trial of an energy management program for people on chronic hemodialysis, as primary aim; and to estimate the effects of the Personal Energy Planning on fatigue in this population, as a secondary aim. The authors concluded that the program could be acceptable to patients and might lead to improvements in life participation. It also sticks well to the CONSORT Checklist for Clinical Trials. However, there are some areas of opportunity in the methodology section that can be improved. Here are some observations that could help improve the quality of your protocol: 1.-Enlarge the description of the intervention 2.-Further expound the details on the outcome measurement 3.-In the intervention section, it is described how study coordinators received training, however, it is recommended to clarify deeply the training and, describe or append the written guidebook. 4.- In the statistical analysis section, the authors described how they categorized the effect size estimates as very small, small, medium, large, or very large. However, there is not a reference to take that decision base on, or if they took that in an arbitrary way. So, it would be better if describe it. 5.- Regarding the statistics, it would be convenient to propose the comparison between the baseline values and the values immediately after treatment and the values at 12 weeks, preferably with an ANOVA of repeated measures or Kruskal Wallis, to be able to know if there was a change over time in each of the groups and not only the magnitude of the effect when comparing control vs. treatment immediately after treatment and at 12 weeks. 6.- In addition, as a pilot study, there is a small sample of participants per group, so it is advisable to make an adjustment to Kohen's d or perform some other test that is better adapted to

	small samples such as Hedges' g to try to reduce the risk of overestimating the effect with Cohen's D test.

REVIEWER	Nair, Devika Vanderbilt University
REVIEW RETURNED	27-Jun-2021

GENERAL COMMENTS	This is a randomized controlled pilot trial of 22 participants receiving chronic hemodialysis aimed to test the efficacy of a self-management program aimed to reduce patient perceptions of fatigue. Strengths of this investigation include the patient-prioritized area of investigation (Jhamb Am J Kidney Dis 2008), and the inclusion of patients with comorbidities and ages representative of the kidney disease population worldwide (Bikbov Lancet 2020). Major: - Would provide qualitative or quantitative detail as to why only 42 out of 159 eligible patients agreed to participate in the intervention as well as the high dropout rate for the intervention - Would provide information as to whether investigators felt that the information delivered to participants in the attention control arm - Using the results from this pilo, how will the investigators decide whether the effect size(s) found are clinically meaningful to patients? - Would include rationale and psychometric properties of each scale used Minor: - In the Abstract, would provide specific effect sizes to distinguish between ‘medium-sized’ and ‘large’ intervention effect - Would provide detail as to what determined whether patients received seven vs nine weeks - Would provide detail on participants’ dialysis recovery time, if possible, in subsequent iterations, as this may serve as an effect modifier for patient-reported symptoms that are used to measure intervention efficacy (Rayner Am J Kidney Dis 2014) - Would explicitly state what were the primary outcomes, secondary outcomes, exploratory outcomes, etc. - Would provide any other comorbidity information that may affect patient perceptions of fatigue (concurrent mood disorders, etc.)

REVIEWER	Finderup, Jeanette Aarhus Universitetshospital, Renal Medicine
REVIEW RETURNED	10-Jul-2021

GENERAL COMMENTS	The authors have conducted a pilot randomized controlled trial for patients on haemodialysis to make the patients work with their fatigue. It is only a pilot study, but the authors are aware of the aim of a pilot study and how to conclude the results. Some minor issues: • I the abstract you state that 159 patients were eligible and afterwards 30 patients met the eligibility criteria. It is confusing that you use the same term twice.

	• In your description of the control, you use the term study nurse coordinator and in data collection section, you use the term trained study coordinators. It is confusing with the different terms, but have the controls received usual care or another intervention by a study nurse coordinator? • In the data collection you write that collected data one week after the PEP program was completed, but other places you write immediately after. What is correct? Did you measure it one week after or immediately after? • I missed some information regarding the measurements and the PEP program but found it in tables. I would like that you flag that information about the measurements and the active components of the PEP program is to be found in table x and table y. • The patient and public involvement statement is not consistent with the short form of GRIPP2. Some language issues: • I do not prefer the term chronic haemodialysis, either haemodialysis or maintenance haemodialysis. • You use the term 'kidney failure population', I prefer the term 'the population with kidney failure' or something like that. • In the background, you write chronic fatigue, I think it is only fatigue, because fatigue is chronic. • You use the abbreviation COPM without explaining it. • One place you use the term renal patients. International guidelines recommend using the term kidney and I prefer, patients with kidney disease.

REVIEWER	Kim, HY New York Medical College, Department of Public Health
REVIEW RETURNED	19-Jul-2021

GENERAL COMMENTS	* Comments 1) Cohen's D is usually based on the mean difference and SD. However, data were presented with median and IQR in Table 3. Did the authors calculate the Cohen's D based on median? If so, what was the formula? 2) in Table 3, there were some strange directions for some change values. For example, in control group, the change in COPM-P in Immediate Post-Treatment Follow-up was -0.3. However, the baseline value was 4.3 and the there was an increase (4.7) in follow-up. There are also other measures with the same issues. Please clarify this. 3) Confidence interval for Cohen's D will also need to be presented in Table 3 to quantify the error imposed on an effect size. 4) The authors concluded that the intervention was associated with improved life participation on some measures. However, the interpretation should be done with a caution due to the very small sample size.

REVIEWER	Peipert, John Northwestern University Feinberg School of Medicine, Medical Social Sciences
REVIEW RETURNED	23-Jul-2021

GENERAL COMMENTS	I am enthusiastic about interventions of this type, which are needed for HD patients. However, there are many elements of this study that leave open whether this particular intervention should be pursued. Specific comments are below. Abstract – mentions medium sized effects, but no data is shown. Please bring-in data on the actual intervention. Abstract – please spell out COPM Introduction – The sentence on p. 5, lines 85-86 is misleading because fatigue is part of quality of life and well-being. Introduction – p. 5, lines 97-98. Please expand on why exercise programs are challenging to promote in this population. Introduction – pp. 5-6. In the paragraph from lines 101-112, please be clearer about whether PEPs improve life participation because they first reduce fatigue (i.e., mediated effect), through another mechanism, or directly. Currently, this causal pathway, and therefore the hypothesis, is unclear. Even if previous evidence does not make the causal pathway(s) totally clear, they should be articulated. Introduction – Currently there is insufficient rationale for the first aim of this pilot. The simple fact that there have not be previous efforts to enroll hemo dialysis patients into PEP interventions doesn't mean we don't know whether it's feasible to do so. Many similarly-structured health services type trials have been conducted with hemodialysis patients, leading to an expectation that patient would indeed be able to recruited and retained in this type of intervention. Please add further justification for this aim. What challenges with this type of intervention in particular would be expected? Methods – p. 7, lines 142-143. What does a score of >4 on items 5, 7, 8, 9 on the Fatigue Severity Scale indicate about fatigue severity? Methods – How was the control arm selected? Why not just usual care? Methods – How was 12 weeks post-intervention chosen? Are the investigators certain that this is sufficient time from intervention to observe effects? Methods – p. 9, line 180-181. Please clarify if questionnaires were administered by the coordinator (i.e., questions read aloud to the patient) or completed by the patient on their own (i.e., patients read and responded without having the coordinator read the question). Methods – The Methods is completely missing a section describing the study measures, including how cognitive

	assessment was performed. Table 2 gives some detail, but not how each is actually scored and what scores mean. This prevents any real interpretation of the results. Methods – A more appropriate reference for the SONG-HD Fatigue is the following: Ju A, Teixeira-Pinto A, Tong A, et al. Validation of a Core Patient-Reported Outcome Measure for Fatigue in Patients Receiving Hemodialysis. Clin J Am Soc Nephrol. 2020;15(11):1614. Methods – It is mentioned that simple and standardized treatment effects were calculated, then Cohen's d was described. By simple, do you just mean raw (non-standardized)? Methods – The description of treatment effect calculation is unclear. Did you examine changes from baseline, or cross-sectional differences at the immediate post-intervention and 12 week time points? If only cross-sectional analysis was conducted, how did you control for the starting levels of the outcome? Methods – There is a concern about the achieved sample size. The authors have cited recommendations to enroll 40 patients into the pilot, but only 30 were enrolled. (Although, despite citing the need for 40, I'm not clear why this number is any better than 30.) Please comment on whether failing to recruit 40 threatens the validity of the study. Results – p. 11. I'm not clear on why the proportion of patients who met eligibility criteria was estimated instead of simply calculated. Results – p. 11. Please provide numbers of patients not meeting each eligibility criterion. Results – p.11. Please comment more on the allowable level of cognitive impairment for participation. Cognitively stable patients were sought for recruitment (thought a definition of cognitively stable is not given), but 30% screened positive for cognitive impairment. Results – p. 15. The rationale for COPM as the primary outcome in a future trial has a logical flaw. The authors cite that the largest effect was observed on the COPM, but that should not be the basis for primary outcome selection. The outcome should be selected on its ability to best reflect whatever the intervention is seeking to change. The size of the effect reflects how successful the intervention is at doing what it intends to do. So, the COPM might be most appropriate if the goal of the intervention is to improve participation in patient-selected activities. Results/Discussion – Related to the last comment, the authors should acknowledge and discuss that fact that effect sizes at 12 weeks for all outcomes besides the COPM scales were either trivial or actually favored the control. This raises serious concerns for me about what the intervention really does and how useful it is. It also goes to my earlier comment about the causal pathway. Based on these results, I do not think the authors can really argue that the intervention impacts fatigue as a mechanism for increasing life participation. Rather, it appears to have an impact on life participation directly, and doing so without decreasing fatigue in the process. The FMQ results really underscore this

	point. Also, why the discrepancy between effects on different life participation measures? It is somewhat worrisome that the effectiveness of the intervention seems to depend completely on which scale is used to evaluate it. Results – p. 15, lines 319-321. Please specify if the clinically-meaningful change of >2 points on the COPM refers to a between group difference or within group change. Table 3. – I think for most variables, the statistics shown are n (%), but this is not stated.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author:

Farragher et al. conducted a pilot randomized clinical trial with the objective of determine the feasibility and value of conducting a randomized controlled trial of an energy management program for people on chronic hemodialysis, as primary aim; and to estimate the effects of the Personal Energy Planning on fatigue in this population, as a secondary aim. The authors concluded that the program could be acceptable to patients and might lead to improvements in life participation. It also sticks well to the CONSORT Checklist for Clinical Trials. However, there are some areas of opportunity in the methodology section that can be improved.
Here are some observations that could help improve the quality of your protocol:

1. Enlarge the description of the intervention.
Thank you; we have included additional description of the intervention in the body of the manuscript. A detailed description of the intervention procedures is also included in Table 1. The intervention has also been described in detail in the published protocol and other published sources, which are now referenced in the current manuscript.
2. Further expound the details on the outcome measurement
Thank you for this suggestion; we have included additional detail about the outcome measures used to facilitate clearer interpretation of results.
3. In the intervention section, it is described how study coordinators received training, however, it is recommended to clarify deeply the training and, describe or append the written guidebook.
Thank you for this suggestion; we have included additional detail about the training provided.

4. In the statistical analysis section, the authors described how they categorized the effect size estimates as very small, small, medium, large, or very large. However, there is not a reference to take that decision base on, or if they took that in an arbitrary way. So, it would be better if describe it.
Thank you for this feedback. The effect size categorizations were informed by the following reference: Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic
This reference has been added to the manuscript. Note that based on your below suggestion, we have changed the effect sizes reported from Cohen's D to Hedge's G. As Hedge's G is a modified form of Cohen's D that adjusts for small sample sizes, both effect sizes are to be interpreted and categorized in the same way.
5. Regarding the statistics, it would be convenient to propose the comparison between the baseline values and the values immediately after treatment and the values at 12 weeks, preferably with an

ANOVA of repeated measures or Kruskal Wallis, to be able to know if there was a change over time in each of the groups and not only the magnitude of the effect when comparing control vs. treatment immediately after treatment and at 12 weeks.
Thank you for this input. We would argue that changes in the outcomes over time within groups are confounded by several factors associated with receiving an intervention and being observed during a study (eg. placebo effect, social desirability bias, Hawthorne effect). Thus, we believe there is little information to be gained from performing statistical tests to determine if changes over time within groups are statistically significant or not.
In addition, this is a pilot study designed to inform a larger randomized controlled trial. As such, this pilot study is not adequately powered to perform tests of statistical significance and derive meaningful information from those tests. This study is instead meant to provide information about whether a larger trial would be valuable or not. That is why we use effect size estimates, rather than tests of statistical significance, throughout the study.
6. In addition, as a pilot study, there is a small sample of participants per group, so it is advisable to make an adjustment to Kohen's d or perform some other test that is better adapted to small samples such as Hedges' g to try to reduce the risk of overestimating the effect with Cohen's D test.

Thank you very much for this helpful input. We have changed all effect size estimates in the study to Hedge's G instead of Cohen's D, as upon further consideration, we agree that Hedge's G is the more appropriate effect size calculation to use.

Reviewer: 2
Comments to the author:
This is a randomized controlled pilot trial of 22 participants receiving chronic hemodialysis aimed to test the efficacy of a self-management program aimed to reduce patient perceptions of fatigue.

Strengths of this investigation include the patient-prioritized area of investigation (Jhamb Am J Kidney Dis 2008), and the inclusion of patients with comorbidities and ages representative of the kidney disease population worldwide (Bikbov Lancet 2020).
Major:

7. Would provide qualitative or quantitative detail as to why only 42 out of 159 eligible patients agreed to participate in the intervention as well as the high dropout rate for the intervention.
Thank you for this feedback. This finding is discussed in the discussion section on page 15.
8. Would provide information as to whether investigators felt that the information delivered to participants in the attention control arm
Although this comment seems to have been unintentionally cut short, we conjecture that the reviewer was asking whether the information from the attention control arm might have affected fatigue management protocols or the outcomes in any way.
The information in the attention control arm was edited to ensure that fatigue was never directly discussed. Although there were some related topics addressed during the attention control, these topics were deemed necessary to maintain blinding among participants (ie. to convince participants in the control arm that they at least plausibly might be receiving the intervention). There was never, however, any direct attempt made to assist participants in changing behaviours or adopting new self-management strategies within the control arm; it consisted purely of information provision only. Thus, we believe the likelihood of behaviour change impacting the outcomes in the control arm to have been reasonably low, and as low as we could have achieved while still maintaining study blinding.
9. Using the results from this pilot, how will the investigators decide whether the effect size(s) found are clinically meaningful to patients?
Thank you for this inquiry. The minimal clinically important difference for the COPM measure has previously been reported as ≥2. This will be used as the standard for clinical meaningfulness of changes in this and future PEP studies.

10. Would include rationale and psychometric properties of each scale used
This information has been provided in Table 2. We also included additional information about measure scoring, to aid interpretation of study results.

Minor:

11. In the Abstract, would provide specific effect sizes to distinguish between 'medium-sized' and 'large' intervention effect.
Thank you for this feedback. Due to word limitations, we are unable to add details about effect sizes for each measure into the abstract. However, the reader can find this information in the results section of the manuscript. Given that testing the intervention's effectiveness is not the primary objective of the study, we believe that this is sufficient.
12. Would provide detail as to what determined whether patients received seven vs nine weeks.
Thank you, this has been added to the manuscript.
13. Would provide detail on participants' dialysis recovery time, if possible, in subsequent iterations, as this may serve as an effect modifier for patient-reported symptoms that are used to measure intervention efficacy (Rayner Am J Kidney Dis 2014)
Thank you for this suggestion. We agree that dialysis recovery time would be an interesting variable to examine in the context of this intervention and its effectiveness, although we would suggest that it should not affect the validity of the measures used, as they assess symptoms over the course of a one week or one month period.
14. Would explicitly state what were the primary outcomes, secondary outcomes, exploratory outcomes, etc.
Thank you, the primary outcome has been highlighted in the abstract. The primary vs. secondary outcomes are also identified in the final paragraph of the introduction section.

15. Would provide any other comorbidity information that may affect patient perceptions of fatigue (concurrent mood disorders, etc.)
Thank you for this suggestion. We measured depression at baseline using a screening tool among all participants to assess the extent of depression in the study sample. We would assert that this would likely provide a more accurate estimate of the prevalence of depression in the study sample than diagnosed mood disorders, given that mood disorders are notoriously underdiagnosed in the hemodialysis population (eg. refer to Chilcot et al., https://doi.org/10.1159/000124749).
Reviewer: 3
The authors have conducted a pilot randomized controlled trial for patients on haemodialysis to make the patients work with their fatigue. It is only a pilot study, but the authors are aware of the aim of a pilot study and how to conclude the results.

Some minor issues:
16. I the abstract you state that 159 patients were eligible and afterwards 30 patients met the eligibility criteria. It is confusing that you use the same term twice.
Thank you for this point of clarification; we have adjusted the description to be clearer.

17. In your description of the control, you use the term study nurse coordinator and in data collection section, you use the term trained study coordinators. It is confusing with the different terms, but have the controls received usual care or another intervention by a study nurse coordinator?
We have removed the term "nurse coordinator" to reduce confusion. The controls received another intervention by a study nurse coordinator, which is described in the methods section.

18. In the data collection you write that collected data one week after the PEP program was completed, but other places you write immediately after. What is correct? Did you measure it one week after or immediately after?

As the patient burden of completing a treatment session, and completing all post-intervention questionnaires, was too great for one session, we broke these activities up into two sessions - with questionnaires being completed one week after the final intervention session.
We have clarified the language throughout the manuscript to be more consistent.

19. I missed some information regarding the measurements and the PEP program but found it in tables. I would like that you flag that information about the measurements and the active components of the PEP program is to be found in table x and table y.
Thank you for this feedback; we have further highlighted the references to Table 1 and Table 2.

20. The patient and public involvement statement is not consistent with the short form of GRIPP2.
Thank you for this suggestion. We have incorporated additional information to increase consistency with GRIPP2-SF.

Some language issues:

21. I do not prefer the term chronic haemodialysis, either haemodialysis or maintenance haemodialysis.
Thank you for this observation; it has been modified to “maintenance hemodialysis” throughout the manuscript.

22. You use the term ‘kidney failure population’, I prefer the term ‘the population with kidney failure’ or something like that.
Thank you for this observation; it has been modified throughout the manuscript to “people with kidney failure”.
23. In the background, you write chronic fatigue, I think it is only fatigue, because fatigue is chronic.
There are also acute instances and types of fatigue (eg. related to acute illness or hospitalizations); thus, we believe specifying that this fatigue is chronic is a valuable clarification.
24. You use the abbreviation COPM without explaining it.
Thank you, we have added in the full name of the measure when it is first mentioned in the manuscript.
25. One place you use the term renal patients. International guidelines recommend using the term kidney and I prefer, patients with kidney disease.
Thank you for this observation – we have corrected the term to people with kidney disease.

Reviewer: 4

26. Cohen’s D is usually based on the mean difference and SD. However, data were presented with median and IQR in Table 3. Did the authors calculate the Cohen’s D based on median? If so, what was the formula?
Cohen’s D was calculated based on mean difference and SD, according to the formulas typically used. Median and IQR were reported in Table 3 as optimal measures of central tendency and variance, based on the small sample size of the intervention and control groups.
However, note that we have since been advised by another reviewer to change our effect size estimates from Cohen’s D to Hedge’s G to correct for the small sample sizes of the study, which we have done in the revised version of the manuscript.
27. in Table 3, there were some strange directions for some change values. For example, in control group, the change in COPM-P in Immediate Post-Treatment Follow-up was -0.3. However, the baseline value was 4.3 and the there was an increase (4.7) in follow-up. There are also other measures with the same issues. Please clarify this.
The median values at individual timepoints do not always correspond to median change values. For example, consider the following set of example values:

Baseline: 1, 2, 4
Post-intervention: 4, 1, 3
Change: +3, -1, -1
The median at baseline is 2 and the median at post-intervention is 3, but the median change value is -1. This is because not only the direction of changes, but also the magnitude of changes, impacts which value ends up as the median value.
28. Confidence interval for Cohen's D will also need to be presented in Table 3 to quantify the error imposed on an effect size.
Thank you for this suggestion. Given that this is a pilot study with a modest sample size, error ranges for standardized effect sizes are wide and thus not particularly informative. As the purpose of this pilot study is not to establish the efficacy of the program, but rather to provide rough estimates of efficacy for planning a future hypothesis-testing trial, we believe the estimates provided are sufficient at this stage. Note that we did include in Appendix A best-case and worst-case scenario estimates for the potential range of Cohen's D (now Hedge's G) to account for study withdrawals. We assumed missing data from study withdrawals in the intervention group were, a. consistent with the average intervention effect size (best-case scenario), and b. consistent with the average control effect size (worst-case scenario).
29. The authors concluded that the intervention was associated with improved life participation on some measures. However, the interpretation should be done with a caution due to the very small sample size.
We agree wholeheartedly with this point. We believe we have been clear in the manuscript that this is a pilot study, and that a larger trial is needed to draw conclusions about the intervention's effects. This point has been emphasized in the abstract, discussion, and conclusions of the manuscript.

Reviewer: 5

I am enthusiastic about interventions of this type, which are needed for HD patients. However, there are many elements of this study that leave open whether this particular intervention should be pursued.
30. Abstract – mentions medium sized effects, but no data is shown. Please bring in data on the actual intervention.
Thank you for this comment. Due to word limits for the abstract, there is not space to include the specific data within the abstract. However, these data are provided in the results section of the paper, which the reader can reference for further information about the study results.

31. Abstract – please spell out COPM
Thank you, we have corrected this with the full name of the assessment tool.

32. Introduction – The sentence on p. 5, lines 85-86 is misleading because fatigue is part of quality of life and well-being.
We agree that there is conceptual overlap; however, fatigue can impact other areas of quality of life and well-being that go beyond the concept of fatigue (eg. social roles, functional disability). We therefore believe the statement in the manuscript is accurate, as supported by the referenced literature.

33. Introduction – p. 5, lines 97-98. Please expand on why exercise programs are challenging to promote in this population.
We have added additional information on why exercise programs are challenging to promote in this population.

34. Introduction – pp. 5-6. In the paragraph from lines 101-112, please be clearer about whether PEPs improve life participation because they first reduce fatigue (i.e., mediated effect), through

another mechanism, or directly. Currently, this causal pathway, and therefore the hypothesis, is unclear. Even if previous evidence does not make the causal pathway(s) totally clear, they should be articulated.
Thank you for this query; we have added information about potential causal pathways into the introduction.

35. Introduction – Currently there is insufficient rationale for the first aim of this pilot. The simple fact that there have not been previous efforts to enroll hemodialysis patients into PEP interventions doesn't mean we don't know whether it's feasible to do so. Many similarly structured health services type trials have been conducted with hemodialysis patients, leading to an expectation that patient would indeed be able to recruited and retained in this type of intervention. Please add further justification for this aim. What challenges with this type of intervention in particular would be expected?
Thank you for this observation. Although many other psychosocial and educational interventions have been trialed in the maintenance hemodialysis population, the acceptability of the energy management approach specifically in the maintenance hemodialysis population has never been explored. There is always the possibility that patients will not respond well to, or connect with, this approach – and the purpose of a pilot trial is to establish more conclusively that an approach is feasible and acceptable to patients. Thus, it was important to explore the feasibility of the energy management approach specifically. We have included an additional sentence which clarifies this in the introduction.

36. Methods – p. 7, lines 142-143. What does a score of >4 on items 5, 7, 8, 9 on the Fatigue Severity Scale indicate about fatigue severity?
These items on the Fatigue Severity Scale assess the impact of fatigue on the participant's life participation specifically. This eligibility requirement was included in the protocol based on early pilot data, which emphasized that participants experiencing life participation difficulties are the most likely to benefit from the PEP program.

37. Methods – How was the control arm selected? Why not just usual care?
Usual care does not control for the effects of additional attention received during an intervention, the placebo effect, the Hawthorne effect (the documented effect of simply being observed on outcomes in a study), or social desirability bias (the desire for participants to report receiving benefit from people who have tried to help them). These are all serious threats to internal validity when subjective outcome measures are being used to assess outcomes in a study (as was the case in this trial). As such, an attention control was deemed the most appropriate methodology to use. We chose this particular control condition (general education about kidney disease) as it controlled for the aforementioned sources of bias, while not directly addressing the outcome of interest (ie. fatigue).

38. Methods – How was 12 weeks post-intervention chosen? Are the investigators certain that this is sufficient time from intervention to observe effects?
12 weeks was chosen to gauge the maintenance of effects of the intervention in a feasible manner for this smaller-scale pilot study. In a larger trial, we would aim to assess longer-term outcomes (ie. 6-12 months) to establish the possibility of extinction of effects over a longer period.

39. Methods – p. 9, line 180-181. Please clarify if questionnaires were administered by the coordinator (i.e., questions read aloud to the patient) or completed by the patient on their own (i.e., patients read and responded without having the coordinator read the question).
As per the instructions of the assessments being used, patients were asked to read and respond to the questionnaires independently. In the event that participants were unable to read the questions due to visual or other barriers, the coordinator was instructed to read the questions aloud to participants to aid questionnaire completion.

40. Methods – The Methods is completely missing a section describing the study measures, including how cognitive assessment was performed. Table 2 gives some detail, but not how each is actually scored and what scores mean. This prevents any real interpretation of the results.
Thank you for this feedback. We have included additional information about the measures in the corresponding table 2 to aid interpretation of the study results.

42. Methods – A more appropriate reference for the SONG-HD Fatigue is the following:
Ju A, Teixeira-Pinto A, Tong A, et al. Validation of a Core Patient-Reported Outcome Measure for Fatigue in Patients Receiving Hemodialysis. Clin J Am Soc Nephrol. 2020;15(11):1614.
Thank you for this suggestion. We have modified the reference.

43. Methods – It is mentioned that simple and standardized treatment effects were calculated, then Cohen's d was described. By simple, do you just mean raw (non-standardized)?
Thank you for this point of clarification - yes, we meant raw (non-standardized) treatment effects. We have modified the language accordingly.

44. Methods – The description of treatment effect calculation is unclear. Did you examine changes from baseline, or cross-sectional differences at the immediate post-intervention and 12 week time points? If only cross-sectional analysis was conducted, how did you control for the starting levels of the outcome?
A paired samples t-test was performed, which compares the magnitude of change from baseline to post-intervention between the two groups on selected outcomes.

45. Methods – There is a concern about the achieved sample size. The authors have cited recommendations to enroll 40 patients into the pilot, but only 30 were enrolled. (Although, despite citing the need for 40, I'm not clear why this number is any better than 30.) Please comment on whether failing to recruit 40 threatens the validity of the study.
As this study is a pilot study that was not designed to examine statistically significant differences in outcomes, the validity of the study is not threatened. A sample size of 40 was chosen to maximize the precision with which we could estimate the intervention effect size, for the purpose of calculating the sample size needed for a larger hypothesis-testing trial. However, there is no consensus for sample sizes for pilot trials, and achieving a smaller sample size just means we have a less precise estimate upon which to base our future trial sample size. It is not a major issue for the purposes of this study. Further explanation and justification of this is provided in the published protocol, which is referenced in the paper. As this information has been published previously, it was not repeated in this article for the purposes of minimizing the length and word count of this paper.

46. Results – p. 11. I'm not clear on why the proportion of patients who met eligibility criteria was estimated instead of simply calculated.
Estimating the proportion allows us to apply the estimate to calculate how many patients would meet eligibility criteria when performing a larger, hypothesis-testing randomized controlled trial.

47. Results – p. 11. Please provide numbers of patients not meeting each eligibility criterion.
This information was not collected for this study due to requirements of the local research ethics board.

48. Results – p.11. Please comment more on the allowable level of cognitive impairment for participation. Cognitively stable patients were sought for recruitment (thought a definition of cognitively stable is not given), but 30% screened positive for cognitive impairment.
The clinical team was consulted for their professional opinion to identify patients who had severe cognitive impairments, that would preclude adequate communication or comprehension to perform basic responsibilities of the trial (eg. filling out questionnaires). These patients were excluded from the

trial without further evaluation. Otherwise, patients were included in the trial and their cognitive status was measured, to enable better interpretation of results and examination of the impact of cognitive dysfunction on study outcomes. In exploratory analyses, the presence of cognitive impairment on the MiniCog was found to have no relationship with study outcomes.

49. Results – p. 15. The rationale for COPM as the primary outcome in a future trial has a logical flaw. The authors cite that the largest effect was observed on the COPM, but that should not be the basis for primary outcome selection. The outcome should be selected on its ability to best reflect whatever the intervention is seeking to change. The size of the effect reflects how successful the intervention is at doing what it intends to do. So, the COPM might be most appropriate if the goal of the intervention is to improve participation in patient-selected activities.
We agree wholeheartedly that outcome measure decisions should be driven by the goal of the intervention. However, other factors such as the acceptability of the measure, and its psychometric performance in the target population, are also important considerations. Given the dearth of psychometric data on fatigue and life participation measures in the kidney disease population in general, this study was valuable in that it demonstrated that the COPM appears able to detect change associated with the intervention (ie. it had sufficient validity, reliability and responsiveness to ultimately capture large and very large-sized effects in life participation associated with the intervention). This study also established that the COPM is acceptable to patients; ie., there were no unexpected challenges with its use in this population. Given that part of the broader purpose of a pilot study is to pilot-test study protocols and ensure no unexpected challenges arise, the information gleaned about the COPM was valuable in that it enabled us to confirm that this was a viable measure to use in a full-scale trial. The other measures used in this study did not demonstrate similar effect sizes, and one potential reason is that they might not have adequate validity, reliability or responsiveness in the kidney disease population. Thus, we were able to rule them out as optimal measures to use for a larger trial.

50. Results/Discussion – Related to the last comment, the authors should acknowledge and discuss that fact that effect sizes at 12 weeks for all outcomes besides the COPM scales were either trivial or actually favored the control. This raises serious concerns for me about what the intervention really does and how useful it is. It also goes to my earlier comment about the causal pathway. Based on these results, I do not think the authors can really argue that the intervention impacts fatigue as a mechanism for increasing life participation. Rather, it appears to have an impact on life participation directly, and doing so without decreasing fatigue in the process. The FMQ results really underscore this point. Also, why the discrepancy between effects on different life participation measures? It is somewhat worrisome that the effectiveness of the intervention seems to depend completely on which scale is used to evaluate it.
Thank you for this valuable inquiry. We believe there are several points of potential explanation for the results observed in this study.
First of all, the validity of other life participation measures has never been established in the chronic kidney disease population; at present, there is a dearth of validated measures of life participation to use in the chronic kidney disease population. As such, measures such as the Reintegration to Normal Living Index or Fatigue Management Questionnaire might not capture relevant areas or aspects of life participation among this population, might not be worded in an understandable way, or might not be responsive enough to capture changes in the outcomes, among other potential explanations. Some of these measures are also less “personalized” and capture broader aspects of life participation, which may or may not be important or meaningful to individual patients. People with kidney disease themselves have recently emphasized, in patient consultation exercises, the importance of using measures of life participation that are idiosyncratic and allow for individual differences in life participation priorities. The COPM measure is unique in that it enables patients to identify their individual life participation goals, and then captures changes in those goals. Thus, it is likely the most valid measure of life participation that can be used. It is therefore perhaps unsurprising that it

performed best in capturing changes associated with the intervention, and not necessarily concerning that other measures did not demonstrate similar positive results, given the previous potential limitations described.
The fatigue measures that were used (and the vast majority that are available and validated in the chronic kidney disease population) do not necessarily capture exertional fatigue, which is the type of fatigue targeted in energy management; when patients report their overall fatigue level, they might perceive and report that their underlying fatigue has not changed, despite experiencing less exertional fatigue while they do activities. Thus, not observing changes in fatigue does not necessarily indicate that the causal pathway (now outlined in the introduction) is not accurate. We would also raise that people on maintenance hemodialysis have identified limitations in life participation as the most important and high-priority aspect of their fatigue experience. Thus, if life participation is improved by the PEP program, it is perhaps less important whether other aspects of fatigue change as well.
The magnitude of any changes favouring the control condition were miniscule, and in all likelihood fall within the measurement error parameters of those measures.
We have included some additional reflections in the discussion section that address these points.

51. Results – p. 15, lines 319-321. Please specify if the clinically-meaningful change of >2 points on the COPM refers to a between group difference or within group change.
This refers to a within-group change.

52. Table 3. – I think for most variables, the statistics shown are n (%), but this is not stated.
This information was included in a footnote below the table.

VERSION 2 – REVIEW

REVIEWER	Nair, Devika Vanderbilt University
REVIEW RETURNED	14-Sep-2021

GENERAL COMMENTS	The authors have adequately responded to my stated suggestions for review and editing