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)	Individually tailored whole-body vibration training to reduce symptoms of chemotherapy-induced peripheral neuropathy: study protocol of a randomized-controlled trial - VANISH
AUTHORS	Streckmann, Fiona; Hess, Viviane; Bloch, Wilhelm; Décard, Bernhard; Ritzmann, Ramona; Lehmann, Helmar; Balke, Maryam; Koliamitra, Christina; Oschwald, Vanessa; Elter, Thomas; Zahner, Lukas; Donath, Lars; Roth, Ralf; Faude, Oliver

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Jennifer Gewandter Institution and Country: University of Rochester; USA Competing interests: None declared
REVIEW RETURNED	13-Jun-2018

GENERAL COMMENTS	The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Reviewer name: Mohsen Zahiri Institution and Country: Baylor College of Medicine Competing interests: None declared
REVIEW RETURNED	04-Jul-2018

GENERAL COMMENTS	This manuscript describes a study protocol for a randomized-controlled trial of the 12 weeks use of a Whole Body Vibration (WBV) to improve the symptoms of CIPN such as lack of peripheral sensation, pain, weakness, etc. The protocol is well written and the rationale for the work is clear from the introduction. However, I have several concerns mentioned in the following: Major concerns: It was not clear that the subjects are patients under chemotherapy or they have already finished the chemo in less than 3 months. Time of chemo can significantly affect the severity of neuropathy. If patients got chemo less than two weeks from the data collection, this timing can influence the results. To have a consistent data collection for the patients under chemotherapy, it is necessary to collect data in the same timeframe and after a specific time of chemotherapy. In addition, to avoid any inconsistency, the patients who finish the chemotherapy in the middle of study should be excluded since the temporary effect of chemo can influence the data and show a significant improvement on the patients. Determine the highest neuromuscular response in the first two visits does not provide the optimal dose of training recommendation. To compare the efficacy of dose, each patient should be exposed to different dose for a specific time period like 4-6 weeks.

	Then, different progress based on different dose can show the optimal dose. For instance, it might be possible that other frequencies, either less than the threshold or larger than a threshold, can have a better influence. Therefore, I am not sure if the claim of defining the optimal dose of training will be proper for this study design. Minor concerns: It might be better if authors talk about the FACT-GOG-Ntx earlier since this factor is the primary outcome of this study. In the strength section, the authors claimed that there is not any treatment option for CIPN. Therefore, this statement brings the idea that the proposed intervention can be used as a treatment for these type of patients. By knowing that there are plenty of studies showing the effectiveness of exercise to improve these symptoms, the authors should either count the exercise as a treatment or not including WBV as a treatment option. As mentioned in the background, reference 18 shows the effectiveness of WBV for CIPN patients with 11 subjects. Therefore, it would be better if authors talk more about the difference between their study and reference 18. Only mentioning “objective and systematic” does not show the difference of this study from the mentioned study. In addition, the terminology of “systematic” is unclear to me. You are conducting several approaches to assess the neuropathy. Several validated methods will be used in this study and I am not sure if it is necessary to measure all of these parameters. Based on the large number of variables you are analyzing, a comment on why you are, or are not, going to use Bonferroni correction (or similar) to reduce the chances of obtaining false-positive results would be appropriate. I have not seen any place talking about the adjustment in the statistics section. Are you going to report the results without any adjustment or estimated variable will be reported based on the adjustment? Since subjects will be between 18-80 years old, the chance of non-normal distribution for this wide range of age is high. So, defining a method to use adjustment is necessary for this section. Is there any reason to use multiple imputations for missing data? Linear mixed models are a valid method used for missing data and allow for testing group difference. I have not seen any place talking about the trial registry of this study in Clinicaltrial.gov. It might be good to mention the trial registry on Clinicaltrial.gov. Suggestion As authors mentioned in the background, exercise is counted as one of the main treatment to improve the CIPN symptoms. The study can be more valuable if authors use exercise for the control group. Since the previous studies have shown the effectiveness of WBV, comparing the WBV group with no intervention group might not add many impacts on this field. As you might now, there are 3 senses that contribute to balance: vestibule, vision, and peripheral sensation. So, having an eyes closed test would be a great challengeable balance test to measure the improvement of sensation for these patients with neuropathy. Is there any reason that you have not included this test to the study design?

REVIEWER	Reviewer name: Andrew Hinde Institution and Country: University of Southampton Competing interests: None declared.
REVIEW RETURNED	08-Aug-2018

GENERAL COMMENTS	This description of a study protocol is largely convincing. My main concern relates to the sample size. While you are correct that, if you are anticipating a medium effect size, a sample of 36 is required to identify it, I think that to assume that recruiting 44 participants will give you 36 valid cases may be a little optimistic. To the extent that effects are less than 'medium' for some secondary outcomes, you are likely to find that your study is underpowered. Therefore I would be inclined EITHER to strengthen the fourth bullet point on page 3 to suggest that 'some secondary endpoints are likely to be underpowered' OR to try to raise the sample size recruited to, say, 50, to give you a bit of extra leeway. It may be too late to increase the sample size, of course, or it may be difficult. But you are cutting things a bit fine with a recruitment of 44. On the question of effect size, you cite f = 0.25 (p. 12, l. 9). But I think you mean f = 0.75. I have checked the calculations and a value of f = 0.75 is consistent with a sample size of 36. On a related point, with 44 cases (reduced by attrition to 36) you should be able to identify whether the whole-body vibration (WBV) training has a significant impact on the symptoms. But I think that establishing the 'optimal dose for training recommendations' (p. 6, l. 12) is optimistic. For a start, on p. 10, ll. 23-27 you suggest that the optimal setting (which is partly defined on the basis of dose) may be specific to each individual. If you find this, then the idea of an 'optimal dose for training recommendations ... with regard to future clinical trials' does not seem very relevant. You might, of course, find that a setting with the same dose produces the highest neuromuscular response in most, or all, patients, but this will not be known until the initial training sessions are completed. On p. 8, ll. 16-26 you describe the assessment procedures. I think clarification would be useful here. How many investigators do you have? And what do you mean by assessments being 'aligned among the investigators' (l. 17)? Later on, you say that '[e]xaminers will be trained by a gold-standard examiner' (l. 22), and refer to 'testers' (l. 25). So we have investigators, examiners, a 'gold-standard examiner' and testers. And these people are doing not investigations, or examinations, or tests, but assessments. I am afraid that I was quite confused by the time I reached l. 26. Why not take a bit more space and explain more slowly how you plan to organise the assessments? p. 10, ll. 23-6. 'After a short familiarization with the training setting ... conditions of WBV within the first two training sessions'. Is this training done on all patients, or only those randomised to the intervention group? The statement on p. 11, ll. 9-11 suggests that it is done on all patients before randomisation. Can you clarify this? Minor points and typographical errors p. 5, l. 3 Delete comma after 'parameters'. p. 6, l. 3 Change 'behold' to 'hold'.

	p. 6, l. 25 Change 'oft he' to 'of the'; and change 'publictaion' to 'publication'. p. 13, l. 5 Delete comma after 'fact'. p. 13, l. 9 Define 'SENIAM'.

REVIEWER	Reviewer name: Yoon Kong Loke Institution and Country: University of East Anglia Competing interests: No competing interests.
REVIEW RETURNED	09-Aug-2018

GENERAL COMMENTS	Please clarify the relationship of the pilot study to this one, and whether the data will be pooled. Are the data from those participants being included in this trial? If this is the case, then this protocol should state that it is the continuation of the pilot study as a full-scale RCT. Please state in abstract and early on in the Methods that this is outcome assessor blinded design. Please also provide a justification for this design rather than double blind - was there no possibility of a sham procedure? Multiple imputation may or may not be reliable for the outcome data that is missing. I suggest that sensitivity analysis should be carried out with, and without, the multiple imputation. How will you consider multiple testing of significance - there are lots of outcomes and time points being tested? For the categorical or dichotomous outcomes - are these being analysed using ANOVA as well?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

> Reviewer Name: Jennifer Gewandter

> Institution and Country: University of Rochester; USA

> Please state any competing interests or state ‘None declared’: None declared

> Please leave your comments for the authors below

> see attachment

Thank you for your time and the effort you put into the review of our paper. We appreciate your valuable comments and have responded to them in your attachment.

Reviewer: 2

> Reviewer Name: Mohsen Zahiri

> Institution and Country: Baylor College of Medicine

> Please state any competing interests or state ‘None declared’: None declared

> This manuscript describes a study protocol for a randomized-controlled trial of the 12 weeks use of a Whole Body Vibration (WBV) to improve the symptoms of CIPN such as lack of peripheral sensation, pain, weakness, etc. The protocol is well written and the rationale for the work is clear from the introduction. However, I have several concerns mentioned in the following:

Response: Thank you for your valuable comments and the effort you put into the detailed review. It is much appreciated.

Major concerns:

1. It was not clear that the subjects are patients under chemotherapy or they have already finished the chemo in less than 3 months. Time of chemo can significantly affect the severity of neuropathy. If patients got chemo less than two weeks from the data collection, this timing can influence the results. To have a consistent data collection for the patients under chemotherapy, it is necessary to collect data in the same timeframe and after a specific time of chemotherapy. In addition, to avoid any inconsistency, the patients who finish the chemotherapy in the middle of study should be excluded since the temporary effect of chemo can influence the data and show a significant improvement on the patients.

Response: Thank you for bringing this to our attention. We hope to have clarified this now: See Abstract, Methods Section, first sentence; as well as p.7/patients

2. Determine the highest neuromuscular response in the first two visits does not provide the optimal dose of training recommendation. To compare the efficacy of dose, each patient should be exposed to different dose for a specific time period like 4-6 weeks. Then, different progress based on different dose can show the optimal dose. For instance, it might be possible that other frequencies, either less than the threshold or larger than a threshold, can have a better influence. Therefore, I am not sure if the claim of defining the optimal dose of training will be proper for this study design.

Response: Thank you very much for this suggestion. We agree, the best way to determine the perfect combination of WBV determinants would be to test a specific WBV setting for several weeks. However, we would like to raise the following aspects that convinced the authors of this study to distinguish an appropriate dose in a cross-sectional setting:

1. The design you proposed to identify the individually tailored WBV training regime would take years and would require the subjects (sample of particular cancer patients under a particular cancer treatment) to keep the disease stage and treatment (i.e. chemotherapy, medication, radiation) over months and years, as well as require them to drive to our site twice a wekk over a long persiod of time, which would further diminish compliance. To our experience (Streckmann et al. 2014; 2018) this is not feasible for patients in the context of cancer therapy.

2. Therefore, we would like to highlight the following aspects:

It is well known that the electromyogram (EMG) of musclse is related to the muscles force production and thus, to the level of contraction. In particular, during a voluntary action, the muscles' EMG activity is related to the extent of the muscle fibers' recruitment (Aagaard 2003 ; Hogrel 2003 ; Milner-Brown et al. 1973a , 1973b ; Riley et al. 2008) and frequency (Aagaard 2003 ; Milner-Brown et al. 1973a). A high EMG activity compared to a low EMG activity results from a higher number of recruited muscle fibers and higher motor unit discharge frequencies (Moritani and Muro 1987) and is accompanied by higher forces generated by the target muscle (Freund et al. 1975 ; Moritani and Muro 1987).

Hence, based on these relations, the EMG activity could be used to easily determine the activation intensity of the muscle in a given set of WBV treatment at least within limits mentioned as cross talk and nonlinearity (Farina et al. 2004 ; Keenan et al. 2005). Within those limitations, the EMG activity can serve as an adequate predictor to estimate the muscles' activation intensity. The limitations will of course be dicsussed adequatly in the final result paper.

3. Designing a WBV-based training regimen with the objective to achieve a distinct type of adaptation (e.g., improved power generated by a specific muscle group) requires a thorough understanding of the influence of the training setup and the selected WBV parameters (Cochran 2011 ; Rittweger 2010). Based on WBV training studies executed in samples of healthy subjects we know that the thigh and shank muscles are similarly affected (Ritzmann et a. 2012, Abercromby et al. 2007). The rise in EMG activity (i.e. the integral and root mean square) with increasing WBV frequency or amplitude is manifested in the plantarfexors, dorsiflexors, knee extensors and flexors. Thus, we are convinced that EMG recordings in a cross-sectional study design can serve as an appropriate alternative to several longitudinal studies to distinguish an appropriate and feasible WBV training setting for cancer patients.

Response: We agree however, that „optimal training setting" may not be the appropriate term and have rephrased it (abstract p2; background p.6; discussion p.12)

Minor concerns:

> It might be better if authors talk about the FACT-GOG-Ntx earlier since this factor is the primary outcome of this study.

Response: Thank you for this advice, though we are not sure where to place it any earlier in the text as it appears immediately as primary endpoint in the abstract as well as as screening tool in the methods and design section. We are open to any suggestions where it could be placed and will gladly change this.

> In the strength section, the authors claimed that there is not any treatment option for CIPN. Therefore, this statement brings the idea that the proposed intervention can be used as a treatment for these type of patients. By knowing that there are plenty of studies showing the effectiveness of exercise to improve these symptoms, the authors should either count the exercise as a treatment or not including WBV as a treatment option.

Response: To our knowledge, there are currently no treatment options for motor AND sensory symptoms of CIPN. There are 5 studies, investigating the positive effects of exercise on CIPN, only 2 of which have investigated WBV. THere are a few studies reporting on the effects of WBV on other entities. Though the reporting on the precise settings of WBV is inconsistent or lacking entirely. We therefore would like to investigate the effects oft he various setting in ordert o be able to make suggestions for future research and training as we believe that WBV could serve as a treatment option, though we need more consistency in reporting and more evidence.

> As mentioned in the background, reference 18 shows the effectiveness of WBV for CIPN patients with 11 subjects. Therefore, it would be better if authors talk more about the difference between their study and reference 18. Only mentioning "objective and systematic" does not show the difference of this study from the mentioned study. In addition, the terminology of "systematic" is unclear to me.

Response: We appreciate your comment. Regarding the difference between the studies and the intention of this stduy, we have added a paragraph (p.5) and hope to have made it more apparent now.

> You are conducting several approaches to assess the neuropathy. Several validated methods will be used in this study and I am not sure if it is necessary to measure all of these parameters.

Response: We felt it is more appropriate to assess several outcome parameters which assess different nerve fiber functions. As mentioned in the text it is difficult and too little is known mechanistically in order to determine CIPN with only one assessment tool. Not all parameters are tested every single time. Elektrophysiology is only perfomed once at the beginning while the clincial short test is performed at every time point but only requests 5mins time. All in all the test require 30mins. We have added this information in the text (Experimental design p.8)

and hope to clarified somewhat?

> Based on the large number of variables you are analyzing, a comment on why you are, or are not, going to use Bonferroni correction (or similar) to reduce the chances of obtaining false-positive results would be appropriate.

Response: We use the exact rANOVA p value for the interaction term as a measure of strength of evidence against the null hypothesis. We do not rely on some arbitrary thresholds. We additionally calculate the point estimates with confidence intervals for the pre-post and pre-follow-up changes scores. We are convinced that this approach is appropriate and in line with current statistical guidelines (Sterne and Davey Smith, BMJ 2001; Greenland et al., Eur J Epidemiol 2016). The revised analysis paragraph starts as follows:

“As primary analysis we will calculate two-factorial repeated measures analyses of variances (rANOVA; factor 1: control vs. intervention group; factor 2: time point of measurement; pre, post, follow-up) to assess the time course of adaptations between groups. The main effect of interest is the rANOVA interaction term between both factors (Hecksteden et al., Front Physiol 2018). We use the exact p value as a measure of the strength of evidence against the null hypothesis, i.e. that there is no intervention effect (Sterne and Davey Smith, BMJ 2001). In order to arrive at an estimation of the size of the observed effects, we calculate the change scores from pre to post as well as pre to follow-up tests together with 90% confidence intervals (Sterne and Davey Smith, BMJ 2001).”

> I have not seen any place talking about the adjustment in the statistics section. Are you going to report the results without any adjustment or estimated variable will be reported based on the adjustment? Since subjects will be between 18-80 years old, the chance of non-normal distribution for this wide range of age is high. So, defining a method to use adjustment is necessary for this section.

Response: Thank you. This is a relevant point. We use age as a stratum in our analysis. Minimisation will allow for the number of participants in each study arm being closely balanced within each stratum (Hecksteden et al., Front Physiol 2018), i.e. for age in our sample. Minimisation is particularly advantageous when small groups should be closely matched with regard to relevant potential confounders (Treasure and MacRae, BMJ 1998). Minimisation may not eliminate bias on all known and, particularly, unknown confounders, but we are convinced that our choice and number of strata (age, neurotoxic agent, study centre) is appropriate in our particular setting. We discussed the issue of controlling for potential confounders in the statistical model. This would, however, have led to an unrealistically large sample size. Therefore, we decided to control for potential confounding at the group allocation stage. We are aware that this approach is not ideal, but as we calculate the point estimates for the change scores with confidence intervals and the groups are likely well balanced (particularly with regard to age), we are convinced that this is an appropriate compromise.

> Is there any reason to use multiple imputations for missing data? Linear mixed models are a valid method used for missing data and allow for testing group difference.

Response: Thank you for this comment. We a priori decided on the statistical model we apply in our study (see comment above). We are convinced that this approach is appropriate and within current consensus (Hecksteden et al., Front Physiol 2018). Although we conduct all tests with great care, it can be expected that there will be some missing data (we are convinced that we can keep the numbers very small). We cannot exclude that some missing data occur non-randomly and therefore a complete-case-analysis may lead to bias results. As we are, however, aware of potential pitfalls with multiple imputation, we conduct, in line with the suggestion of another reviewer, a sensitivity analysis by comparing the results after imputation with the complete-case results.

> I have not seen any place talking about the trial registry of this study in Clinicaltrial.gov. It might be good to mention the trial registry on Clinicaltrial.gov.

Response: Thank you for bringing this to our attention. In addition to the mentioning below the abstract, we have also added the information in the main part: (Experimental design, p.7)

> Suggestion

> As authors mentioned in the background, exercise is counted as one of the main treatment to improve the CIPN symptoms. The study can be more valuable if authors use exercise for the control group. Since the previous studies have shown the effectiveness of WBV, comparing the WBV group with no intervention group might not add many impacts on this field.

Response: Thank you for this suggestion. However, the effectiveness for WBV on symptoms of CIPN has only been shown in one pilot study (Streckmann et al. 2018). The other study (Schönsteiner et al.) had a multimodal approach and don't report on the precise setting, making it difficult to compare and create evidence. As we know from sensorimotor training for instance, the setting hast o be between 20-40sec. Or it will not have the desired effect. We can only assume that the settings for WBV could be equally sensitive and require a precise combination tob e most effective. It could be possible, that patients with CIPN reach the highest effect on a setting that may not be feasible for these patients due to pain for instance or that there is a setting that ist he most appropriate. To date the pilot study used settings between 18-30Hz, 2 or 4 mm and 60sec. Aim of this study ist o determine the effect of the various settings and the impact during training. We have tried to clarify this (Strength section, p.3)

As the mechanistics and effects of specific exercise intensities remain tob e elucidated, we are not sure which exercise would not influence the results and therfore chose not to exercise with the control group immediately, but offer the training tot hem after completion of the study.

> As you might now, there are 3 senses that contribute to balance: vestibule, vision, and peripheral sensation. So, having an eyes closed test would be a great challengeable balance test to measure the improvement of sensation for these patients with neuropathy. Is there any reason that you have not included this test to the study design?

Response: You are absolutely right. Additionally, Kneis et al. 2017, showed the importance and grading of the various conditions for patients with CIPN. We therfore always include the eyes closed condition in our balance test. Sorry for not clarifying this. We only mentioned our balance test by standardized protocol and usually describe the precise protocol in the result paper. We have added an example in the method section under balance test (p.10) and hope it is sufficient at this point?

Reviewer: 3

> Reviewer Name: Andrew Hinde

> Institution and Country: University of Southampton

> Please state any competing interests or state ‘None declared’: None declared.

> This description of a study protocol is largely convincing. My main concern relates to the sample size. While you are correct that, if you are anticipating a medium effect size, a sample of 36 is required to identify it, I think that to assume that recruiting 44 participants will give you 36 valid cases may be a little optimistic. To the extent that effects are less than 'medium' for some secondary outcomes, you are likely to find that your study is underpowered. Therefore I would be inclined EITHER to strengthen the fourth bullet point on page 3 to suggest that 'some secondary endpoints are likely to be underpowered' OR to try to raise the sample size recruited to, say, 50, to give you a bit of extra leeway. It may be too late to increase the sample size, of course, or it may be difficult. But you are cutting things a bit fine with a recruitment of 44.

Response: Thank you. We partly agree. However, the expected 20% drop out rate is based on experiences from similar prior studies. We changed the fourth bullet point as follows:

“Due to the estimated sample size for the primary outcome, it is possible that the analysis of some secondary endpoints may be underpowered.”

> On the question of effect size, you cite f = 0.25 (p. 12, l. 9). But I think you mean f = 0.75. I have checked the calculations and a value of f = 0.75 is consistent with a sample size of 36.

Response: We did the sample size estimation with G*power. An f value of 0.25 corresponds to a partial eta squared (the common effect size measure for an rANOVA) of 0.06, i.e. a medium effect size according to common guidelines. We are convinced that our sample size estimation is appropriate. Here is the corresponding protocol:

[1] -- Thursday, October 25, 2018 -- 09:20:52

F tests - ANOVA: Repeated measures, within-between interaction

Analysis: A priori: Compute required sample size

Input: Effect size f = 0.25

α err prob = 0.05

Power (1-β err prob) = 0.90

Number of groups = 2

Number of measurements = 3

Corr among rep measures = 0.5

Nonsphericity correction ε = 1

Output: Noncentrality parameter λ = 13.5000000

Critical F = 3.1316720

Numerator df = 2.0000000

Denominator df = 68.0000000

Total sample size = 36

Actual power = 0.9060304

> On a related point, with 44 cases (reduced by attrition to 36) you should be able to identify whether the whole-body vibration (WBV) training has a significant impact on the symptoms. But I think that establishing the 'optimal dose for training recommendations' (p. 6, l. 12) is optimistic. For a start, on p. 10, ll. 23-27 you suggest that the optimal setting (which is partly defined on the basis of dose) may be specific to each individual. If you find this, then the idea of an 'optimal dose for training recommendations ... with regard to future clinical trials' does not seem very relevant. You might, of course, find that a setting with the same dose produces the highest neuromuscular response in most, or all, patients, but this will not be known until the initial training sessions are completed.

Response: We agree with the reviewer that the study is not designed to establish an optimal training dosage. This is an exploratory part of our study. But we think, it is from a clinical and practical perspective a relevant question. We are convinced that our results will allow for a careful first recommendation on how to apply whole-body vibration training exactly in neuropathic patients. We revised the corresponding passages in the manuscript as follows:

“(d) detect an appropriate and feasible dose for training recommendations”

> On p. 8, ll. 16-26 you describe the assessment procedures. I think clarification would be useful here. How many investigators do you have? And what do you mean by assessments being 'aligned among the investigators' (l. 17)? Later on, you say that '[e]xaminers will be trained by a gold-standard examiner' (l. 22), and refer to 'testers' (l. 25). So we have investigators, examiners, a 'gold-standard examiner' and testers. And these people are doing not investigations, or examinations, or tests, but assessments. I am afraid that I was quite confused by the time I reached l. 26. Why not take a bit more space and explain more slowly how you plan to organise the assessments?

Response: Thank you for this comment and we are sorry to have caused such confusion. We hope to have clarified this now and have added more detailed paragraph (p.8, assessment procedures)

> p. 10, ll. 23-6. 'After a short familiarization with the training setting ... conditions of WBV within the first two training sessions'. Is this training done on all patients, or only those randomised to the intervention group? The statement on p. 11, ll. 9-11 suggests that it is done on all patients before randomisation. Can you clarify this?

Thank you for bringing this to our attention and we regret that this was not stated clearly enough. We have added, that only the intervention group performed this and hopw to have clarified it by doing so.

> Minor points and typographical errors

> p. 5, l. 3 Delete comma after 'parameters'. (We have deleted it)

> p. 6, l. 3 Change 'behold' to 'hold'. (changed)

> p. 6, l. 25 Change 'oft he' to 'of the'; and change 'publictaion' to 'publication'. (changed)

> p. 13, l. 5 Delete comma after 'fact'. (has been deleted)

> p. 13, l. 9 Define 'SENIAM'. (we have added the definition)

Reviewer: 4

> Reviewer Name: Yoon Kong Loke

> Institution and Country: University of East Anglia

> Please state any competing interests or state ‘None declared’: No competing interests.

> Please clarify the relationship of the pilot study to this one, and whether the data will be pooled. Are the data from those participants being included in this trial? If this is the case, then this protocol should state that it is the continuation of the pilot study as a full-scale RCT.

Thank you for this comment. Unfortunately, we are not quite sure which pilot study you are reffering to as this type of study has never been done. We intend to perfom a larger RCT, involving the results of this study in future, as to date, there is no consistent reporting on the training settings of WBV and it is therfore also not evaluated how patients with CIPN should train in future studies.

> Please state in abstract and early on in the Methods that this is outcome assessor blinded design.

Response: Thank you for bringing this to our attention. We have changed single-blind to assessor blinded design.

> Please also provide a justification for this design rather than double blind - was there no possibility of a sham procedure?

Response: Thank you for this suggestion. Unfortunately in this type of intervention sham procedure is not possible. To date, the effect of WBV to any extent in these patients is not yet evaluated sufficiently. We thefore cannot outrule, that even low frequency settings for instance would not ahve an affect. We furthermore believe that vibration plates are to well known menawhile in order to place patients on them without switching them on, which is why we chose to do it with a control group and make a larger effort to maintain all other components as clear as possible (assessors do not know the randomization result, trainers have no contact to assessors and patients are only informed themselves after the first assessment in order not to give away any information accidentaly.

> Multiple imputation may or may not be reliable for the outcome data that is missing. I suggest that sensitivity analysis should be carried out with, and without, the multiple imputation.

Response: Thank you very much. This is a very good point. We will conduct the suggested sensitivity analysis and revised the paper as follows:

“We will also conduct a sensitivity analysis by comparing the imputed results with the complete-case results.”

> How will you consider multiple testing of significance - there are lots of outcomes and time points being tested?

Response: We use the exact rANOVA p value for the interaction term as a measure of strength of evidence against the null hypothesis. We do not rely on some arbitrary thresholds. We additionally calculate the point estimates with confidence intervals for the pre-post and pre-follow-up changes scores. We are convinced that this approach is appropriate and in line with current statistical guidelines (Sterne and Davey Smith, BMJ 2001; Greenland et al., Eur J Epidemiol 2016). The revised analysis paragraph starts as follows:

“As primary analysis we will calculate two-factorial repeated measures analyses of variances (rANOVA; factor 1: control vs. intervention group; factor 2: time point of measurement; pre, post, follow-up) to assess the time course of adaptations between groups.

The main effect of interest is the rANOVA interaction term between both factors (Hecksteden et al., Front Physiol 2018). We use the exact p value as a measure of the strength of evidence against the null hypothesis, i.e. that there is no intervention effect (Sterne and Davey Smith, BMJ 2001). In order to arrive at an estimation of the size of the observed effects, we calculate the change scores from pre to post as well as pre to follow-up tests together with 90% confidence intervals (Sterne and Davey Smith, BMJ 2001)."

> For the categorical or dichotomous outcomes - are these being analysed using ANOVA as well?

Response: Thank you, but as the TNS score (primary endpoint) results in an overall score, we don't have any dichotomous variables and will therefore be performing parametric testing.

VERSION 2 – REVIEW

REVIEWER	Reviewer name: Mohsen Zahiri Institution and Country: Baylor College of Medicine Competing interests: None declared
REVIEW RETURNED	25-Nov-2018

GENERAL COMMENTS	The answers were convincing and the authors clarify some parts that caused confusion. The manuscript can be considered as accepted from my point.

REVIEWER	Reviewer name: Andrew Hinde Institution and Country: University of Southampton Competing interests: None declared
REVIEW RETURNED	14-Nov-2018

GENERAL COMMENTS	Thank you for your response to my comments. The sections that needed clarification are all now much clearer. I am happy with your calculations of sample size and I hope that, in the event, the numbers you recruit will prove sufficient to establish the results you seek. I wish you well with your study.