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)	Study protocol of a randomized control trial in subacute patients with visuospatial neglect: Is Congruent Movement Training more effective than standard Visual Scanning Therapy to ameliorate symptoms of neglect?
AUTHORS	Elshout, J.A.; Nijboer, T.C.W.; van der Stigchel, S.

VERSION 1 – REVIEW

REVIEWER	Arnaud Saj Psychology Department, University of Montréal, QC, Canadna
REVIEW RETURNED	11-Jun-2019

GENERAL COMMENTS	Van der Stigchel's team show an interesting project on the remapping combining the visual an motor effect in spatial neglect. This is an original project wich respect the special issue, taking into account the cited literature. The manuscript is well-written and the results expected will be interessting. There are some concerns, in particularly with the population. By resolving these concerns, the quality of the project will be improved for publication. Detailed comments are given below: 1) Patients are with right and left hemispheric stroke; 2) The neuropsychological VSN tests are light (it miss a lot of tests); why 'and/or' for the CBS in the text; it misses the motor neglect assessment; 3) The limit of age should be 75-80 years and have a cognitive global measure (MOCA but if the neglect limited the score but used the MOCA for the visual disorders); 4) Why patient group patient control and healthy control subject, if the experimental protocol is AB-BA for the two groups of neglect, it's better; 5) In patients group, the authors will have to take into account the presence of frontal or parietal lesion (Saj et al., 2018). I look forward to seeing the paper!

REVIEWER	Kimberly Hreha UTMB, USA
REVIEW RETURNED	05-Jul-2019

GENERAL COMMENTS	I am not sure if it is appropriate to publish a study that hasn't been carried out yet. However, if that is OK with the editor, then I would just suggest the following minor edits: Line 104 – I think the authors meant to write "visuospatial neglect (VSN)".

	Line 232- I would ask that the authors be more specifics regarding the outcome measures. For example: what do they mean by “among others”? Lines 254-257- I don’t understand this paragraph. The title is patient and public involvement however then the authors write that the patients and public will not be directly involved. Was this a typo? There are many reports now on the importance of patient engagement prior to the study design. Is this what the authors are alluding to? At what point in time will the assessments be completed? This information would be helpful to add to the methodology.

REVIEWER	Teresa Neeman Australian National University Canberra, ACT, Australia
REVIEW RETURNED	09-Aug-2019

GENERAL COMMENTS	Suggested improvements to study design: (a) allocation concealment: the protocol should describe how patient selection and screening will be carried out without knowledge of the expected treatment assignment. (b) evaluator blinding: the protocol should describe which evaluators will be blinded. For example, the study nurse assessing ADL should be blinded to treatment. There may be other tests where the evaluator could potentially bias the outcome by knowledge of the treatment assignment. These situations should be considered and blinding plans discussed in the protocol. (c) covariate adjustment or stratification: The authors plan to adjust the analysis by level of education. If education level is the key confounder, they should be more specific as to what levels will be differentiated. Also, if education is an important confounder, the authors may consider a stratified randomised design, ie randomising within education levels, to get a more precise measure of the treatment effect (compared with post-stratifying). (d) On the other hand, it is more usual that the baseline VSN is an important predictor of improvement. That is, patients with high levels of neglect may be more resistant to therapy than patients with lower levels of neglect. Conversely, patients with very low levels of neglect may show little improvements due to ceiling effects. If these are relevant issues, then the authors may consider a stratified randomised design, stratifying (instead of (c)) by baseline VSN level, to ensure that treatments are compared within patients with similar baseline VSN. (e) Collecting baseline of control subjects: I understand the rationale for wanting to have negative and positive baseline controls, but I don't think they will end up being useful in the statistical analysis. The healthy controls will be problematic because it's not clear if they represent the "normal" version of the patients on the study. For example, if they are from a different age group, then any differences may be attributed to age. It's also not clear what the purpose of the patient control group is, since patients will be their own controls (post - pre). There is nothing in the analysis plan about how these data will be incorporated to address the research question. The authors should think carefully why these data will be informative. I doubt that they will be, but if they disagree, they should describe their proposed analysis in the analysis plan.

	(f) It would be very helpful for both the authors and the readers if they included a study schedule, see e.g. Figure 1 in https://www.bmj.com/content/346/bmj.e7586 . One could then see in a single table the set of tests and assessments at each timepoint in the study. (g) sample size calculation: More consideration needs to be included in the sample size calculation. The authors should present (i) mean improvement observed from other studies, and (ii) postulate what additional improvement they expect to observe in this study. (iii) There should also be some assessment of between-patient variation. Simply using a Cohen's D of 1 obscures the important thinking that is needed by a good sample size assessment.

REVIEWER	Zhiying You University of Colorado Denver Anschutz Medical Campus Aurora, CO 80045
REVIEW RETURNED	09-Aug-2019

GENERAL COMMENTS	Overall, the manuscript clearly describes a randomized controlled clinical trial. However, there is some space for further improvement as follows. 1. Please specify the start date and end date of the study. 2. Please provide justifications why the patient control group (without VSN) and healthy control group were included. 3. Lines 205-216: While more than one primary outcome was specified, please specify how the study will consider these outcomes. In other words, please specify they are co-primary outcomes or multiple primary outcomes. The FDA has provided guideline on this topic because of the complex when involving multiple primary outcome. 4. Lines 243-244: All covariates to be adjusted for in data analysis should be specified before looking as any data. It should not be determined after exploring any collected data from the study.

VERSION 1 – AUTHOR RESPONSE

Response to reviewers:

Reviewer 1:

1) Patients are with right and left hemispheric stroke;

Response: The reviewer is correct that we include both left and right hemispheric stroke. We have now mentioned this more explicit in the method section with subheading “Subjects”.

VSN patients are included (in chronological order) if they (1) are clinical diagnosed with symptomatic stroke (left or right ischemic or intracerebral haemorrhagic lesion), (2) show signs of VSN based on one of the neuropsychological VSN tests (shape cancellation, line bisection, Catherine Bergego Scale), (3) are between 18-85 years of age, (4) have sufficient comprehension and communication, (5) have sufficient motivation to participate and (6) give written informed consent.

2) The neuropsychological VSN tests are light (it miss a lot of tests); why 'and/or' for the CBS in the text; it misses the motor neglect assessment;

Response: We use the neglect tests and observation scale that are currently incorporated in the rehabilitation centre (i.e. shape cancellation tests, line bi-section test and Catherine Bergego Scale). Motor neglect can be assessed with the Catherine Bergego Scale. The omission difference score of a shape cancellation test is the most sensitive test to indicate an attentional imbalance (Ferber & Karnath, 2001; Husain and Rorden, 2003). Nevertheless, we will include patients if at least one of these tests indicates neglect (which was indicated by the 'and/or'). We have now rewritten this sentence to clarify this in the text.

VSN patients are included (in chronological order) if they (1) are clinical diagnosed with symptomatic stroke (left or right ischemic or intracerebral haemorrhagic lesion), (2) show signs of VSN based on one of the neuropsychological VSN tests (shape cancellation, line bisection, Catherine Bergego Scale), (3) are between 18-85 years of age, (4) have sufficient comprehension and communication, (5) have sufficient motivation to participate and (6) give written informed consent.

3) The limit of age should be 75-80 years and have a cognitive global measure (MOCA but if the neglect limited the score but used the MOCA for the visual disorders);

Response: We are not sure why the age should be limited to 75-80. There are also some older (>80) still fit and active patients in the rehabilitation centres where we conduct the study.
As mentioned in the paragraph "demographical and stroke related parameters" we will collect MOCA scores for each patient (line 202). The reviewer is right that we should take the MOCA scores into account. Therefore, we will compare the MOCA scores of the two groups and will correct for that if the scores are different between groups. As patients are randomly assigned to each group we expect an equal distribution between groups.

4) Why patient group patient control and healthy control subject, if the experimental protocol is AB-BA for the two groups of neglect, it's better;

Response: We thank the reviewer for raising these important points. We now made more explicit why we include the patient control and healthy control groups (Method section subheading "Baseline performances")

Since we also include new (secondary) outcome measures and tests (eye-tracking, cookie theft picture for free exploration, visual discrimination task, virtual reality), we need to compare the data of the neglect patients to stroke patients without neglect and healthy control to study whether baseline performance 1) deviates from normal range 2) is stroke or neglect specific. In addition, the control data allows us to examine whether performance after training change to values that can be considered as normal on these tasks.

We also agree with the reviewer that an AB-BA is a good design to test our research question. However, in the current setting (patients in the subacute phase admitted to the rehabilitation centre) it is impossible to use an AB-BA design, as most patients will be discharged from the rehabilitation centre before they have finished the study protocol. We have added this to the method section in the subheading "Design".

Since patients are admitted to the rehabilitation centre, which aim to discharge patients from the rehabilitation centre as soon as possible, we chose to compare two independent groups (A vs B) rather than a cross-over design (AB-BA), which will have led to a large dropout.

5) In patients group, the authors will have to take into account the presence of frontal or parietal lesion (Saj et al., 2018).

Response: We thank the reviewer for this suggestion. We will use all information available (MR scan, medical record) to describe each lesion in detail. We will first perform the group analysis (patients are randomly assigned to each group), but will check for individual differences in lesions location. Based on the reviewers' suggestion (and paper) we will be specifically keen to check whether the performance of patients with frontal lesions (if we have these patients in our sample) are different from patients with no frontal lesions.

Reviewer 2:

1. Line 104 – I think the authors meant to write "visuospatial neglect (VSN)".

Response: We thank the reviewer for finding this typo and have rewritten the sentence accordingly.

These results in healthy controls suggest new approaches to treatment for patients with asymmetric attentional deficits such as VSN.

2. Line 232- I would ask that the authors be more specifics regarding the outcome measures. For example: what do they mean by "among others"?

Response: There is a lot of data available for each VR session. As this is a more explorative test we gave three specific examples that we could analyse and leave room for additional analyses by adding "among other". However, as we have not planned additional analyses yet, we have removed "among others".

3. Lines 254-257- I don't understand this paragraph. The title is patient and public involvement however then the authors write that the patients and public will not be directly involved. Was this a typo? There are many reports now on the importance of patient engagement prior to the study design. Is this what the authors are alluding to?

Response: BMJ open publish study protocols and we follow their manuscript formatting policies. It is on the journals request to include a paragraph under subheading "patient and public involvement" even if no patients were involved in the design of the study.

4. At what point in time will the assessments be completed? This information would be helpful to add to the methodology.

Response: We thank the reviewer for raising this point. We have added the planned end date of the study to the method section.

Patients will be recruited in De Hoogstraat Rehabilitation and De Parkgraaf Rehabilitation between May 2018 and January 2020.

Reviewer 3:

1. allocation concealment: the protocol should describe how patient selection and screening will be carried out without knowledge of the expected treatment assignment.

2. evaluator blinding: the protocol should describe which evaluators will be blinded. For example, the study nurse assessing ADL should be blinded to treatment. There may be other tests where the evaluator could potentially bias the outcome by knowledge of the treatment assignment. These situations should be considered and blinding plans discussed in the protocol.

Response to 1 and 2: We thank the reviewer for raising these important two points. We now mention more explicit how the selection process take place and have added the blinding procedure.

At the time of the neuropsychological screening to assess whether a patient has VSN, it is still unknown whether a patient can be included in the study and which treatment he/she will receive if he/she can be included. After inclusion, in chronological order, VSN patients will be randomly assigned to one of two groups, based on a predetermined list generated using Matlab, that has paired one of the training variants to a participant number. Patients are not explicitly informed about the nature of their treatment (i.e. whether they receive the experimental CMT or control VST treatment). The nurses who assess the CBS are blinded to training variant. The researchers who administer the training and the tests are not blinded, since they have to explain the training and tests. To minimize any (unintended) bias, the same task instructions are read aloud for each patient before each test. All data is collected and saved automatically by the computer program so that no changes can be made after task completion.

3. covariate adjustment or stratification: The authors plan to adjust the analysis by level of education. If education level is the key confounder, they should be more specific as to what levels will be differentiated. Also, if education is an important confounder, the authors may consider a stratified randomised design, ie randomising within education levels, to get a more precise measure of the treatment effect (compared with post-stratifying).

Response: We agree with the reviewer that a stratified randomised design would have been a useful design. However, our main question is not to study what the effect of education level is on these training paradigms. In addition, since we do not know beforehand what the variability in the level of education will be in our patient group, there might be difficulties by filling each group within the time window we have to conduct this study. Also, we then have to decide how many groups we want to compare and which cut-off scores to use. This is beyond the scope of our study. Therefore, we have not chosen for a stratified design but will correct for level of education in the main analysis. We will use the Verhage scores (Verhage, 1964) to indicate level of education. We have added his specification to the manuscript.

Level of education (Verhage score) will be added as covariate.

4. On the other hand, it is more usual that the baseline VSN is an important predictor of improvement. That is, patients with high levels of neglect may be more resistant to therapy than patients with lower levels of neglect. Conversely, patients with very low levels of neglect may show little improvements due to ceiling effects. If these are relevant issues, then the authors may consider a stratified randomised design, stratifying (instead of (c)) by baseline VSN level, to ensure that treatments are compared within patients with similar baseline VSN.

Response: We agree with the reviewer that severity of neglect might be an important predictor of improvement. Yet, patients are randomly assigned to one of the two groups, so we expect that both

groups will be comparable in terms of severity of neglect. A difference in severity of neglect between the two groups will be based on chance. However, if there is a difference in severity of neglect, we will correct for that to address our main research question.
As this is a good suggestion of the reviewer, we will conduct additional analysis to relate severity of neglect with training potential in general. We have added this to the manuscript.
With respect to the stratified design, we refer to our response to point three.

Patients are randomly assigned to one of the two groups, so we expect that both groups will be comparable in terms of severity of neglect. A difference in severity of neglect between the two groups will be based on chance. However, if there is a difference in severity of neglect, we will correct for that to address our main research question. In addition, we will conduct additional analysis to relate severity of neglect with training potential in general.

5. Collecting baseline of control subjects: I understand the rationale for wanting to have negative and positive baseline controls, but I don't think they will end up being useful in the statistical analysis. The healthy controls will be problematic because it's not clear if they represent the "normal" version of the patients on the study. For example, if they are from a different age group, then any differences may be attributed to age. It's also not clear what the purpose of the patient control group is, since patients will be their own controls (post - pre). There is nothing in the analysis plan about how these data will be incorporated to address the research question. The authors should think carefully why these data will be informative. I doubt that they will be, but if they disagree, they should describe their proposed analysis in the analysis plan.

Response: We now made more explicit why we include the patient control and healthy control groups (Method section subheading “Baseline performances”). As mentioned in this section, the control groups will be age-matched.
For example, we also collect eye tracking data during the star cancellation task (primary outcome measure) to study the eye movement patterns (e.g. number of fixations right vs. left). To know whether the eye movement patterns of the neglect patients during this task deviate from normal range (with or without stroke), we need the patient and healthy control groups.

Since we also include new (secondary) outcome measures and tests (eye-tracking data, cookie theft picture for free exploration, visual discrimination task, virtual reality), we need to compare the data of the neglect patients to stroke patients without neglect and healthy control to study whether baseline performance 1) deviates from normal range 2) is stroke or neglect specific. In addition, the control data allows us to examine whether performance after training change to values that can be considered as normal on these tasks.
Therefore, only baseline measurements will be collected for a patient control group and healthy control group and these two groups will not receive any training. Performance on all outcome measures of both VSN intervention groups will be compared to the performance of a patient control group (stroke patients without VSN, n=15) and age-matched healthy controls (n=15).

Also, we now mention more explicit that we conduct the same analysis for the secondary outcome measures and added the analysis with the comparisons to the control groups.

Performance on all primary and secondary outcome measures will be compared.

To explore the performance on the new outcome measures (eye tracking data: number of fixations, fixation duration, direction of first saccade) and tasks (cookie theft and Virtual Supermarket) we compare baseline performances of the neglect patients to the patient and healthy control groups using parametric t-tests, or non-parametric tests in case of non-normal distributed data. If these data

are deviating from the control groups we conduct additional analyses, similar as performed for the primary outcome measures (Repeated Measures Analyses (ANCOVA), with session (baseline-post training) as within-subject variable and treatment (CMT, VST) as between-subject variable), to test whether these new outcome measures improve more during CMT than VST training.

6. It would be very helpful for both the authors and the readers if they included a study schedule, see e.g. Figure 1 in https://www.bmj.com/content/346/bmj.e7586 . One could then see in a single table the set of tests and assessments at each timepoint in the study.

Response: We thank the reviewer for this suggestion. We have added the table below to the manuscript.

	STUDY PERIOD												
	Enrollment	Baseline	Training										Evaluation
TIMEPOINT	$-t_1$	t_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	t_1
ENROLLMENT:													
Eligibility assessment	X												
Written Informed Consent	X												
Allocation	X												
INTERVENTIONS (neglect):													
Visual Scanning Therapy		X	X	—	—	—	—	—	—	—	—	X	X
Congruent Movement Training		X	X	—	—	—	—	—	—	—	—	X	X
CONTROL GROUPS:													
Patiënt Control Group		X											
Healthy Control Group		X											
ASSESSMENTS:													
Shape Cancellation Task		X											X
Line Bi-section Task		X											X
Catherine Bergego Scale		X											X
Visual Discrimination Task		X											X
Visual Exploration Task		X											X
Virtual Supermarket (VR)		X											X
Perimetry Test		X											X

7. sample size calculation: More consideration needs to be included in the sample size calculation. The authors should present (i) mean improvement observed from other studies, and (ii) postulate what additional improvement they expect to observe in this study. (iii) There should also be some assessment of between-patient variation. Simply using a Cohen's D of 1 obscures the important thinking that is needed by a good sample size assessment.

Response: Effect sizes reported in literature vary considerably among studies, ranging from 0 to 2.84, with the highest effect size for combined training methods. As mentioned in the sample size calculation (line 251-257), we based a Cohen's D of 1 on the study by Polanowska who compared somatosensory stimulation + visual scanning training to visual scanning training alone showing an effect size of 1.58. This study design and training paradigm is very comparable to our design and training. While their amount of training was higher we are more conservative and chose a Cohen's D of 1, which seems plausible to us.

We have rewritten our paragraph about sample size calculation and added means (SDs) of two comparable studies:

We performed an a priori power analysis (G-Power 3.1) to calculate the sample size. Effect sizes reported in literature vary considerably among studies, ranging from 0 to 2.84, with the highest effect size for combined training methods. A study by Polanowska et al compared somatosensory stimulation + visual scanning training to visual scanning training alone (mean improvement experimental group = 58.4 (20.6) vs control group =17,35 (30,3)) showing an effect size of 1.58. Another study by Schroder et al (2008) treated neglect patients with TENS or OKS compared to a control group that received VST. This study reported an effect size of 0.83 (TENS) and 1.56 (OKS) (mean improvement experimental group TENS= 1,298 (1,23), experimental group OKS= 1,938 (0,89) vs control group =0,264 (1,25)).
While both study designs and training paradigms are very comparable to our design and training, their amount of training was a bit higher. Therefore, we are more conservative and chose a Cohen's D of 1. A power analysis with power set to 0.8 and alpha set to 0.05 estimated the sample size at 14 patients per group (28 in total) for sufficient statistical power.

Reviewer 4:

1. Please specify the start date and end date of the study.

Response: We thank the reviewer for raising this point. We have added the planned end date of the study to the method section.

Patients will be recruited in De Hoogstraat Rehabilitation and De Parkgraaf Rehabilitation between May 2018 and January 2020.

2. Please provide justifications why the patient control group (without VSN) and healthy control group were included.

Response: We now made more explicit why we include the patient control and healthy control groups (Method section subheading "Baseline performances")

Since we also include new (secondary) outcome measures and tests (eye-tracking data, cookie theft picture for free exploration, visual discrimination task, virtual reality), we need to compare the data of the neglect patients to stroke patients without neglect and healthy control to study whether baseline performance 1) deviates from normal range 2) is stroke or neglect specific. In addition, the control data allows us to examine whether performance after training change to values that can be considered as normal on these tasks.

3. Lines 205-216: While more than one primary outcome was specified, please specify how the study will consider these outcomes. In other words, please specify they are co-primary outcomes or multiple primary outcomes. The FDA has provided guideline on this topic because of the complex when involving multiple primary outcome.

Response: We thank the reviewer for the suggestion to follow the FDA guideline. We have 4 separate primary outcome measures: change in performance (post training - baseline) on 1) shape cancellation (omission difference left vs right), 2) line bisection (deviation from centre), 3) visual discrimination task (% correct) and 4) Catherine Bergego Scale (score). We will use the False Discovery Rate (FDR) approach to correct for multiple comparisons. We have added this to the method section.

The False Discovery Rate (FDR) approach will be used to correct for multiple comparisons.

4. Lines 243-244: All covariates to be adjusted for in data analysis should be specified before looking as any data. It should not be determined after exploring any collected data from the study.

Response: Patients are randomly assigned to one of the two groups. Therefore, we expect that both groups will be comparable in terms of severity of neglect and the demographical and stroke related parameters (defined in the paragraph “demographical and stroke related parameters”). Any difference between one of these demographical or stroke related parameters or severity of neglect will be based on chance. However, if there is a difference in one of these demographical or stroke related parameters between groups, we will correct for that to address our main research question.

VERSION 2 – REVIEW

REVIEWER	Teresa Neeman Australian National University Australia
REVIEW RETURNED	21-Sep-2019

GENERAL COMMENTS	The authors have addressed my concerns - thanks!